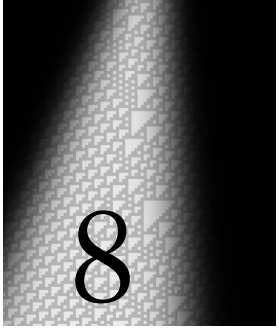
EXCERPTED FROM

STEPHEN WOLFRAM

A NEW KIND OF SCIENCE

CHAPTER 8

Implications for Everyday Systems



8

Implications for Everyday Systems

Issues of Modelling

In the previous chapter I showed how various general forms of behavior that are common in nature can be understood by thinking in terms of simple programs. In this chapter what I will do is to take what we have learned, and look at a sequence of fairly specific kinds of systems in nature and elsewhere, and in each case discuss how the most obvious features of their behavior arise.

The majority of the systems I consider are quite familiar from everyday life, and at first one might assume that the origins of their behavior would long ago have been discovered. But in fact, in almost all cases, rather little turns out to be known, and indeed at any fundamental level the behavior that is observed has often in the past seemed quite mysterious. But what we will discover in this chapter is that by thinking in terms of simple programs, the fundamental origins of this behavior become much less mysterious.

It should be said at the outset that it is not my purpose to explain every detail of all the various kinds of systems that I discuss. And in fact, to do this for even just one kind of system would most likely take at least another whole book, if not much more.

But what I do want to do is to identify the basic mechanisms that are responsible for the most obvious features of the behavior of each kind of system. I want to understand, for example, how in general

snowflakes come to have the intricate shapes they do. But I am not concerned, for example, with details such as what the precise curvature of the tips of the arms of the snowflake will be.

In most cases the basic approach I take is to try to construct the very simplest possible model for each system. From the intuition of traditional science we might think that if the behavior of a system is complex, then any model for the system must also somehow be correspondingly complex.

But one of the central discoveries of this book is that this is not in fact the case, and that at least if one thinks in terms of programs rather than traditional mathematical equations, then even models that are based on extremely simple underlying rules can yield behavior of great complexity. And in fact in the course of this chapter, I will construct a whole sequence of remarkably simple models that do rather well at reproducing the main features of complex behavior in a wide range of everyday natural and other systems.

Any model is ultimately an idealization in which only certain aspects of a system are captured, and others are ignored. And certainly in each kind of system that I consider here there are many details that the models I discuss do not address. But in most cases there have in the past never really been models that can even reproduce the most obvious features of the behavior we see. So it is already major progress that the models I discuss yield pictures that look even roughly right.

In many traditional fields of science any model which could yield such pictures would immediately be considered highly successful. But in some fields—especially those where traditional mathematics has been used the most extensively—it has come to be believed that in a sense the only truly objective or scientific way to test a model is to look at certain rather specific details.

Most often what is done is to extract a small set of numbers from the observed behavior of a system, and then to see how accurately these numbers can be reproduced by the model. And for systems whose overall behavior is fairly simple, this approach indeed often works quite well. But when the overall behavior is complex, it becomes impossible to characterize it in any complete way by just a few numbers.

And indeed in the literature of traditional science I have quite often seen models which were taken very seriously because they could be made to reproduce a few specific numbers, but which are shown up as completely wrong if one works out the overall behavior that they imply. And in my experience by far the best first step in assessing a model is not to look at numbers or other details, but rather just to use one's eyes, and to compare overall pictures of a system with pictures from the model.

If there are almost no similarities then one can reasonably conclude that the model is wrong. But if there are some similarities and some differences, then one must decide whether or not the differences are crucial. Quite often this will depend, at least in part, on how one intends to use the model. But with appropriate judgement it is usually not too difficult from looking at overall behavior to get at least some sense of whether a particular model is on the right track.

Typically it is not a good sign if the model ends up being almost as complicated as the phenomenon it purports to describe. And it is an even worse sign if when new observations are made the model constantly needs to be patched in order to account for them.

It is usually a good sign on the other hand if a model is simple, yet still manages to reproduce, even quite roughly, a large number of features of a particular system. And it is an even better sign if a fair fraction of these features are ones that were not known, or at least not explicitly considered, when the model was first constructed.

One might perhaps think that in the end one could always tell whether a model was correct by explicitly looking at sufficiently low-level underlying elements in a system and comparing them with elements in the model. But one must realize that a model is only ever supposed to provide an abstract representation of a system—and there is nothing to say that the various elements in this representation need have any direct correspondence with the elements of the system itself.

Thus, for example, a traditional mathematical model might say that the motion of a planet is governed by a set of differential equations. But one does not imagine that this means that the planet itself contains a device that explicitly solves such equations. Rather, the idea is that

the equations provide some kind of abstract representation for the physical effects that actually determine the motion of the planet.

When I have discussed models like the ones in this chapter with other scientists I have however often encountered great confusion about such issues. Perhaps it is because in a simple program it is so easy to see the underlying elements and the rules that govern them. But countless times I have been asked how models based on simple programs can possibly be correct, since even though they may successfully reproduce the behavior of some system, one can plainly see that the system itself does not, for example, actually consist of discrete cells that, say, follow the rules of a cellular automaton.

But the whole point is that all any model is supposed to do—whether it is a cellular automaton, a differential equation, or anything else—is to provide an abstract representation of effects that are important in determining the behavior of a system. And below the level of these effects there is no reason that the model should actually operate like the system itself.

Thus, for example, a cellular automaton can readily be set up to represent the effect of an inhibition on growth at points on the surface of a snowflake where new material has recently been added. But in the cellular automaton this effect is just implemented by some rule for certain configurations of cells—and there is no need for the rule to correspond in any way to the detailed dynamics of water molecules.

So even though there need not be any correspondence between elements in a system and in a model, one might imagine that there must still be some kind of complete correspondence between effects. But the whole point of a model is to have a simplified representation of a system, from which those features in which one is interested can readily be deduced or understood. And the only way to achieve this is to pick out only certain effects that are important, and to ignore all others.

Indeed, in practice, the main challenge in constructing models is precisely to identify which effects are important enough that they have to be kept, and which are not. In some simple situations, it is sometimes possible to set up experiments in which one can essentially isolate each individual effect and explicitly measure its importance. But

in the majority of cases the best evidence that some particular set of effects are in fact the important ones ultimately comes just from the success of models that are based on these effects.

The systems that I discuss in this chapter are mostly complicated enough that there are at least tens of quite different effects that could contribute to their overall behavior. But in trying to construct the simplest possible models, I have always picked out just a few effects that I believe will be the most important. Inevitably there will be phenomena that depend on other effects, and which are therefore not correctly reproduced by the models I consider. So if these phenomena are crucial to some particular application, then there will be no choice but to extend the model for that application.

But insofar as the goal is to understand the basic mechanisms that are responsible for the most obvious features of overall behavior, it is important to keep the underlying model as simple as possible. For even with just a few extensions models usually become so complicated that it is almost impossible to tell where any particular feature of behavior really comes from.

Over the years I have been able to watch the progress of perhaps a dozen significant models that I have constructed—though in most cases never published—for a variety of kinds of systems with complex behavior. My original models have typically been extremely simple. And the initial response to them has usually been great surprise that such simple models could ever yield behavior that has even roughly the right features. But experts in the particular types of systems involved have usually been quick to point out that there are many details that my models do not correctly reproduce.

Then after an initial period where the models are often said to be too simplistic to be worth considering, there begin to be all sorts of extensions added that attempt to capture more effects and more details. The result of this is that after a few years my original models have evolved into models that are almost unrecognizably complex. But these models have often then been used with great success for many practical purposes. And at that point, with their success established, it sometimes happens that the models are examined more carefully—and

it is then discovered that many of the extensions that were added were in fact quite unnecessary, so that in the end, after perhaps a decade has passed, it becomes recognized that models equivalent to the simple ones I originally proposed do indeed work quite well.

One might have thought that in the literature of traditional science new models would be proposed all the time. But in fact the vast majority of what is done in practically every field of science involves not developing new models but rather accumulating experimental data or working out consequences of existing models.

And among the models that have been used, almost all those that have gone beyond the level of being purely descriptive have ended up being formulated in very much the same kind of way: typically as collections of mathematical equations. Yet as I emphasized at the very beginning of this book, this is, I believe, the main reason that in the past it has been so difficult to find workable models for systems whose behavior is complex. And indeed it is one of the central ideas of this book to go beyond mathematical equations, and to consider models that are based on programs which can effectively involve rules of any kind.

It is in many respects easier to work with programs than with equations. For once one has a program, one can always find out what its behavior will be just by running it. Yet with an equation one may need to do elaborate mathematical analysis in order to find out what behavior it can lead to. It does not help that models based on equations are often stated in a purely implicit form, so that rather than giving an actual procedure for determining how a system will behave—as a program does—they just give constraints on what the behavior must be, and provide no particular guidance about finding out what, if any, behavior will in fact satisfy these constraints.

And even when models based on equations can be written in an explicit form, they still typically involve continuous variables which cannot for example be handled directly by a practical computer. When their overall behavior is sufficiently simple, complete mathematical formulas to describe this behavior can sometimes be found. But as soon as the behavior is more complex there is usually no choice but to use some form of approximation. And despite many attempts over the past

fifty or so years, it has almost never been possible to demonstrate that results obtained from such approximations even correctly reproduce what the original mathematical equations would imply.

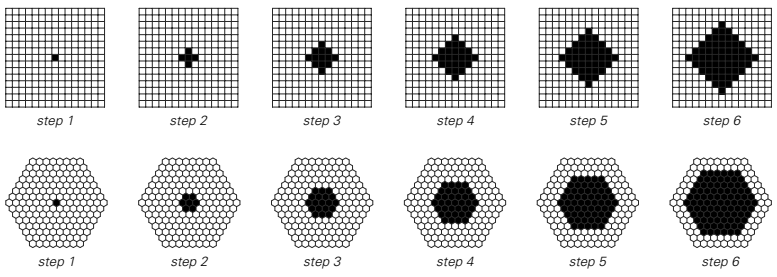
Models based on simple programs, however, suffer from no such problems. For essentially all of them involve only discrete elements which can be handled quite directly on a practical computer. And this means that it becomes straightforward in principle—and often highly efficient in practice—to work out at least the basic consequences of such models.

Many of the models that I discuss in this chapter are actually based on some of the very simplest kinds of programs that I consider anywhere in this book. But as we shall see, even these models appear quite sufficient to capture the behavior of a remarkably wide range of systems from nature and elsewhere—establishing beyond any doubt, I believe, the practical value of thinking in terms of simple programs.

The Growth of Crystals

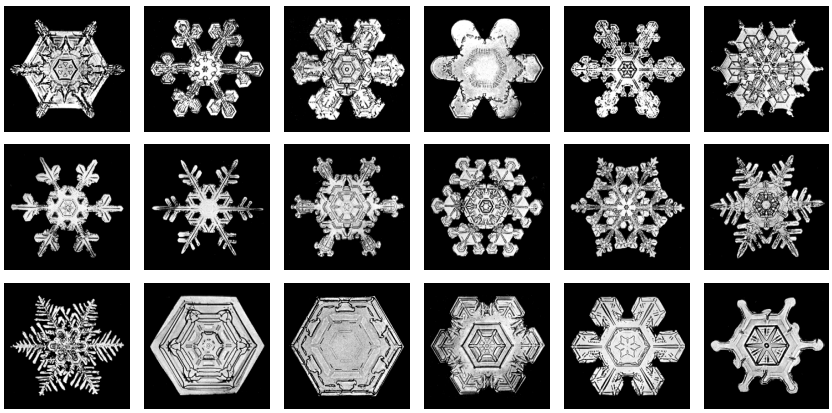
At a microscopic level crystals consist of regular arrays of atoms laid out much like the cells in a cellular automaton. A crystal forms when a liquid or gas is cooled below its freezing point. Crystals always start from a seed—often a foreign object such as a grain of dust—and then grow by progressively adding more atoms to their surface.

As an idealization of this process, one can consider a cellular automaton in which black cells represent regions of solid and white cells represent regions of liquid or gas. If one assumes that any cell which is adjacent to a black cell will itself become black on the next step, then one gets the patterns of growth shown below.



Cellular automata with rules that specify that a cell should become black if any of its neighbors are already black. The patterns produced have a simple faceted form that reflects directly the structure of the underlying lattice of cells.

The shapes produced in each case are very simple, and ultimately consist just of flat facets arranged in a way that reflects directly the structure of the underlying lattice of cells. And many crystals in nature—including for example most gemstones—have similarly simple faceted forms. But some do not. And as one well-known example, snowflakes can have highly intricate forms, as illustrated below.

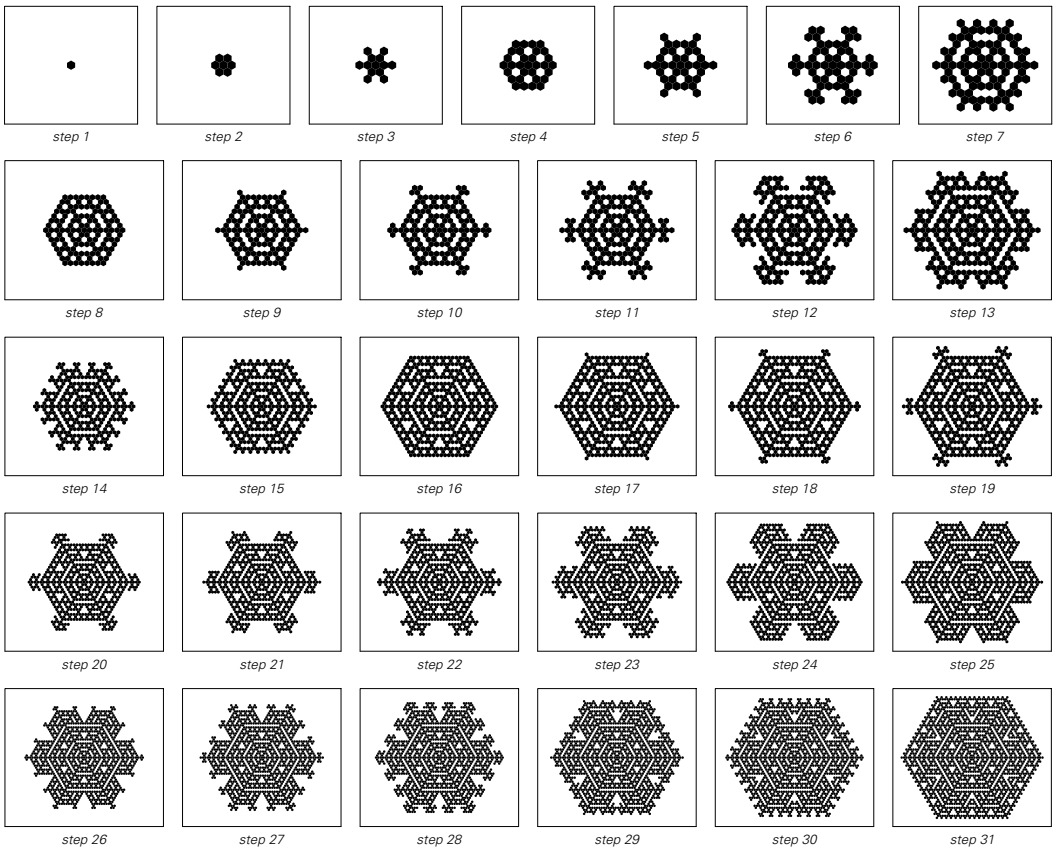


Examples of typical forms of snowflakes. Note that the scales for different pictures are different.

To a good approximation, all the molecules in a snowflake ultimately lie on a simple hexagonal grid. But in the actual process of snowflake growth, not every possible part of this grid ends up being filled with ice. The main effect responsible for this is that whenever a piece of ice is added to the snowflake, there is some heat released, which then tends to inhibit the addition of further pieces of ice nearby.

One can capture this basic effect by having a cellular automaton with rules in which cells become black if they have exactly one black neighbor, but stay white whenever they have more than one black neighbor. The pictures on the facing page show a sequence of steps in the evolution of such a cellular automaton. And despite the simplicity of its underlying rules, what one sees is that the patterns it produces are strikingly similar to those seen in real snowflakes.

From looking at the behavior of the cellular automaton, one can immediately make various predictions about snowflakes. For example,



The evolution of a cellular automaton in which each cell on a hexagonal grid becomes black whenever exactly one of its neighbors was black on the step before. This rule captures the basic growth inhibition effect that occurs in snowflakes. The resulting patterns obtained at different steps look remarkably similar to many real snowflakes.

one expects that during the growth of a particular snowflake there should be alternation between tree-like and faceted shapes, as new branches grow but then collide with each other.

And if one looks at real snowflakes, there is every indication that this is exactly what happens. And in fact, in general the simple cellular automaton shown above seems remarkably successful at reproducing all sorts of obvious features of snowflake growth. But inevitably there are many details that it does not capture. And indeed some of the photographs on the facing page do not in the end look much like patterns produced at any step in the evolution shown above.

But it turns out that as soon as one tries to make a more complete model, there are immediately an immense number of issues that arise, and it is difficult to know which are really important and which are not. At a basic level, one knows that snowflakes are formed when water vapor in a cloud freezes into ice, and that the structure of a given snowflake is determined by the temperature and humidity of the environment in which it grows, and the length of time it spends there.

The growth inhibition mentioned above is a result of the fact that when water or water vapor freezes into ice, it releases a certain amount of latent heat—as the reverse of the phenomenon that when ice is warmed to 0°C it still needs heat applied before it will actually melt.

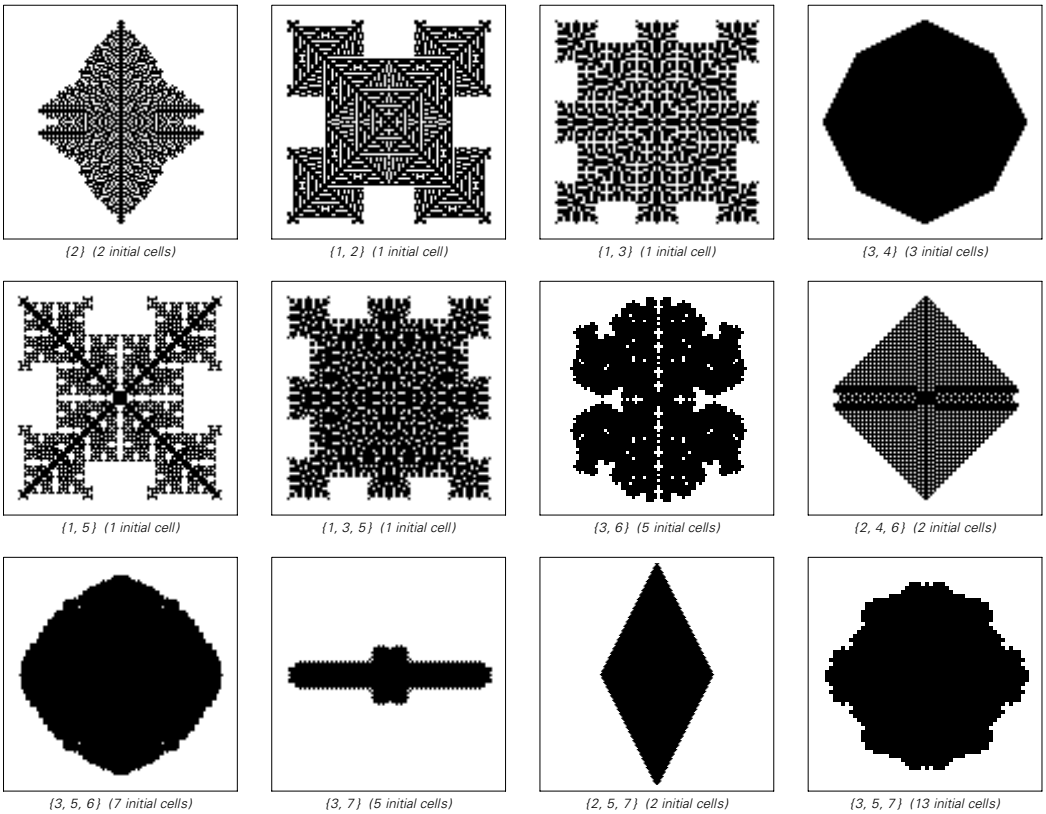
But there are also many effects. The freezing temperature, for example, effectively varies with the curvature of the surface. The rate of heat conduction differs in different directions on the hexagonal grid. Convection currents develop in the water vapor around the snowflake. Mechanical stresses are produced in the crystal as it grows.

Various models of snowflake growth exist in the standard scientific literature, typically focusing on one or two of these effects. But in most cases the models have at some basic level been rather unsuccessful. For being based on traditional mathematical equations they have tended to be able to deal only with what amount to fairly simple smooth shapes—and so have never really been able to address the kind of intricate structure that is so striking in real snowflakes.

But with models based on simple programs such as cellular automata, there is no problem in dealing with more complicated shapes, and indeed, as we have seen, it is actually quite easy to reproduce the basic features of the overall behavior that occurs in real snowflakes.

So what about other types of crystals?

In nature a variety of forms are seen. And as the pictures on the facing page demonstrate, the same is true even in cellular automata with very simple rules. Indeed, much as in nature, the diversity of behavior is striking. Sometimes simple faceted forms are produced. But in other cases there are needle-like forms, tree-like or dendritic forms, as well as rounded forms, and forms that seem in many respects random.



Examples of patterns produced by two-dimensional cellular automata set up to mimic the growth of crystals. The rules in each case take a cell to become black if the specified number of its neighbors (including diagonals) on a square grid are black on the step before. These rules are such that once a cell has become black, corresponding to solid, it never reverts to white again. In each case a row of initial black cells of the specified length was used.

The occurrence of these last forms is at first especially surprising. For one might have assumed that any apparent randomness in the final shape of something like a crystal must always be a consequence of randomness in its original seed, or in the environment in which it grew.

But in fact, as the pictures above show—and as we have seen many times in this book—it is also possible for randomness to arise intrinsically just through the application of simple underlying rules. And contrary to what has always been assumed, I suspect that this is actually how the apparent randomness that one sometimes sees in shapes formed by crystalline materials often comes about.

The Breaking of Materials

In everyday life one of the most familiar ways to generate randomness is to break a solid object. For although the details vary from one material to another it is almost universally the case that the line or surface along which fracture actually occurs seems rough and in many respects random.

So what is the origin of this randomness? At first one might think that it must be a reflection of random small-scale irregularities within the material. And indeed it is true that in materials that consist of many separate crystals or grains, fractures often tend to follow the boundaries between such elements.

But what happens if one takes for example a perfect single crystal—say a standard highly pure industrial silicon crystal—and breaks it? The answer is that except in a few special cases the pattern of fracture one gets seems to look just as random as in other materials.

And what this suggests is that whatever basic mechanism is responsible for such randomness, it cannot depend on the details of particular materials. Indeed, the fact that almost indistinguishable patterns of fracture are seen both at microscopic scales and in geological systems on scales of order kilometers is another clue that there must be a more general mechanism at work.

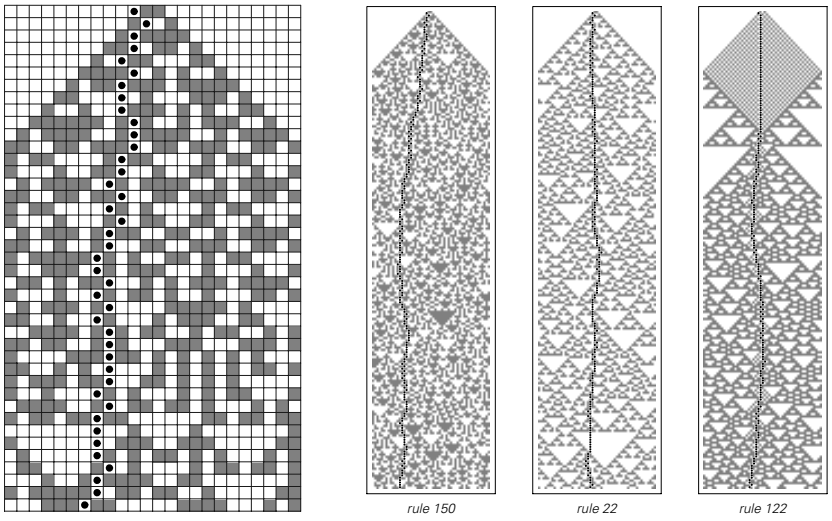
So what might this mechanism be?

When a solid material breaks what typically happens is that a crack forms—usually at the edge of the material—and then spreads. Experience with systems from hand-held objects to engineering structures and earthquakes suggests that it can take a while for a crack to get started, but that once it does, the crack tends to move quickly and violently, usually producing a lot of noise in the process.

One can think of the components of a solid—whether at the level of atoms, molecules, or pieces of rock—as being bound together by forces that act a little like springs. And when a crack propagates through the solid, this in effect sets up an elaborate pattern of vibrations in these springs. The path of the crack is then in turn determined by where the springs get stretched so far that they break.

There are many factors which affect the details of displacements and vibrations in a solid. But as a rough approximation one can perhaps assume that each element of a solid is either displaced or not, and that the displacements of neighboring elements interact by some definite rule—say a simple cellular automaton rule.

The pictures below show the behavior that one gets with a simple model of this kind. And even though there is no explicit randomness inserted into the model in any way, the paths of the cracks that emerge nevertheless appear to be quite random.



A very simple cellular automaton model for fracture. At each step, the color of each cell, which roughly represents the displacement of an element of the solid, is updated according to a cellular automaton rule. The black dot, representing the location of a crack, moves from one cell to another based on the displacements of neighboring cells, at each step setting the cell it reaches to be white. Even though no randomness is inserted from outside, the paths of the cracks that emerge from this model nevertheless appear to a large extent random. There is some evidence from physical experiments that dislocations around cracks can form patterns that look similar to the gray and white backgrounds above.

There are certainly many aspects of real materials that this model does not even come close to capturing. But I nevertheless suspect that even when much more realistic models for specific materials are used, the fundamental mechanisms responsible for randomness will still be very much the same as in the extremely simple model shown here.

Fluid Flow

A great many striking phenomena in nature involve the flow of fluids like air and water—as illustrated on the facing page. Typical of what happens is what one sees when water flows around a solid object. At sufficiently slow speeds, the water in effect just slides smoothly around, yielding a very simple laminar pattern of flow. But at higher speeds, there starts to be a region of slow-moving water behind the object, and a pair of eddies are formed as the water swirls into this region.

As the speed increases, these eddies become progressively more elongated. And then suddenly, when a critical speed is reached, the eddies in effect start breaking off, and getting carried downstream. But every time one eddy breaks off, another starts to form, so that in the end a whole street of eddies are seen in the wake behind the object.

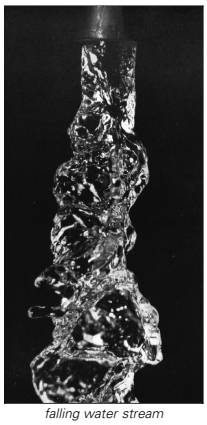
At first, these eddies are arranged in a very regular way. But as the speed of the flow is increased, glitches begin to appear, at first far behind the object, but eventually throughout the wake. Even at the highest speeds, some overall regularity nevertheless remains. But superimposed on this is all sorts of elaborate and seemingly quite random behavior.

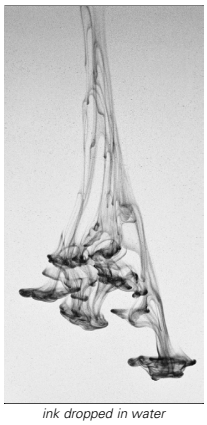
But this is just one example of the very widespread phenomenon of fluid turbulence. For as the pictures on the facing page indicate—and as common experience suggests—almost any time a fluid is made to flow rapidly, it tends to form complex patterns that seem in many ways random.

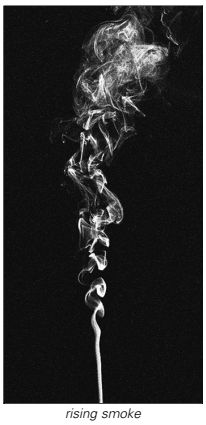
So why fundamentally does this happen?

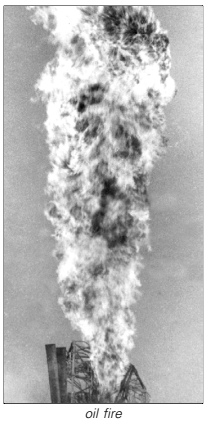
Traditional science, with its basis in mathematical equations, has never really been able to provide any convincing underlying explanation. But from my discovery that complex and seemingly random behavior is in a sense easy to get even with very simple programs, the phenomenon of fluid turbulence immediately begins to seem much less surprising.

But can simple programs really reproduce the particular kinds of behavior we see in fluids? At a microscopic level, physical fluids consist of large numbers of molecules moving around and colliding with each other. So as a simple idealization, one can consider having a large number of particles move around on a fixed discrete grid, and undergo collisions governed by simple cellular-automaton-like rules.

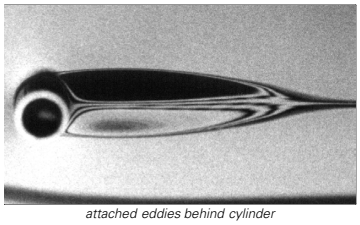

falling water stream

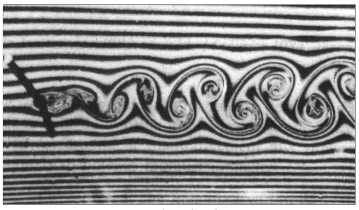

ink dropped in water

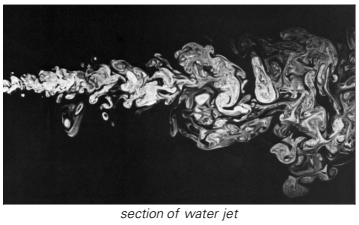

rising smoke

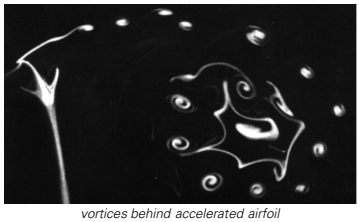

oil fire

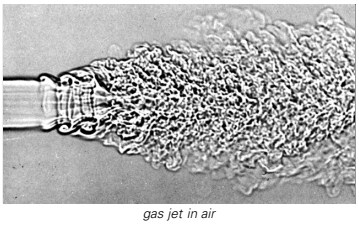

soap film turbulence

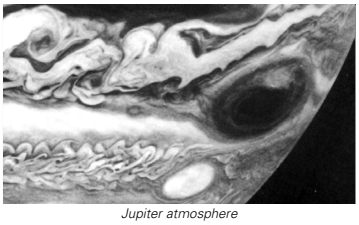

attached eddies behind cylinder

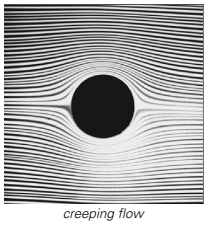

vortex street

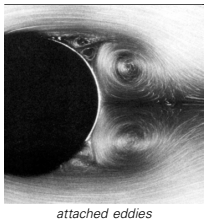

section of water jet

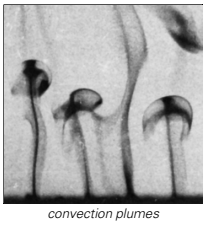

vortices behind accelerated airfoil

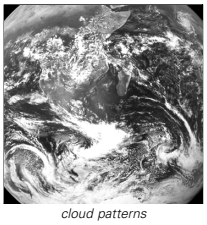

gas jet in air

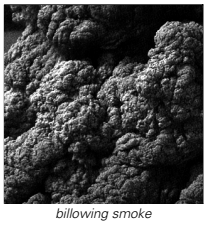

Jupiter atmosphere


creeping flow

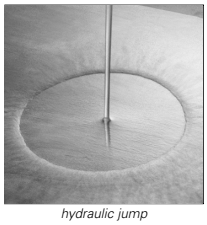

attached eddies

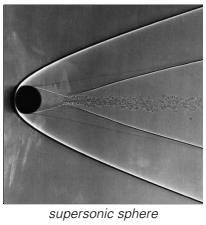

convection plumes

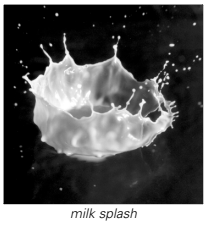

cloud patterns

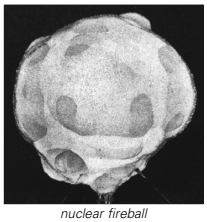

billowing smoke


convection cells


hydraulic jump

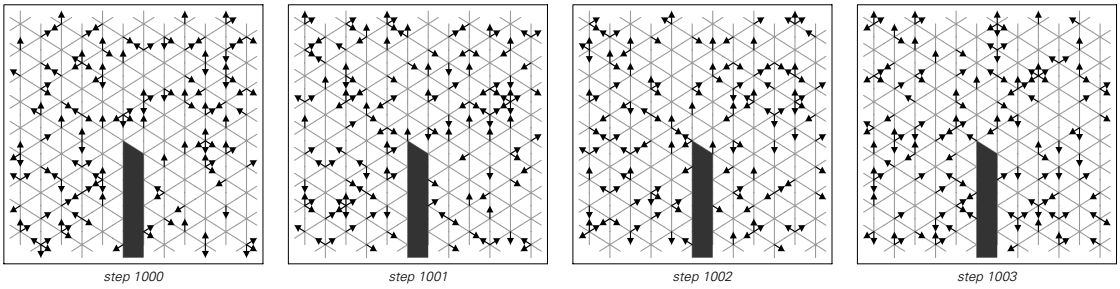

supersonic sphere

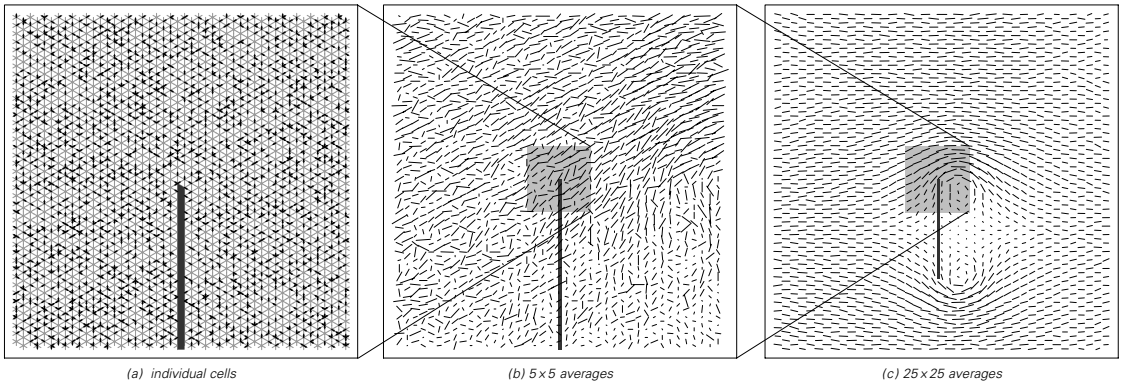

milk splash

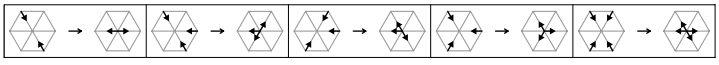

nuclear fireball

Examples of typical patterns generated in various kinds of fluid flow. Note the frequent occurrence of seemingly random turbulence.

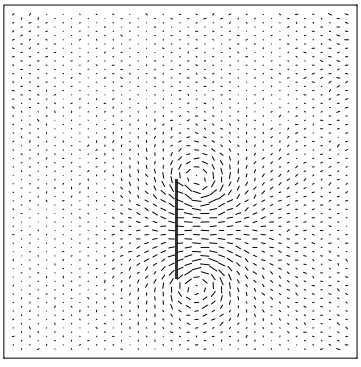
The pictures below give an example of such a system. In the top row of pictures—as well as picture (a)—all one sees is a collection of discrete particles bouncing around. But if one zooms out, and looks at average motion of increasingly large blocks of particles—as in pictures (b) and (c)—then what begins to emerge is behavior that seems smooth and continuous—just like one expects to see in a fluid.







A simple cellular automaton system set up to emulate the microscopic behavior of molecules in a fluid. At each step the configuration of particles is updated according to the simple collision rules shown above. Particles are reflected whenever they hit the plate. A steady stream of particles is inserted in a regular way far to the left, with an average speed 3/10 of the maximum possible. Picture (a) shows the configuration of individual particles; pictures (b) and (c) show total velocities of successively larger blocks of particles. Picture (d) is obtained by transforming to a reference frame in which the fluid is on average at rest.



This happens for exactly the same reason as in a real fluid, or, for that matter, in various examples that we saw in Chapter 7: even though at an underlying level the system consists of discrete particles, the effective randomness of the detailed microscopic motions of these particles makes their large-scale average behavior seem smooth and continuous.

We know from physical experiments that the characteristics of fluid flow are almost exactly the same for air, water, and all other ordinary fluids. Yet at an underlying level these different fluids consist of very different kinds of molecules, with very different properties. But somehow the details of such microscopic structure gets washed out if one looks at large-scale fluid-like behavior.

Many times in this book we have seen examples where different systems can yield very much the same overall behavior, even though the details of their underlying rules are quite different. But in the particular case of systems like fluids, it turns out that one can show—as I will discuss in the next chapter—that so long as certain physical quantities such as particle number and momentum are conserved, then whenever there is sufficient microscopic randomness, it is almost inevitable that the same overall fluid behavior will be obtained.

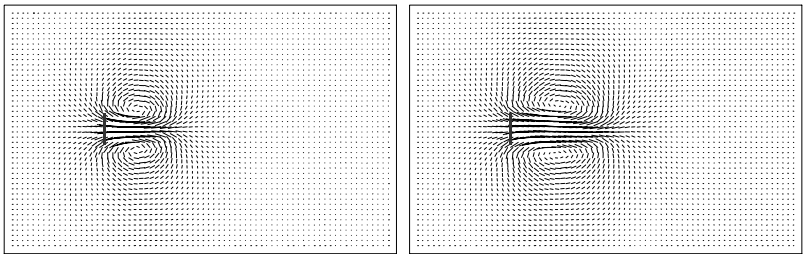
So what this means is that to reproduce the observed properties of physical fluids one should not need to make a model that involves realistic molecules: even the highly idealized particles on the facing page should give rise to essentially the same overall fluid behavior.

And indeed in pictures (c) and (d) one can already see the formation of a pair of eddies, just as in one of the pictures on page 377.

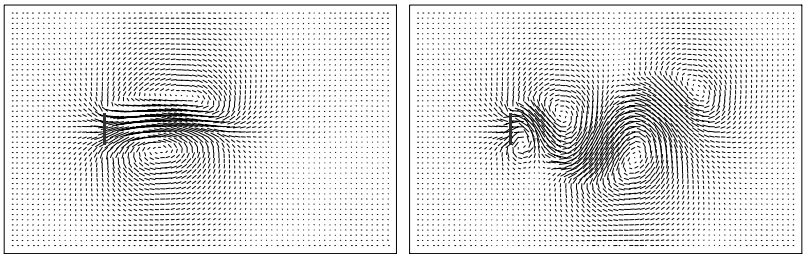
So what happens if one increases the speed of the flow? Does one see the same kinds of phenomena as on page 377? The pictures on the next page suggest that indeed one does. Below a certain critical speed, a completely regular array of eddies is formed. But at the speed used in the pictures on the next page, the array of eddies has begun to show random irregularities just like those associated with turbulence in real fluids.

So where does this randomness come from?

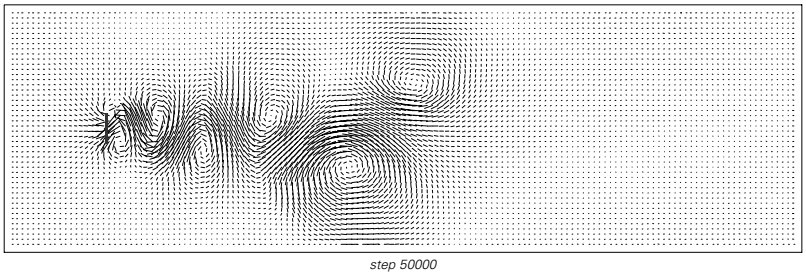
In the past couple of decades it has come to be widely believed that randomness in turbulent fluids must somehow be associated with



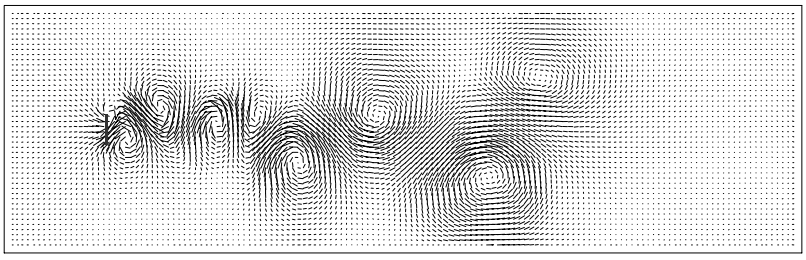
step 10000 step 20000



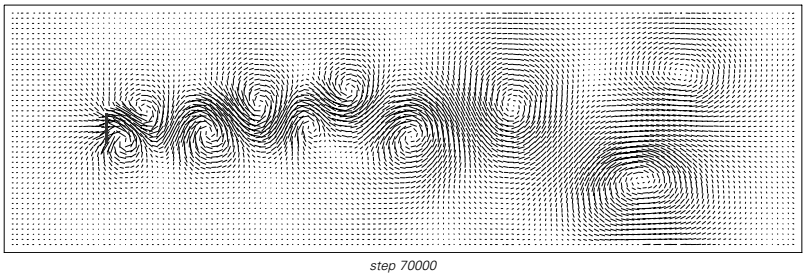
step 30000 step 40000



step 50000



step 60000



step 70000

A larger example of the cellular automaton system shown on the previous page. In each picture there are a total of 30 million underlying cells. The individual velocity vectors drawn correspond to averages over 20 × 20 blocks of cells. Particles are inserted in a regular way at the left-hand end so as to maintain an overall flow speed equal to about 0.4 of the maximum possible. To make the patterns of flow easier to see, the velocities shown are transformed so that the fluid is on average at rest, and the plate is moving. The underlying density of particles is approximately 1 per cell, or 1/6 the maximum possible—a density which more or less minimizes the viscosity of the fluid. The Reynolds number of the flow shown is then approximately 100. The agreement with experimental results on actual fluid flows is striking.

sensitive dependence on initial conditions, and with the chaos phenomenon that we discussed in Chapter 4.

But while there are certainly mathematical equations that exhibit this phenomenon, none of those typically investigated have any close connection to realistic descriptions of fluid flow.

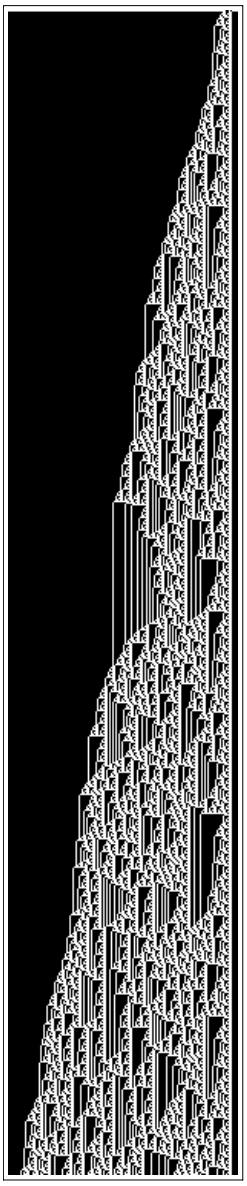
And in the model on the facing page it turns out that there is essentially no sensitive dependence on initial conditions, at least at the level of overall fluid behavior. If one looks at individual particles, then changing the position of even one particle will typically have an effect that spreads rapidly. But if one looks instead at the average behavior of many particles, such effects get completely washed out. And indeed when it comes to large-scale fluid behavior, it seems to be true that in almost all cases there is no discernible difference between what happens with different detailed initial conditions.

So is there ever sensitive dependence on initial conditions?

Presumably there do exist situations in which there is some kind of delicate balance—say of whether the first eddy is shed at the top or bottom of an object—and in which small changes in initial conditions can have a substantial effect. But such situations appear to be very much the exception rather than the rule. And in the vast majority of cases, small changes instead seem to damp out rapidly—just as one might expect from everyday experience with viscosity in fluids.

So what this means is that the randomness we observe in fluid flow cannot simply be a reflection of randomness that is inserted through the details of initial conditions. And as it turns out, in the pictures on the facing page, the initial conditions were specifically set up to be very simple. Yet despite this, there is still apparent randomness in the overall behavior that is seen.

And so, once again, just as for many other systems that we have studied in this book, there is little choice but to conclude that in a turbulent fluid most of the randomness we see is not in any way inserted from outside but is instead intrinsically generated inside the system itself. In the pictures on page 378 considerable randomness was already evident at the level of individual particles. But since changes in the configurations of such particles do not seem to have any discernible



A cellular automaton (rule 225) whose behavior is reminiscent of turbulent fluid flow.

effect on overall patterns of flow, one cannot realistically attribute the large-scale randomness that one sees in a turbulent fluid to randomness that exists at the level of individual particles.

Instead, what seems to be happening is that intrinsic randomness generation occurs directly at the level of large-scale fluid motion. And as an example of a simple approach to modelling this, one can consider having a collection of discrete eddies that occur at discrete positions in the fluid, and interact through simple cellular automaton rules.

The picture on the left shows an example of what can happen. And although many details are different from what one sees in real fluids, the overall mixture of regularity and randomness is strikingly similar.

One consequence of the idea that there is intrinsic randomness generation in fluids and that it occurs at the level of large-scale fluid motion is that with sufficiently careful preparation it should be possible to produce patterns of flow that seem quite random but that are nevertheless effectively repeatable—so that they look essentially the same on every successive run of an experiment.

And even if one looks at existing experiments on fluid flow, there turn out to be quite a few instances—particularly for example involving interactions between small numbers of vortices—where there are known patterns of fluid flow that look intricate, but are nevertheless essentially repeatable. And while none of these yet look complicated enough that they might reasonably be called random, I suspect that in time similar but vastly more complex examples will be found.

Among the patterns of fluid flow on page 377 each has its own particular details and characteristics. But while some of the simpler ones have been captured quite completely by methods based on traditional mathematical equations, the more complex ones have not. And in fact from the perspective of this book this is not surprising.

But now from the experience and intuition developed from the discoveries in this book, I expect that there will in fact be remarkably simple programs that can be found that will successfully manage to reproduce the main features of even the most intricate and apparently random forms of fluid flow.

Fundamental Issues in Biology

Biological systems are often cited as supreme examples of complexity in nature, and it is not uncommon for it to be assumed that their complexity must be somehow of a fundamentally higher order than other systems.

And typically it is thought that this must be a consequence of the rather unique processes of adaptation and natural selection that operate in biological systems. But despite all sorts of discussion over the years, no clear understanding has ever emerged of just why such processes should in the end actually lead to much complexity at all.

And in fact what I have come to believe is that many of the most obvious examples of complexity in biological systems actually have very little to do with adaptation or natural selection. And instead what I suspect is that they are mainly just another consequence of the very basic phenomenon that I have discovered in this book in the context of simple programs: that in almost any kind of system many choices of underlying rules inevitably lead to behavior of great complexity.

The general idea of thinking in terms of programs is, if anything, even more obvious for biological systems than for physical ones. For in a physical system the rules of a program must normally be deduced indirectly from the laws of physics. But in a biological organism there is genetic material which can be thought of quite directly as providing a program for the development of the organism.

Most of the programs that I have discussed in this book, however, have been very simple. Yet the genetic program for every biological organism known today is long and complicated: in humans, for example, it presumably involves millions of separate rules—making it by most measures as complex as large practical software systems like *Mathematica*.

So from this one might think that the complexity we see in biological organisms must all just be a reflection of complexity in their underlying rules—making discoveries about simple programs not really relevant. And certainly the presence of many different types of organs and other elements in a typical complete organism seems likely to be related to the presence of many separate sets of rules in the underlying

program. But what if one looks not at a complete organism but instead just at some part of an organism?

Particularly on a microscopic scale, the forms one sees are often highly regular and quite simple, as in the pictures on the facing page. And when one looks at these, it seems perfectly reasonable to suppose that they are in effect produced by fairly simple programs.

But what about the much more complicated forms that one sees in biological systems? On the basis of traditional intuition one might assume that such forms could never be produced by simple programs. But from the discoveries in this book we now know that in fact it is possible to get remarkable complexity even from very simple programs.

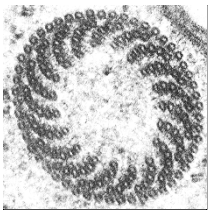
So is this what actually happens in biological systems?

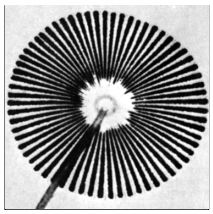
There is certainly no dramatic difference between the underlying types of cells or other elements that occur in complex biological forms and in the forms on the facing page. And from this one might begin to suspect that in the end the kinds of programs which generate all these forms are quite similar—and all potentially rather simple.

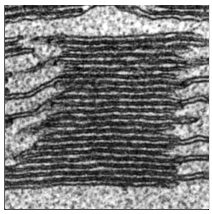
For even though the complete genetic program for an organism is long and complicated, the subprograms which govern individual aspects of an organism can still be simple—and there are now plenty of specific simple examples where this is known to be the case. But still one might assume that to get significant complexity would require something more. And indeed at first one might think that it would never really be possible to say much at all about complexity just by looking at parts of organisms.

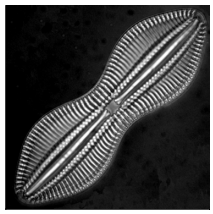
But in fact, as it turns out, a rather large fraction of the most obvious examples of biological complexity seem to involve only surprisingly limited parts of the organisms. Elaborate pigmentation patterns, for instance, typically exist just on an outer skin, and are made up of only a few types of cells. And the vast majority of complicated

Examples of highly regular forms occurring in biological systems. Most of these forms are simple enough that it seems immediately plausible that they could in effect be generated by simple programs. The majority show either simple geometrical shapes, or repetition of identical elements. A few, however, show various types of nesting. Note that there seems to be no obvious correlation between the sophistication of a form and when in geological time it first appeared. ▸

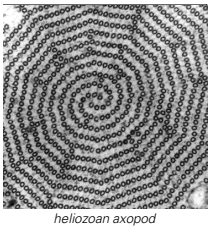

protist microspine section

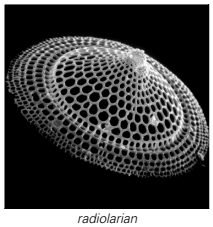

alga

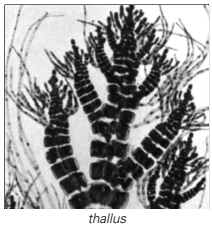

chloroplast

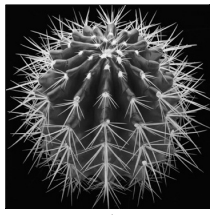

diatom

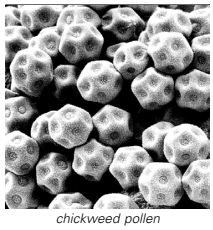

coccolithophorid

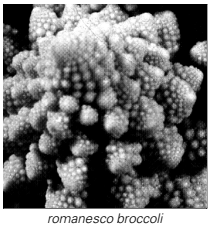

heliozoan axopod

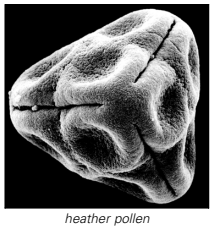

radiolarian

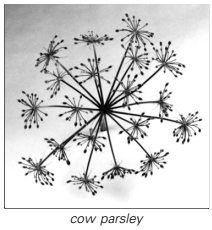

thallus

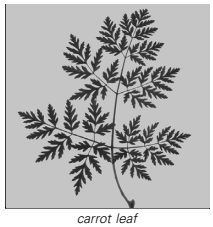

cactus

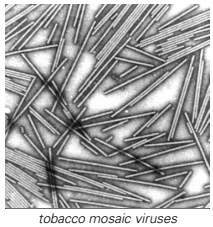

chickweed pollen

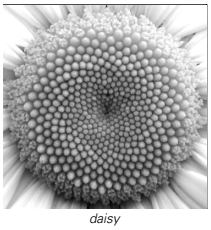

romanesco broccoli

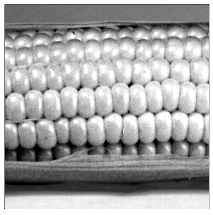

heather pollen

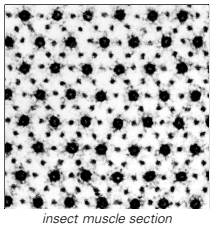

cow parsley

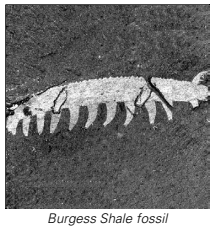

carrot leaf

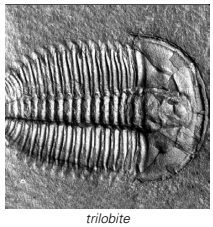

tobacco mosaic viruses


daisy

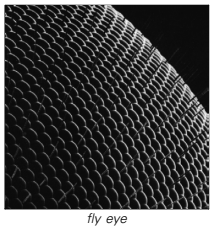

corn

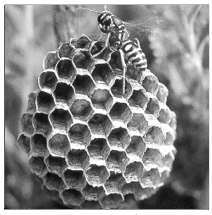

insect muscle section

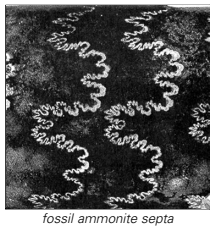

Burgess Shale fossil

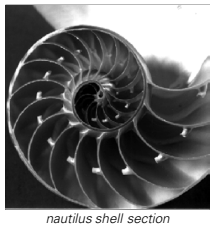

trilobite

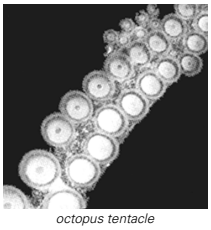

barnacle

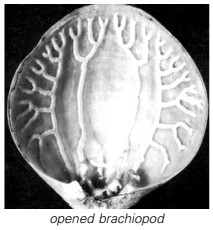

fly eye

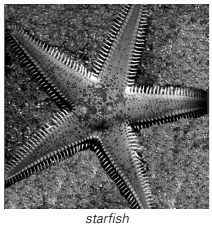

wasp nest


fossil ammonite septa

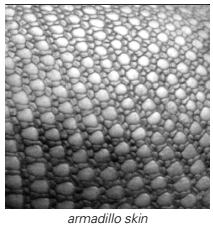

nautilus shell section


octopus tentacle


opened brachiopod


starfish


sea urchin


armadillo skin

morphological structures get their forms from arrangements of very limited numbers of types of cells or other elements.

But just how are the programs for these and other features of organisms actually determined? Over the past century or so it has become almost universally believed that at some level these programs must end up being the ones that maximize the fitness of the organism, and the number of viable offspring it produces.

The notion is that if a line of organisms with a particular program typically produce more offspring, then after a few generations there will inevitably be vastly more organisms with this program than with other programs. And if one assumes that the program for each new offspring involves small random mutations then this means that over the course of many generations biological evolution will in effect carry out a random search for programs that maximize the fitness of an organism.

But how successful can one expect such a search to be?

The problem of maximizing fitness is essentially the same as the problem of satisfying constraints that we discussed at the end of Chapter 7. And what we found there is that for sufficiently simple constraints—particularly continuous ones—iterative random searches can converge fairly quickly to an optimal solution. But as soon as the constraints are more complicated this is no longer the case. And indeed even when the optimal solution is comparatively simple it can require an astronomically large number of steps to get even anywhere close to it.

Biological systems do appear to have some tricks for speeding up the search process. Sexual reproduction, for example, allows large-scale mixing of similar programs, rather than just small-scale mutation. And differentiation into organs in effect allows different parts of a program to be updated separately. But even with a whole array of such tricks, it is still completely implausible that the trillion or so generations of organisms since the beginning of life on Earth would be sufficient to allow optimal solutions to be found to constraints of any significant complexity.

And indeed one suspects that in fact the vast majority of features of biological organisms do not correspond to anything close to optimal solutions: rather, they represent solutions that were fairly easy to find, but are good enough not to cause fatal problems for the organism.

The basic notion that organisms tend to evolve to achieve a maximum fitness has certainly in the past been very useful in providing a general framework for understanding the historical progression of species, and in yielding specific explanations for various fairly simple properties of particular species.

But in present-day thinking about biology the notion has tended to be taken to an extreme, so that especially among those not in daily contact with detailed data on biological systems it has come to be assumed that essentially every feature of every organism can be explained on the basis of it somehow maximizing the fitness of the organism.

It is certainly recognized that some aspects of current organisms are in effect holdovers from earlier stages in biological evolution. And there is also increasing awareness that the actual process of growth and development within an individual organism can make it easier or more difficult for particular kinds of structures to occur.

But beyond this there is a surprisingly universal conviction that any significant property that one sees in any organism must be there because it in essence serves a purpose in maximizing the fitness of the organism.

Often it is at first quite unclear what this purpose might be, but at least in fairly simple cases, some kind of hypothesis can usually be constructed. And having settled on a supposed purpose it often seems quite marvellous how ingenious biology has been in finding a solution that achieves that purpose.

Thus, for example, the golden ratio spiral of branches on a plant stem can be viewed as a marvellous way to minimize the shading of leaves, while the elaborate patterns on certain mollusc shells can be viewed as marvellous ways to confuse the visual systems of supposed predators.

But it is my strong suspicion that such purposes in fact have very little to do with the real reasons that these particular features exist. For instead, as I will discuss in the next couple of sections, what I believe is that these features actually arise in essence just because they are easy to produce with fairly simple programs. And indeed as one looks at more and more complex features of biological organisms—notably texture and pigmentation patterns—it becomes increasingly difficult to find any credible purpose at all that would be served by the details of what one sees.

In the past, the idea of optimization for some sophisticated purpose seemed to be the only conceivable explanation for the level of complexity that is seen in many biological systems. But with the discovery in this book that it takes only a simple program to produce behavior of great complexity, a quite different—and ultimately much more predictive—kind of explanation immediately becomes possible.

In the course of biological evolution random mutations will in effect cause a whole sequence of programs to be tried. And the point is that from what we have discovered in this book, we now know that it is almost inevitable that a fair fraction of these programs will yield complex behavior.

Some programs will presumably lead to organisms that are more successful than others, and natural selection will cause these programs eventually to dominate. But in most cases I strongly suspect that it is comparatively coarse features that tend to determine the success of an organism—not all the details of any complex behavior that may occur.

Thus in a very simple case it is easy to imagine for example that an organism might be more likely to go unnoticed by its predators, and thus survive and be more successful, if its skin was a mixture of brown and white, rather than, say, uniformly bright orange. But it could then be that most programs which yield any mixture of colors also happen to be such that they make the colors occur in a highly complex pattern.

And if this is so, then in the course of random mutation, the chances are that the first program encountered that is successful enough to survive will also, quite coincidentally, exhibit complex behavior.

On the basis of traditional biological thinking one would tend to assume that whatever complexity one saw must in the end be carefully crafted to satisfy some elaborate set of constraints. But what I believe instead is that the vast majority of the complexity we see in biological systems actually has its origin in the purely abstract fact that among randomly chosen programs many give rise to complex behavior.

In the past it tends to have been implicitly assumed that to get substantial complexity in a biological system must somehow be fundamentally very difficult. But from the discoveries in this book I have come to the conclusion that instead it is actually rather easy.

So how can one tell if this is really the case?

One circumstantial piece of evidence is that one already sees considerable complexity even in very early fossil organisms. Over the course of the past billion or so years, more and more organs and other devices have appeared. But the most obvious outward signs of complexity, manifest for example in textures and other morphological features, seem to have already been present even from very early times.

And indeed there is every indication that the level of complexity of individual parts of organisms has not changed much in at least several hundred million years. So this suggests that somehow the complexity we see must arise from some straightforward and general mechanism—and not, for example, from a mechanism that relies on elaborate refinement through a long process of biological evolution.

Another circumstantial piece of evidence that complexity is in a sense easy to get in biological systems comes from the observation that among otherwise very similar present-day organisms features such as pigmentation patterns often vary from quite simple to highly complex.

Whether one looks at fishes, butterflies, molluscs or practically any other kind of organism, it is common to find that across species or even within species organisms that live in the same environment and have essentially the same internal structure can nevertheless exhibit radically different pigmentation patterns. In some cases the patterns may be simple, but in other cases they are highly complex.

And the point is that no elaborate structural changes and no sophisticated processes of adaptation seem to be needed in order to get these more complex patterns. And in the end it is, I suspect, just that some of the possible underlying genetic programs happen to produce complex patterns, while others do not.

Two sections from now I will discuss a rather striking potential example of this: if one looks at molluscs of various types, then it turns out that the range of pigmentation patterns on their shells corresponds remarkably closely with the range of patterns that are produced by simple randomly chosen programs based on cellular automata.

And examples like this—together with many others in the next couple of sections—provide evidence that the kind of complexity we see in biological organisms can indeed successfully be reproduced by short

and simple underlying programs. But there still remains the question of whether actual biological organisms really use such programs, or whether somehow they instead use much more complicated programs.

Modern molecular biology should soon be able to isolate the specific programs responsible, say, for the patterns on mollusc shells, and see explicitly how long they are. But there are already indications that these programs are quite short.

For one of the consequences of a program being short is that it has little room for inessential elements. And this means that almost any mutation or change in the program—however small—will tend to have a significant effect on at least the details of patterns it produces.

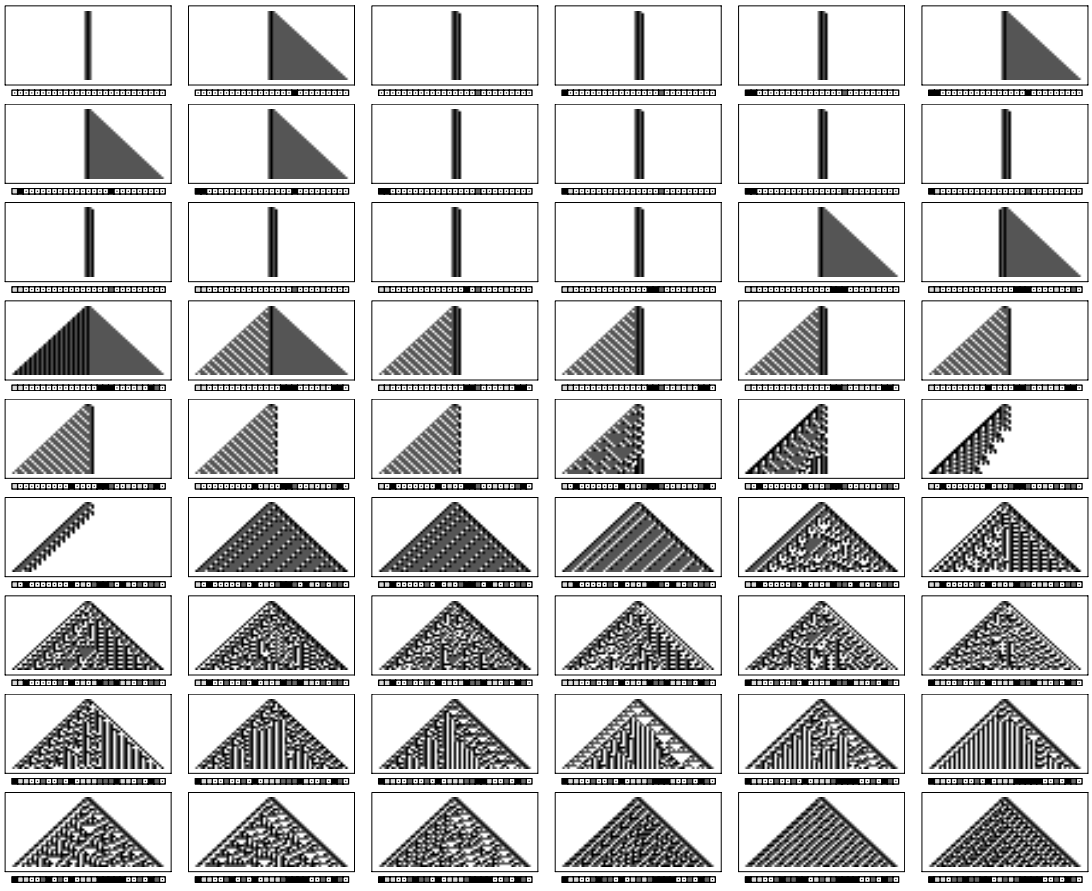
Sometimes it is hard to tell whether changes in patterns between organisms within a species are truly of genetic origin. But in cases where they appear to be it is common to find that different organisms show a considerable variety of different patterns—supporting the idea that the programs responsible for these patterns are indeed short.

So what about the actual process of biological evolution? How does it pick out which programs to use? As a very simple idealization of biological evolution, one can consider a sequence of cellular automaton programs in which each successive program is obtained from the previous one by a random mutation that adds or modifies a single element.

The pictures on the facing page then show a typical example of what happens with such a setup. If one starts from extremely short programs, the behavior one gets is at first quite simple. But as soon as the underlying programs become even slightly longer, one immediately sees highly complex behavior.

Traditional intuition would suggest that if the programs were to become still longer, the behavior would get ever richer and more complex. But from the discoveries in this book we know that this will not in general be the case: above a fairly low threshold, adding complexity to an underlying program does not fundamentally change the kind of behavior that it can produce.

And from this one concludes that biological systems should in a sense be capable of generating essentially arbitrary complexity by using short programs formed by just a few mutations.



The behavior of a sequence of cellular automaton programs obtained by successive random mutations. The first program contains no rules for changing the color of a cell with any neighborhood. Mutations in successive programs add rules for changing the colors of cells with specific neighborhoods, or modify these rules. Each program in the sequence differs from the previous one by a single mutation, made completely at random. The sequence provides a very simple idealization of biological evolution without explicit natural selection. The cellular automata shown here all have 3 possible colors and nearest-neighbor rules. The label for each picture gives a representation of the rules for each of the 27 possible 3-cell neighborhoods. A dot signifies that the rule does not change the color of the center cell in the neighborhood.

But if complexity is this easy to get, why is it not even more widespread in biology? For while there are certainly many examples of elaborate forms and patterns in biological systems, the overall shapes and many of the most obvious features of typical organisms are usually quite simple.

So why should this be? My guess is that in essence it reflects limitations associated with the process of natural selection. For while

natural selection is often touted as a force of almost arbitrary power, I have increasingly come to believe that in fact its power is remarkably limited. And indeed, what I suspect is that in the end natural selection can only operate in a meaningful way on systems or parts of systems whose behavior is in some sense quite simple.

If a particular part of an organism always grows, say, in a simple straight line, then it is fairly easy to imagine that natural selection could succeed in picking out the optimal length for any given environment. But what if an organism can grow in a more complex way, say like in the pictures on the previous page? My strong suspicion is that in such a case natural selection will normally be able to achieve very little.

There are several reasons for this, all somewhat related.

First, with more complex behavior, there are typically a huge number of possible variations, and in a realistic population of organisms it becomes infeasible for any significant fraction of these variations to be explored.

Second, complex behavior inevitably involves many elaborate details, and since different ones of these details may happen to be the deciding factors in the fates of individual organisms, it becomes very difficult for natural selection to act in a consistent and definitive way.

Third, whenever the overall behavior of a system is more complex than its underlying program, almost any mutation in the program will lead to a whole collection of detailed changes in the behavior, so that natural selection has no opportunity to pick out changes which are beneficial from those which are not.

Fourth, if random mutations can only, say, increase or decrease a length, then even if one mutation goes in the wrong direction, it is easy for another mutation to recover by going in the opposite direction. But if there are in effect many possible directions, it becomes much more difficult to recover from missteps, and to exhibit any form of systematic convergence.

And finally, as the results in Chapter 7 suggest, for anything beyond the very simplest forms of behavior, iterative random searches rapidly tend to get stuck, and make at best excruciatingly slow progress towards any kind of global optimum.

In a sense it is not surprising that natural selection can achieve little when confronted with complex behavior. For in effect it is being asked to predict what changes would need to be made in an underlying program in order to produce or enhance a certain form of overall behavior. Yet one of the main conclusions of this book is that even given a particular program, it can be very difficult to see what the behavior of the program will be. And to go backwards from behavior to programs is a still much more difficult task.

In writing this book it would certainly have been convenient to have had a systematic way to be able to find examples of programs that exhibit specified forms of complex behavior. And indeed I have tried hard to develop iterative search procedures that would do this. But even using a whole range of tricks suggested by biology—as well as quite a number that are not—I have never been successful. And in fact in every single case I have in the end reverted either to exhaustive or to purely random searches, with no attempt at iterative improvement.

So what does this mean for biological organisms? It suggests that if a particular feature of an organism is successfully going to be optimized for different environments by natural selection, then this feature must somehow be quite simple.

And no doubt that is a large part of the reason that biological organisms always tend to consist of separate organs or other parts, each of which has at least some attributes that are fairly simple. For in this way there end up being components that are simple enough to be adjusted in a meaningful fashion by natural selection.

It has often been claimed that natural selection is what makes systems in biology able to exhibit so much more complexity than systems that we explicitly construct in engineering. But my strong suspicion is that in fact the main effect of natural selection is almost exactly the opposite: it tends to make biological systems avoid complexity, and be more like systems in engineering.

When one does engineering, one normally operates under the constraint that the systems one builds must behave in a way that is readily predictable and understandable. And in order to achieve this one typically

limits oneself to constructing systems out of fairly small numbers of components whose behavior and interactions are somehow simple.

But systems in nature need not in general operate under the constraint that their behavior should be predictable or understandable. And what this means is that in a sense they can use any number of components of any kind—with the result, as we have seen in this book, that the behavior they produce can often be highly complex.

However, if natural selection is to be successful at systematically molding the properties of a system then once again there are limitations on the kinds of components that the system can have. And indeed, it seems that what is needed are components that behave in simple and somewhat independent ways—much as in traditional engineering.

At some level it is not surprising that there should be an analogy between engineering and natural selection. For both cases can be viewed as trying to create systems that will achieve or optimize some goal.

Indeed, the main difference is just that in engineering explicit human effort is expended to find an appropriate form for the system, whereas in natural selection an iterative random search process is used instead. But the point is that the conditions under which these two approaches work turn out to be not so different.

In fact, there are even, I suspect, similarities in quite detailed issues such as the kinds of adjustments that can be made to individual components. In engineering it is common to work with components whose properties can somehow be varied smoothly, and which can therefore be analyzed using the methods of calculus and traditional continuous mathematics.

And as it turns out, much as we saw in Chapter 7, this same kind of smooth variation is also what tends to make iterative search methods such as natural selection be successful.

In biological systems based on discrete genetic programs, it is far from clear how smooth variation can emerge. Presumably in some cases it can be approximated by the presence of varying numbers of repeats in the underlying program. And more often it is probably the result of combinations of large numbers of elements that each produce fairly random behavior.

But the possibility of smooth variation seems to be important enough to the effectiveness of natural selection that it is extremely common in actual biological systems. And indeed, while there are some traits—such as eye color and blood type in humans—that are more or less discrete, the vast majority of traits seen, say, in the breeding of plants and animals, show quite smooth variation.

So to what extent does the actual history of biological evolution reflect the kinds of simple characteristics that I have argued one should expect from natural selection?

If one looks at species that exist today, and at the fossil record of past species, then one of the most striking features is just how much is in common across vast ranges of different organisms. The basic body plans for animals, for example, have been almost the same for hundreds of millions of years, and many organs and developmental pathways are probably even still older.

In fact, the vast majority of structurally important features seem to have changed only quite slowly and gradually in the course of evolution—just as one would expect from a process of natural selection that is based on smooth variations in fairly simple properties.

But despite this it is still clear that there is considerable diversity, at least at the level of visual appearance, in the actual forms of biological organisms that occur. So how then does such diversity arise?

One effect, to be discussed at greater length in the next section, is essentially just a matter of geometry. If the relative rates of growth of different parts of an organism change even slightly, then it turns out that this can sometimes have dramatic consequences for the overall shape of the organism, as well as for its mechanical operation.

And what this means is that just by making gradual changes in quantities such as relative rates of growth, natural selection can succeed in producing organisms that at least in some respects look very different.

But what about other differences between organisms? To what extent are all of them systematically determined by natural selection?

Following the discussion earlier in this section, it is my strong suspicion that at least many of the visually most striking differences—

associated for example with texture and pigmentation patterns—in the end have almost nothing to do with natural selection.

And instead what I believe is that such differences are in essence just reflections of completely random changes in underlying genetic programs, with no systematic effects from natural selection.

Particularly among closely related species of organisms there is certainly quite a contrast between the dramatic differences often seen in features such as pigmentation patterns and the amazing constancy of other features. And most likely those features in which a great degree of constancy is seen are precisely the ones that have successfully been molded by natural selection.

But as I mentioned earlier, it is almost always those features which change most rapidly between species that show the most obvious signs of complexity. And this observation fits precisely with the idea that complexity is easy to get by randomly sampling simple programs, but is hard for natural selection to handle in any kind of systematic way.

So in the end, therefore, what I conclude is that many of the most obvious features of complexity in biological organisms arise in a sense not because of natural selection, but rather in spite of it.

No doubt it will for many people be difficult to abandon the idea that natural selection is somehow crucial to the presence of complexity in biological organisms. For traditional intuition makes one think that to get the level of complexity that one sees in biological systems must require great effort—and the long and ponderous course of evolution revealed in the fossil record seems like just the kind of process that should be involved.

But the point is that what I have discovered in this book shows that in fact if one just chooses programs at random, then it is easy to get behavior of great complexity. And it is this that I believe lies at the heart of most of the complexity that we see in nature, both in biological and non-biological systems.

Whenever natural selection is an important determining factor, I suspect that one will inevitably see many of the same simplifying features as in systems created through engineering. And only when natural selection is not crucial, therefore, will biological systems be

able to exhibit the same level of complexity that one observes for example in many systems in physics.

In biology the presence of long programs with many separate parts can lead to a certain rather straightforward complexity analogous to having many physical objects of different kinds collected together. But the most dramatic examples of complexity in biology tend to occur in individual parts of systems—and often involve patterns or structures that look remarkably like those in physics.

Yet if biology samples underlying genetic programs essentially at random, why should these programs behave anything like programs that are derived from specific laws of physics?

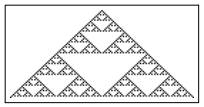
The answer, as we have seen many times in this book, is that across a very wide range of programs there is great universality in the behavior that occurs. The details depend on the exact rules for each program, but the overall characteristics remain very much the same.

And one of the important consequences of this is that it suggests that it might be possible to develop a rather general predictive theory of biology that would tell one, for example, what basic forms are and are not likely to occur in biological systems.

One might have thought that the traditional idea that organisms are selected to be optimal for their environment would already long ago have led to some kind of predictive theory. And indeed it has for example allowed some simple numerical ratios associated with populations of organisms to be successfully derived. But about a question such as what forms of organisms are likely to occur it has much less to say.

There are a number of situations where fairly complicated structures appear to have arisen independently in several very different types of organisms. And it is sometimes claimed that this kind of convergent evolution occurs because these structures are in some ultimate sense optimal, making it inevitable that they will eventually be produced.

But I would be very surprised if this explanation were correct. And instead what I strongly suspect is that the reason certain structures appear repeatedly is just that they are somehow common among programs of certain kinds—just as, for example, we have seen that the



An example of a basic pattern that is produced in several variants by a wide range of simple programs.

intricate nested pattern shown on the left arises from many different simple programs.

Ever since the original development of the theory of evolution, there has been a widespread belief that the general trend seen in the fossil record towards the formation of progressively more complicated types of organisms must somehow be related to an overall increase in optimality.

Needless to say, we do not know what a truly optimal organism would be like. But if optimality is associated with having as many offspring as possible, then very simple organisms such as viruses and protozoa already seem to do very well.

So why then do higher organisms exist at all? My guess is that it has almost nothing to do with optimality, and that instead it is essentially just a consequence of strings of random mutations that happened to add more and more features without introducing fatal flaws.

It is certainly not the case—as is often assumed—that natural selection somehow inevitably leads to organisms with progressively more elaborate structures and progressively larger numbers of parts.

For a start, some kinds of organisms have been subject to natural selection for more than a billion years, but have never ended up becoming much more complicated. And although there are situations where organisms do end up becoming more complicated, they also often become simpler.

A typical pattern—remarkably similar, as it happens, to what occurs in the history of technology—is that at some point in the fossil record some major new capability or feature is suddenly seen. At first there is then rapid expansion, with many new species trying out all sorts of possibilities that have been opened up. And usually some of these possibilities get quite ornate and elaborate. But after a while it becomes clear what makes sense and what does not. And typically things then get simpler again.

So what is the role of natural selection in all of this? My guess is that as in other situations, its main systematic contribution is to make things simpler, and that insofar as things do end up getting more complicated, this is almost always the result of essentially random

sampling of underlying programs—without any systematic effect of natural selection.

For the more superficial aspects of organisms—such as pigmentation patterns—it seems likely that among programs sampled at random a fair fraction will produce results that are not disastrous for the organism. But when one is dealing with the basic structure of organisms, the vast majority of programs sampled at random will no doubt have immediate disastrous consequences. And in a sense it is natural selection that is responsible for the fact that such programs do not survive.

But the point is that in such a case its effect is not systematic or cumulative. And indeed it is my strong suspicion that for essentially all purposes the only reasonable model for important new features of organisms is that they come from programs selected purely at random.

So does this then mean that there can never be any kind of general theory for all the features of higher organisms? Presumably the pattern of exactly which new features were added when in the history of biological evolution is no more amenable to general theory than the specific course of events in human history. But I strongly suspect that the vast majority of significant new features that appear in organisms are at least at first associated with fairly short underlying programs. And insofar as this is the case the results of this book should allow one to develop some fairly general characterizations of what can happen.

So what all this means is that much of what we see in biology should correspond quite closely to the typical behavior of simple programs as we have studied them in this book—with the main caveat being just that certain aspects will be smoothed and simplified by the effects of natural selection. Seeing in earlier chapters of this book all the diverse things that simple programs can do, it is easy to be struck by analogies to books of biological flora and fauna. Yet what we now see is that in fact such analogies may be quite direct—and that many of the most obvious features of actual biological organisms may in effect be direct reflections of typical behavior that one sees in simple programs.

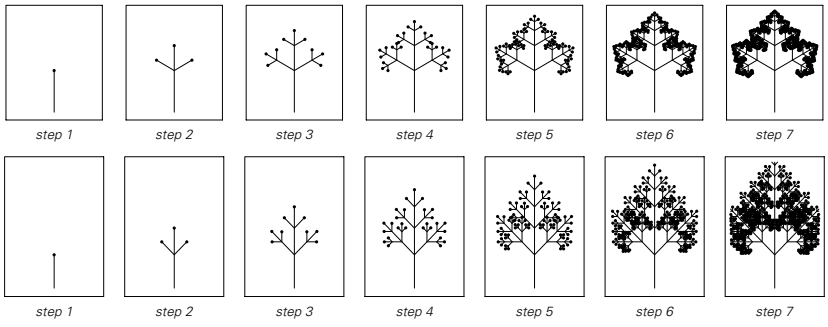
Growth of Plants and Animals

Looking at all the elaborate forms of plants and animals one might at first assume that the underlying rules for their growth must be highly complex. But in this book we have discovered that even by following very simple rules it is possible to obtain forms of great complexity. And what I have come to believe is that in fact most aspects of the growth of plants and animals are in the end governed by remarkably simple rules.

As a first example of biological growth, consider the stem of a plant. It is usually only at the tip of a stem that growth can occur, and much of the time all that ever happens is that the stem just gets progressively longer. But the crucial phenomenon that ultimately leads to much of the structure we see in many kinds of plants is that at the tip of a stem it is possible for new stems to form and branch off. And in the simplest cases these new stems are in essence just smaller copies of the original stem, with the same basic rules for growth and branching.

With this setup the succession of branchings can then be represented by steps in the evolution of a neighbor-independent substitution system in which the tip of each stem is at each step replaced by a collection of smaller stems in some fixed configuration.

Two examples of such substitution systems are shown in the pictures below. In both cases the rules are set up so that every stem in effect just branches into exactly three new stems at each step. And this



Steps in the evolution of substitution systems that provide simple models for the growth of plants. At each step every growing stem is replaced by a collection of three new stems according to the rules shown. For individual stems this type of branching is known in botany as monopodial.

means that the network of connections between stems necessarily has a very simple nested form. But if one looks at the actual geometrical arrangement of stems there is no longer such simplicity; indeed, despite the great simplicity of the underlying rules, considerable complexity is immediately evident even in the pictures at the bottom of the facing page.

The pictures on the next page show patterns obtained with various sequences of choices for the lengths and angles of new stems. In a few cases the patterns are quite simple; but in most cases they turn out to be highly complex—and remarkably diverse.

The pictures immediately remind one of the overall branching patterns of all sorts of plants—from algae to ferns to trees to many kinds of flowering plants. And no doubt it is from such simple rules of growth that most such overall branching patterns come.

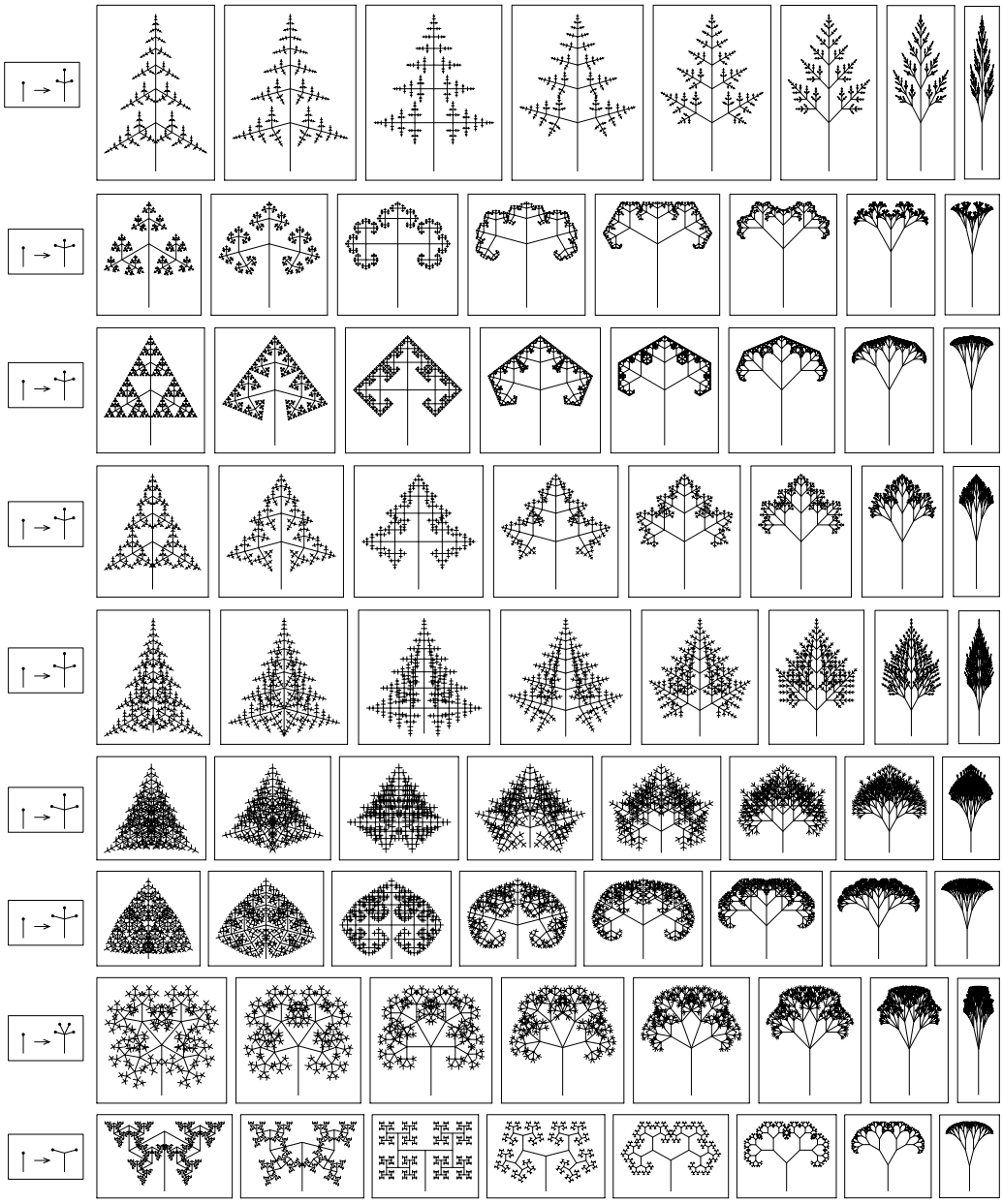
But what about more detailed features of plants? Can they also be thought of as consequences of simple underlying rules of growth?

For many years I wondered in particular about the shapes of leaves. For among different plants there is tremendous diversity in such shapes—as illustrated in the pictures on page 403. Some plants have leaves with simple smooth boundaries that one might imagine could be described by traditional mathematical functions. Others have leaves with various configurations of sharp points. And still others have leaves with complex and seemingly somewhat random boundaries.

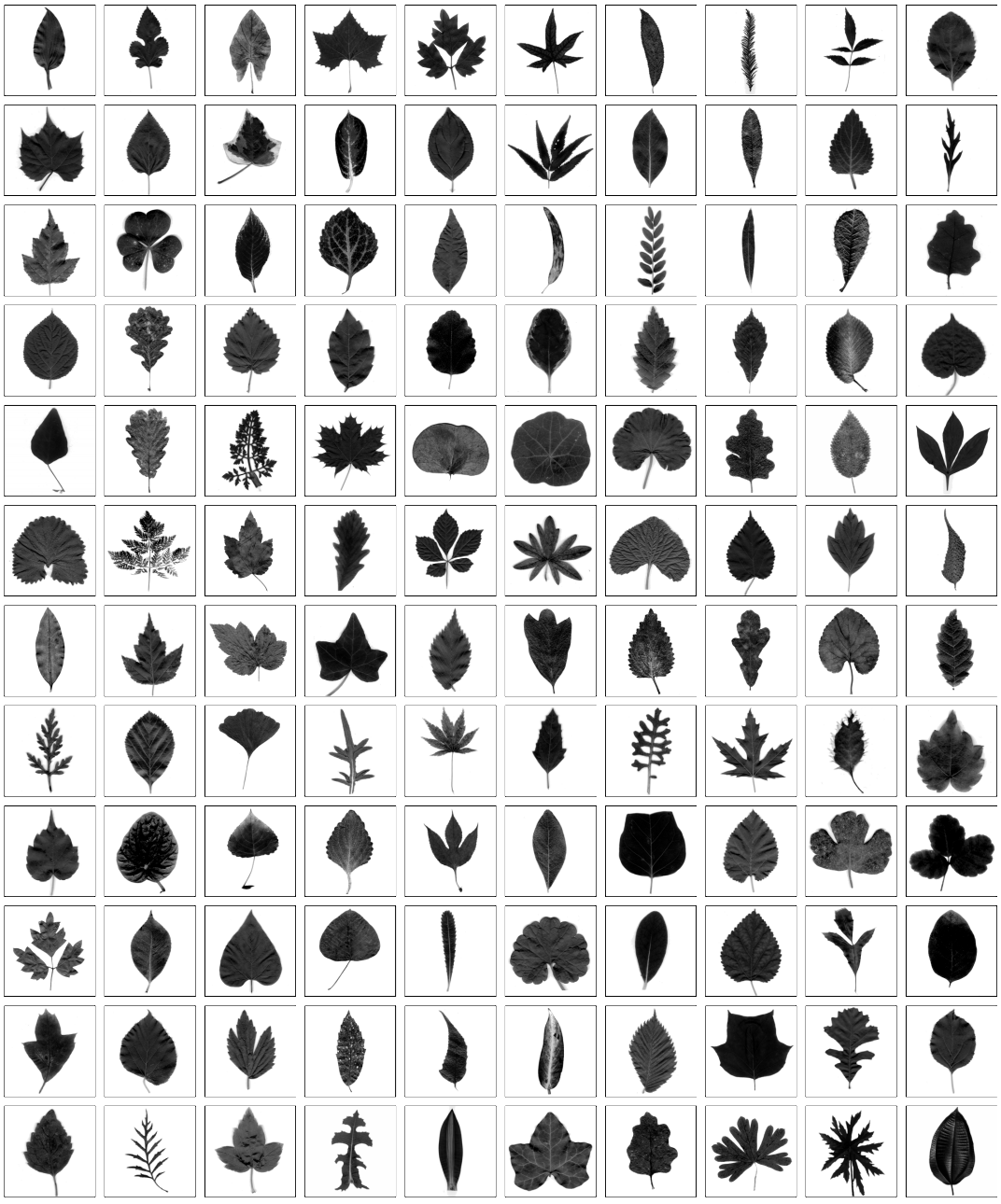
So given this diversity one might at first suppose that no single kind of underlying rule could be responsible for what is seen. But looking at arrays of pictures like the ones on the next page one makes a remarkable discovery: among the patterns that can be generated by simple substitution systems are ones whose outlines look extremely similar to those of a wide variety of types of leaves.

There are patterns with smooth edges that look like lily pads. There are patterns with sharp points that look like prickly leaves of various kinds. And there are patterns with intricate and seemingly somewhat random shapes that look like sycamore or grape leaves.

It has never in the past been at all clear how leaves get the shapes they do. Presumably most of the processes that are important take place while leaves are still folded up inside buds, and are not yet very solid.



Limiting patterns produced by substitution systems of the type shown in the previous picture. The patterns on each row are obtained from rules that are set up to give branches with particular relative lengths. The angles between the branches are taken to increase by 15° in successive pictures across the row. Note that pictures shown on different rows are scaled differently—so that the initial vertical stem does not always appear with the same height. The similarity between pictures on this page and overall branching patterns and shapes of leaves in many kinds of plants is striking.



Examples of different kinds of leaves, mostly from common flowering plants. The diversity of shapes is remarkable, as is the similarity to the forms shown on the facing page. The leaves range in size from under an inch to many feet.

For although leaves typically expand significantly after they come out, the basic features of their shapes almost never seem to change.

There is some evidence that at least some aspects of the pattern of veins in a leaf are laid down before the main surface of the leaf is filled in, and perhaps the stems in the branching process I describe here correspond to precursors of structures related to veins. Indeed, the criss-crossing of veins in the leaves of higher plants may be not unrelated to the fact that stems in the pictures two pages ago often cross over—although certainly many of the veins in actual full-grown leaves are probably added long after the shapes of the leaves are determined.

One might at the outset have thought that leaves would get their shapes through some mechanism quite unrelated to other aspects of plant growth. But I strongly suspect that in fact the very same simple process of branching is ultimately responsible both for the overall forms of plants, and for the shapes of their leaves.

Quite possibly there will sometimes be at least some correspondence between the lengths and angles that appear in the rules for overall growth and for the growth of leaves. But in general the details of all these rules will no doubt depend on very specific characteristics of individual plants.

The distance before a new stem appears is, for example, probably determined by the rates of production and diffusion of plant hormones and related substances, and these rates will inevitably depend both on the thickness and mechanical structure of the stem, as well as on all kinds of biochemical properties of the plant. And when it comes to the angles between old and new stems I would not be surprised if these were governed by such microscopic details as individual shapes of cells and individual sequences of cell divisions.

The traditional intuition of biology would suggest that whenever one sees complexity—say in the shape of a leaf—it must have been generated for some particular purpose by some sophisticated process of natural selection. But what the pictures on the previous pages demonstrate is that in fact a high degree of complexity can arise in a sense quite effortlessly just as a consequence of following certain simple rules of growth.

No doubt some of the underlying properties of plants are indeed guided by natural selection. But what I strongly suspect is that in the

vast majority of cases the occurrence of complexity—say in the shapes of leaves—is in essence just a side effect of the particular rules of growth that happen to result from the underlying properties of the plant.

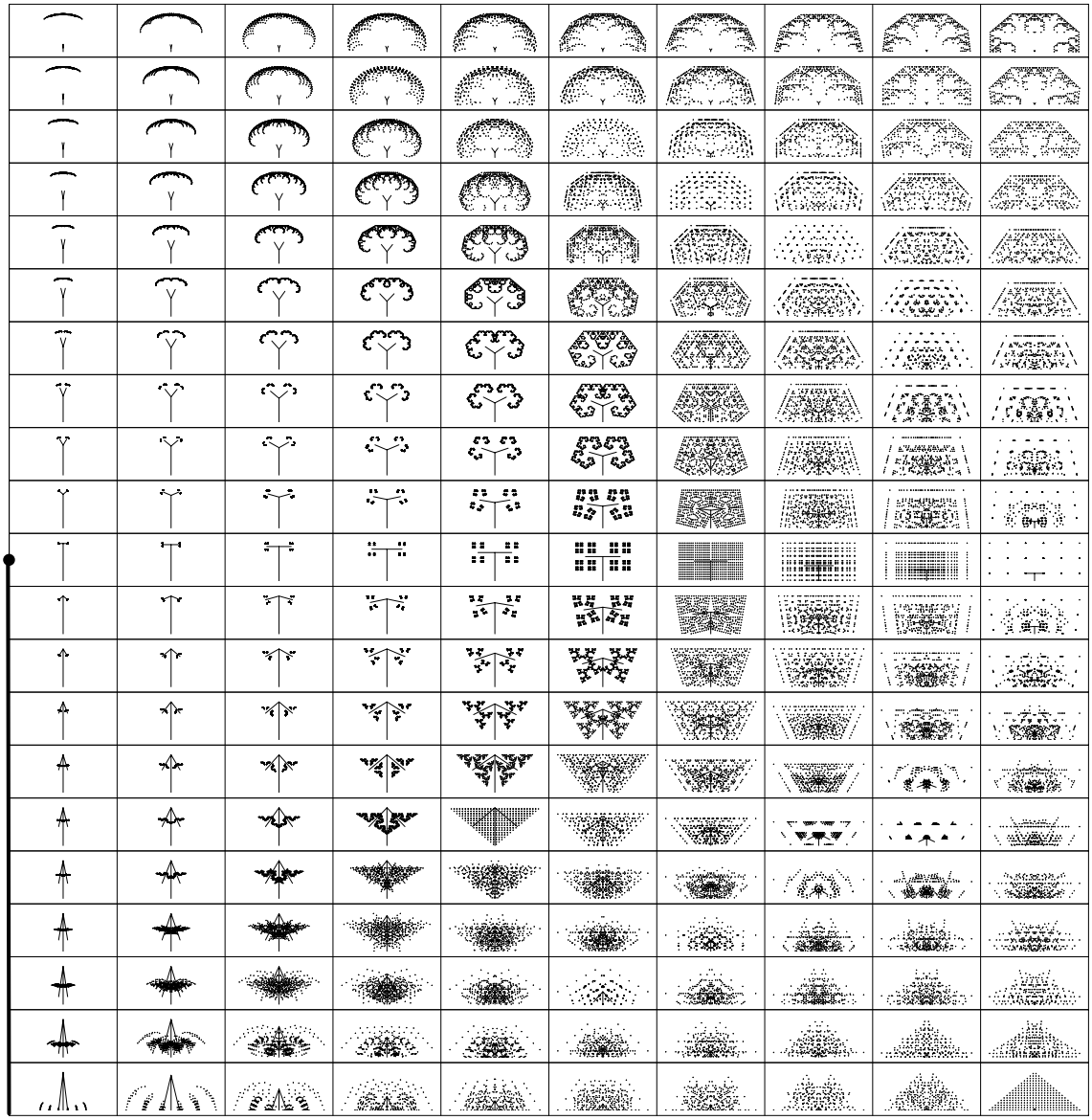
The pictures on the next page show the array of possible forms that can be produced by rules in which each stem splits into exactly two new stems at each step. The vertical black line on the left-hand side of the page represents in effect the original stem at each step, and the pictures are arranged so that the one which appears at a given position on the page shows the pattern that is generated when the tip of the right-hand new stem goes to that position relative to the original stem shown on the left.

In some cases the patterns obtained are fairly simple. But even in these cases the pictures show that comparatively small changes in underlying rules can lead to much more complex patterns. And so if in the course of biological evolution gradual changes occur in the rules, it is almost inevitable that complex patterns will sometimes be seen.

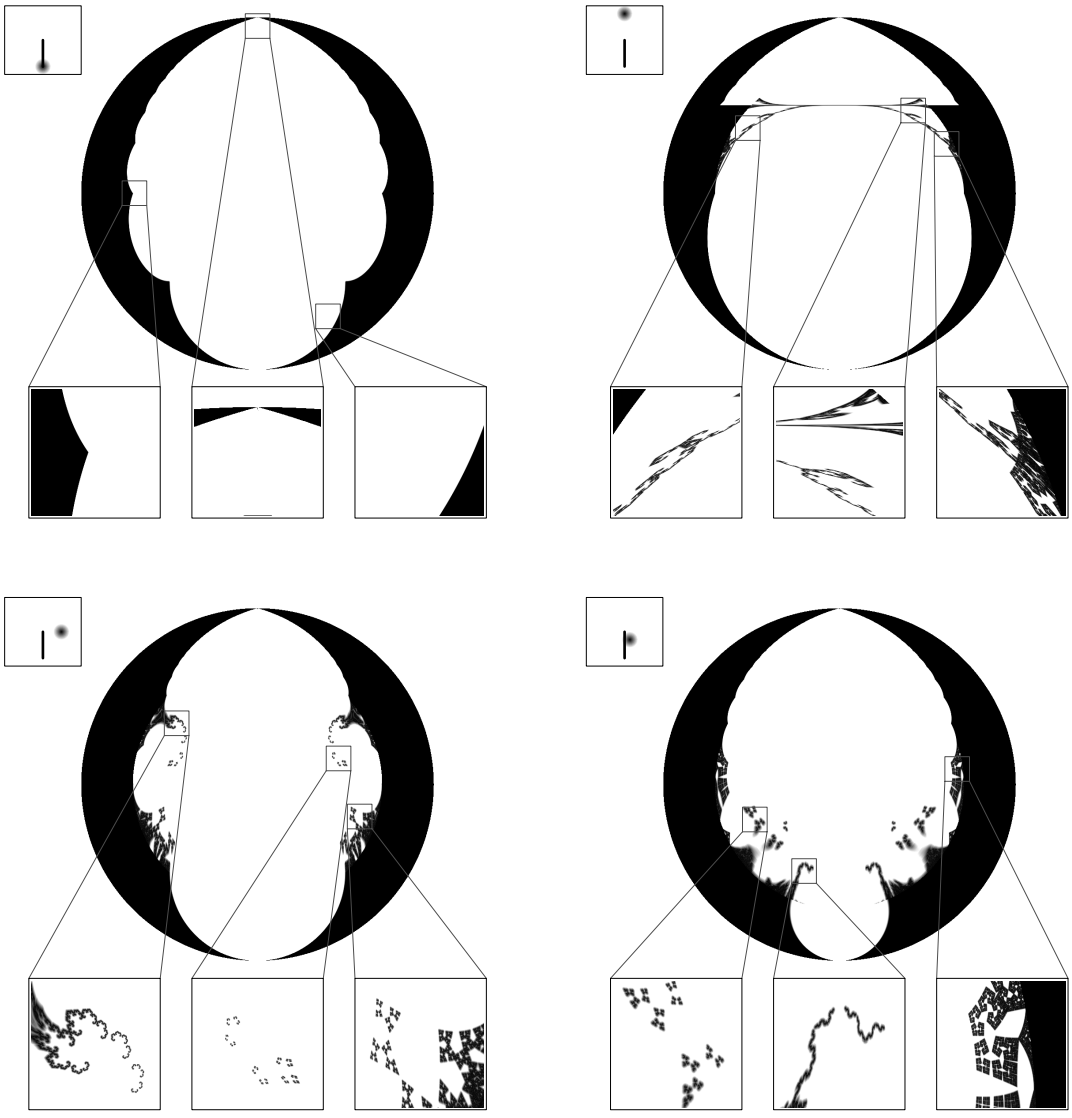
But just how suddenly can the patterns change? To get some idea of this one can construct a kind of limit of the array on the next page in which the total number of pictures is in effect infinite, but only a specific infinitesimal region of each picture is shown. Page 407 gives results for four choices of the position of this region relative to the original stem. And instead of just displaying black or white depending on whether any part of the pattern lies in the region, the picture uses gray levels to indicate how close it comes.

The areas of solid black thus correspond to ranges of parameters in the underlying rule for which the patterns obtained always reach a particular position. But what we see is that at the edges of these areas there are often intricate structures with an essentially nested form. And the presence of such structures implies that at least with some ranges of parameters, even very small changes in underlying rules can lead to large changes in certain aspects of the patterns that are produced.

So what this suggests is that it is almost inevitable that features such as the shapes of leaves can sometimes change greatly even when the underlying properties of plants change only slightly. And I suspect



The full array of patterns that can be produced by simple substitution systems in which each stem branches into exactly two symmetrical stems at each step. The patterns are arranged on the page so that the pattern shown at a particular position corresponds to what is obtained with a rule in which the tip of the right-hand stem goes to that position (corrected for the aspect ratio of the array) relative to the original stem shown as a vertical line on the left-hand side of the page. In each case the result of 10 steps of evolution is shown, and the pictures are scaled so that all points above the bottom of the original stem can be included. Note that for rules outside of a distorted semicircle centered on the dot at the left-hand side of the page, and touching the three other sides of the page, the patterns generated grow at each step, rather than tending to a limit of fixed size.



Maps of where in the space of parameters for the substitution systems on the facing page the patterns obtained overlap the region indicated in the icon at the top left of each picture. Black corresponds to complete overlap, while white corresponds to no overlap. The maps shown can be thought of as being made by taking an infinitely dense limit of the array of pictures on the facing page, but keeping only what one sees in each picture by looking through a peephole at a particular position relative to the original stem.

that this is precisely why such diverse shapes of leaves are occasionally seen even in plants that otherwise appear very similar.

But while features such as the shapes of leaves typically differ greatly between different plants, there are also some seemingly quite sophisticated aspects of plants that typically remain almost exactly the same across a huge range of species.

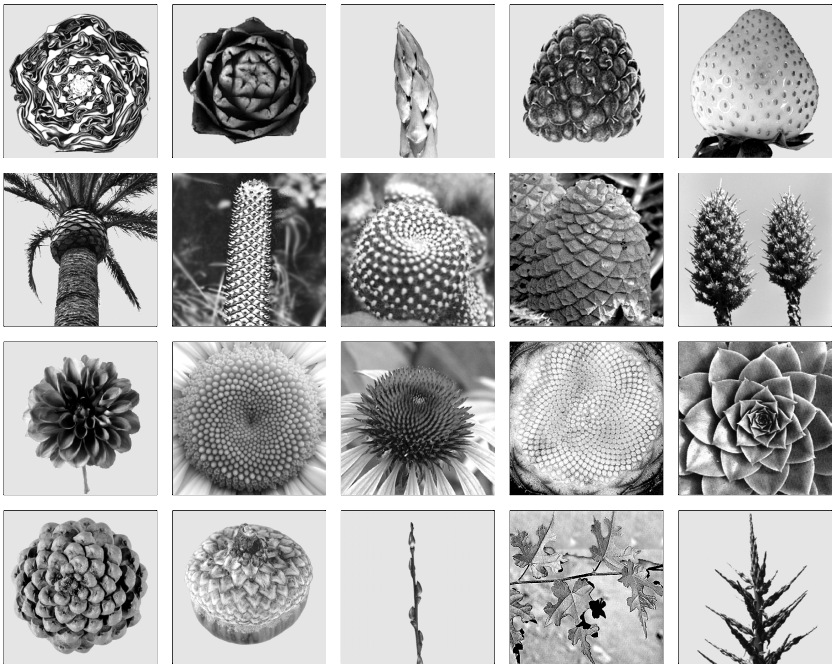
One example is the arrangement of sequences of plant organs or other elements around a stem. In some cases successive leaves, say, will always come out on opposite sides of a stem—180° apart. But considerably more common is for leaves to come out less than 180° apart, and in most plants the angle turns out to be essentially the same, and equal to almost exactly 137.5°.

It is already remarkable that such a definite angle arises in the arrangement of leaves—or so-called phyllotaxis—of so many plants. But it turns out that this very same angle also shows up in all sorts of other features of plants, as shown in the pictures at the top of the facing page. And although the geometry is different in different cases, the presence of a fixed angle close to 137.5° always leads to remarkably regular spiral patterns.

Over the years, much has been written about such patterns, and about their mathematical properties. For it turns out that an angle between successive elements of about 137.5° is equivalent to a rotation by a number of turns equal to the so-called golden ratio $(1 + \sqrt{5})/2 \simeq 1.618$ which arises in a wide variety of mathematical contexts—notably as the limiting ratio of Fibonacci numbers.

And no doubt in large part because of this elegant mathematical connection, it has usually come to be assumed that the 137.5° angle and the spiral patterns to which it leads must correspond to some kind of sophisticated optimization found by an elaborate process of natural selection.

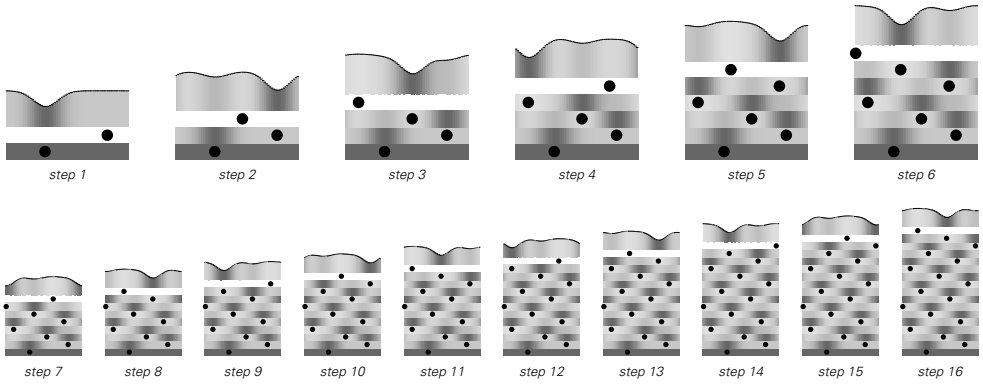
But I do not believe that this is in fact the case. And instead what I strongly suspect is that the patterns are just inevitable consequences of a rather simple process of growth not unlike one that was already discussed, at least in general terms, nearly a century ago.



Examples of spiral arrangements of elements in various plant systems. The details of the final geometry are different in different cases. But in all cases it turns out that the original angle between successive elements is almost exactly 137.5°. The first row shows red cabbage (cut open), artichoke, asparagus, raspberry and strawberry. The first two objects on the last row are a pinecone and an acorn.

The positions of new plant organs or other elements around a stem are presumably determined by what happens in a small ring of material near the tip of the growing stem. And what I suspect is that a new element will typically form at a particular position around the ring if at that position the concentration of some chemical has reached a certain critical level.

But as soon as an element is formed, one can expect that it will deplete the concentration of the chemical in its local neighborhood, and thus inhibit further elements from forming nearby. Nevertheless, general processes in the growing stem will presumably make the concentration steadily rise throughout the ring of active material, and eventually this concentration will again get high enough at some position that it will cause another element to be formed.

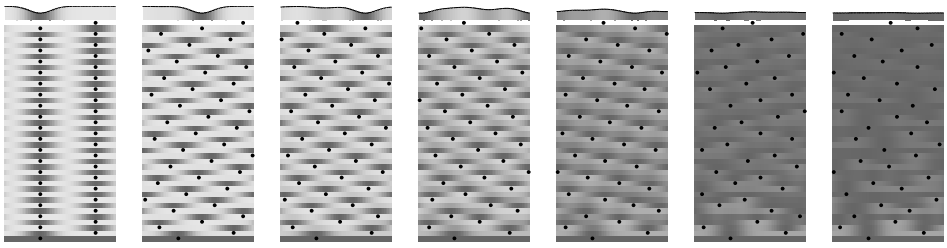


A simple model for the arrangement of leaves or other elements produced at the growing tip of a plant stem. The stem is taken to grow up the page, and for purposes of display it is unrolled into a line. The positions of leaves or other elements are indicated by black dots. The concentration of a chemical is indicated by gray level, and for the top line at each step, it is also plotted. The rule for the system places a new black dot at whatever position this concentration is largest. The black dot is then assumed to deplete the concentration around it, but the overall concentration is uniformly increased before the next step. It turns out that successive black dots rapidly become spaced at almost exactly 137.5°.

The pictures above show an example of this type of process. For purposes of display the ring of active material is unrolled into a line, and successive states of this line are shown one on top of each other going up the page. At each step a new element, indicated by a black dot, is taken to be generated at whatever position the concentration is maximal. And around this position the new element is then taken to produce a dip in concentration that is gradually washed out over the course of several steps.

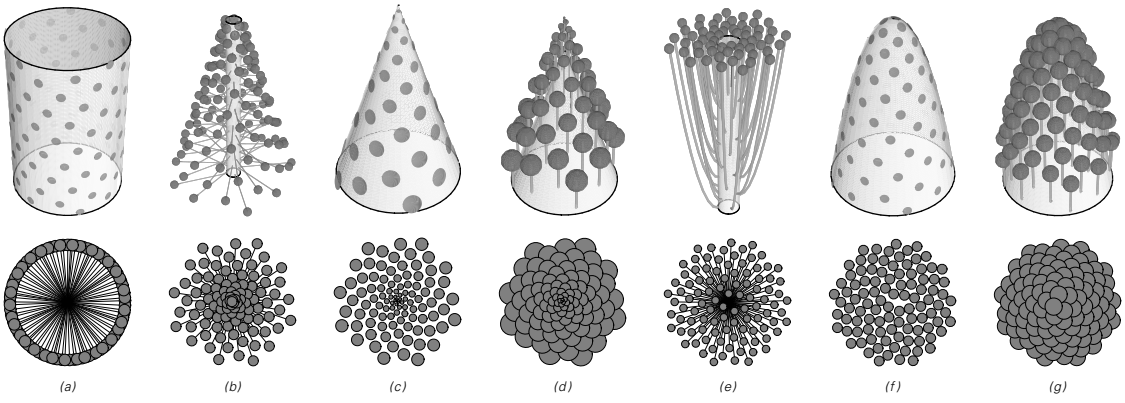
The way the pictures are drawn, the angles between successive elements correspond to the horizontal distances between them. And although these distances vary somewhat for the first few steps, what we see in general is remarkably rapid convergence to a fixed distance—which turns out to correspond to an angle of almost exactly 137.5°.

So what happens if one changes the details of the model? In the extreme case where all memory of previous behavior is immediately damped out the first picture at the top of the facing page shows that successive elements form at 180° angles. And in the case where there is very little damping the last two pictures show that at least for a while elements can form at fairly random angles. But in the majority of cases one sees rather rapid convergence to almost precisely 137.5°.

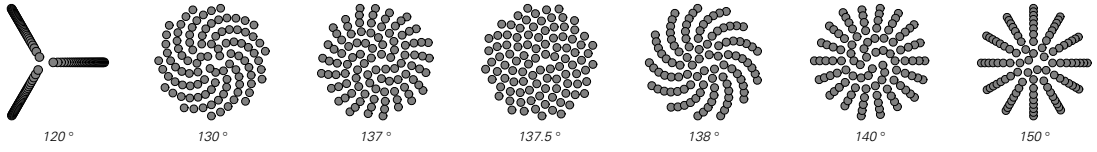


Examples of changing the amount of damping used in the model on the facing page. 100% damping corresponds to increasing the overall concentration at each step so much that no memory of previous steps remains. 0% corresponds to no increase in overall concentration at each step. Away from these extreme cases, rapid convergence is seen to a spacing between black dots of almost exactly 137.5°.

So just how does this angle show up in actual plant systems? As the top pictures below demonstrate, the details depend on the geometry and relative growth rates of new elements and of the original stem. But in all cases very characteristic patterns are produced.



Examples of structures formed in various geometries by successively adding elements at a golden ratio angle 137.5°. Each of these structures is seen in one type of plant growth or another, as illustrated on page 409.



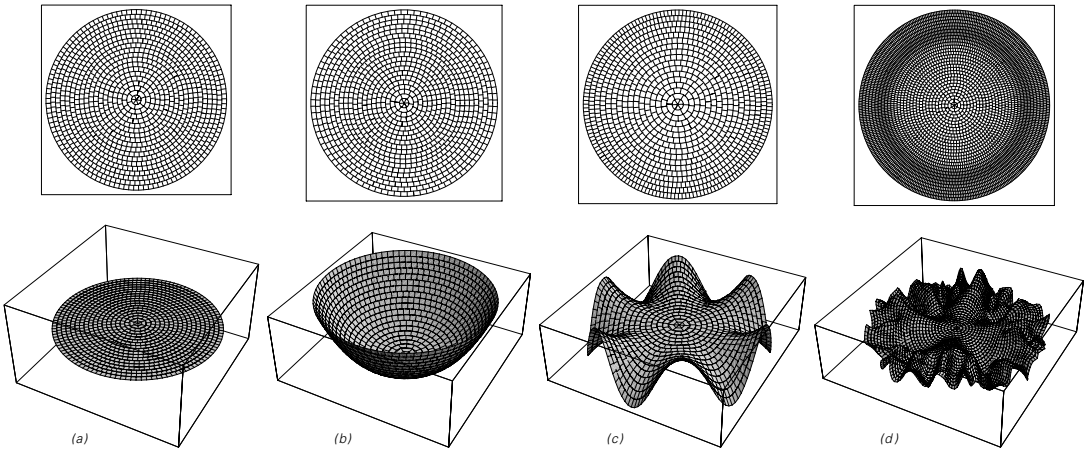
Overall patterns formed by successively adding elements at a variety of different angles. In each case the n^{th} element appears at coordinates $\sqrt{n}\ \{Cos[n\,\theta],\ Sin[n\,\theta]\}$. Stripes are seen if θ/π (with θ in radians) is easy to approximate by a rational number. (The size of the region before stripes appear depends on *Length[ContinuedFraction[θ/π]]*.)

And as the bottom pictures on the previous page demonstrate, the forms of these patterns are very sensitive to the precise angle of successive elements: indeed, even a small deviation leads to patterns that are visually quite different. At first one might have assumed that to get a precise angle like 137.5° would require some kind of elaborate and highly detailed process. But just as in so many other situations that we have seen in this book, what we have seen is that in fact a very simple rule is all that is in the end needed.

One of the general features of plants is that most of their cells tend to develop fairly rigid cellulose walls which make it essentially impossible for new material to be added inside the volume of the plant, and so typically force new growth to occur only on the outside of the plant—most importantly at the tips of stems.

But when plants form sheets of material as in leaves or petals there is usually some flexibility for growth to occur within the sheet. And the pictures below show examples of what can happen if one starts with a flat disk and then adds different amounts of material in different places.

If more material is added near the center than near the edge, as in case (b), then the disk is forced to take on a cup shape similar to many



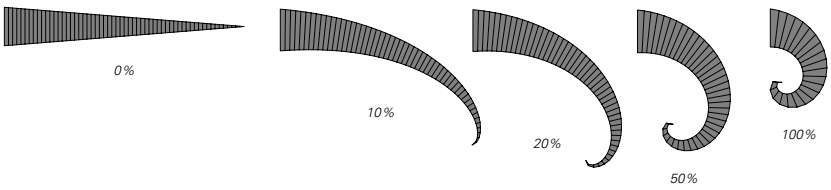
Disks with varying amounts of material at different distances from their centers. In the top row the disks are always flat, forcing the cells of material to vary in size and shape. In the bottom row, the disks form shapes in three dimensions in which all cells are the same size and shape. Relative to case (a), the amount of material going out from the center decreases linearly in case (b), increases linearly in case (c), and increases exponentially in case (d).

flowers. But if more material is added near the edge than near the center, as in case (c), then the sheet will become wavy at the edge, much like some leaves. And if the amount of material increases sufficiently rapidly from the center to the edge, as in case (d), then the disk will be forced to become highly corrugated, somewhat like a lettuce leaf.

So what about animals? To what extent are their mechanisms of growth the same as plants? If one looks at air passages or small blood vessels in higher animals then the patterns of branching one sees look similar to those in plants. But in most of their obvious structural features animals do not typically look much like plants at all. And in fact their mechanisms of growth mostly turn out to be rather different.

As a first example, consider a horn. One might have thought that, like a stem in a plant, a horn would grow by adding material at its tip. But in fact, like nails and hair, a horn instead grows by adding material at its base. And an immediate consequence of this is that the kind of branching that one sees in plants does not normally occur in horns.

But on the other hand coiling is common. For in order to get a structure that is perfectly straight, the rate at which material is added must be exactly the same on each side of the base. And if there is any difference, one edge of the structure that is produced will always end up being longer than the other, so that coiling will inevitably result, as in the pictures below.

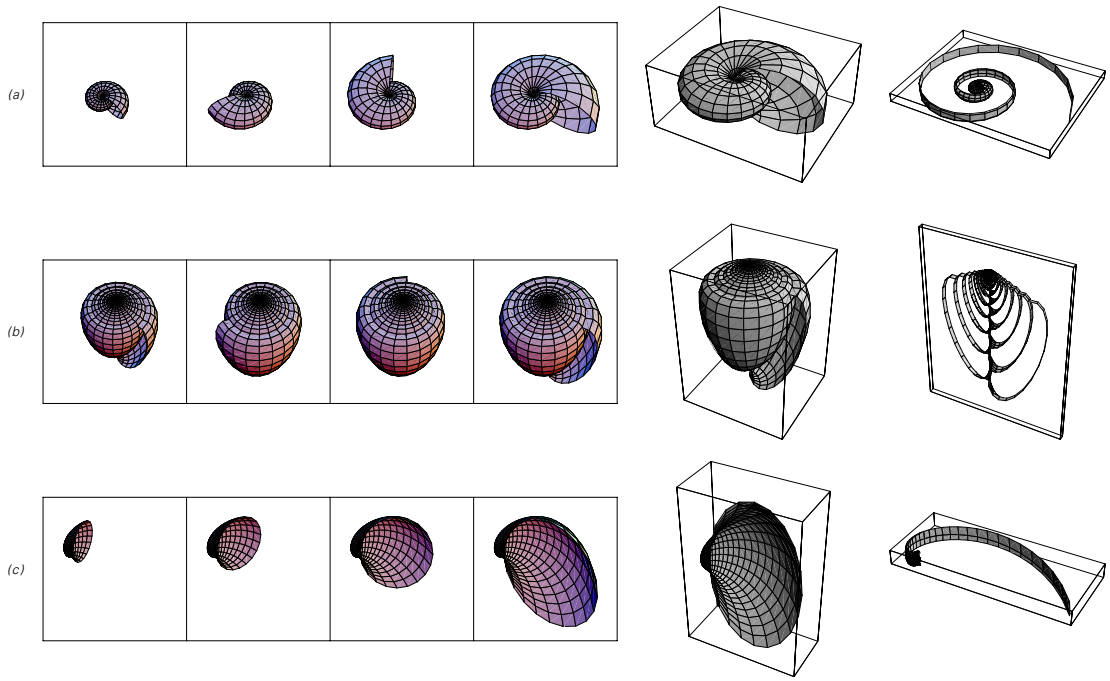


Idealized horns generated by progressively adding new material, with the amount of material on the upper edge of the base always being the specified percentage larger than the amount on the lower edge. These pictures can be viewed as one-dimensional analogs of those on the facing page.

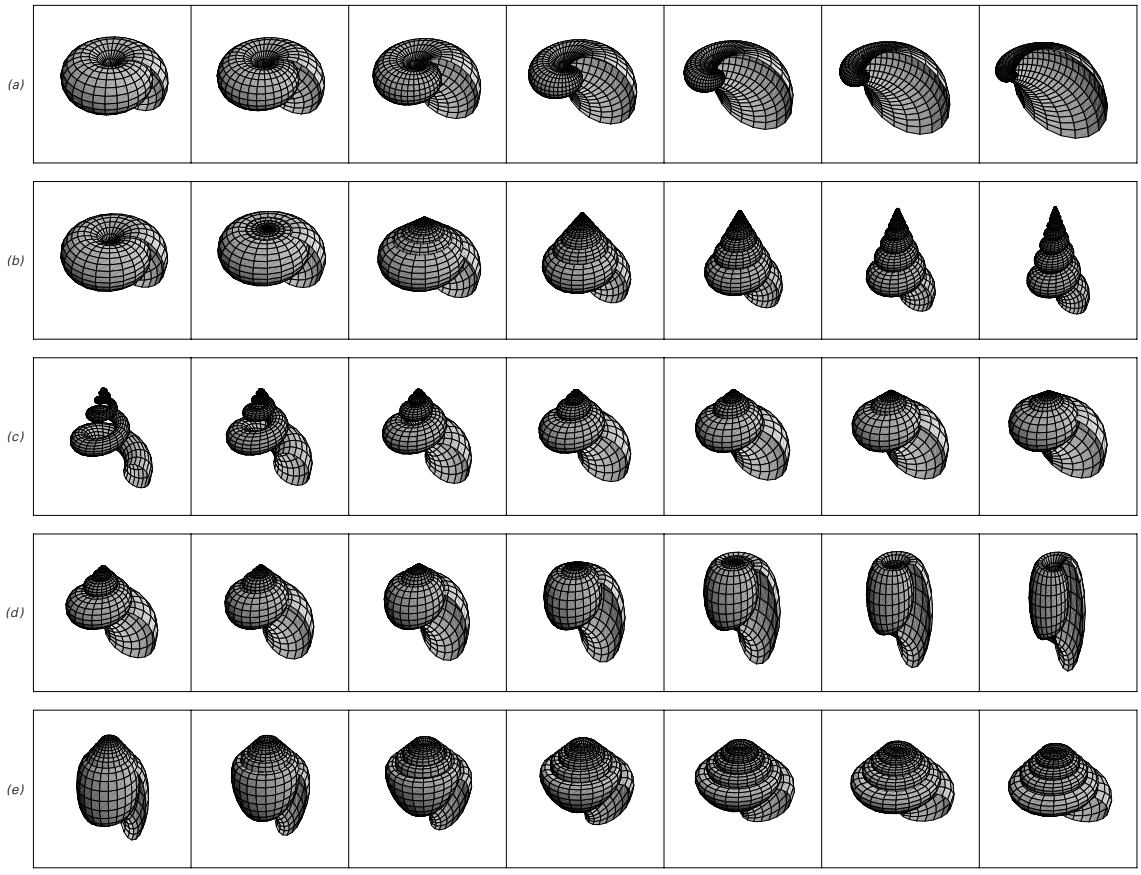
And as has been thought for several centuries, it turns out that a three-dimensional version of this phenomenon is essentially what leads to the elaborate coiled structures that one sees in mollusc shells. For in a typical case, the animal which lives at the open end of the shell

secretes new shell material faster on one side than the other, causing the shell to grow in a spiral. The rates at which shell material is secreted at different points around the opening are presumably determined by details of the anatomy of the animal. And it turns out that—much as we saw in the case of branching structures earlier in this section—even fairly small changes in such rates can have quite dramatic effects on the overall shape of the shell.

The pictures below show three examples of what can happen, while the facing page shows the effects of systematically varying certain growth rates. And what one sees is that even though the same very simple underlying model is used, there are all sorts of visually very different geometrical forms that can nevertheless be produced.



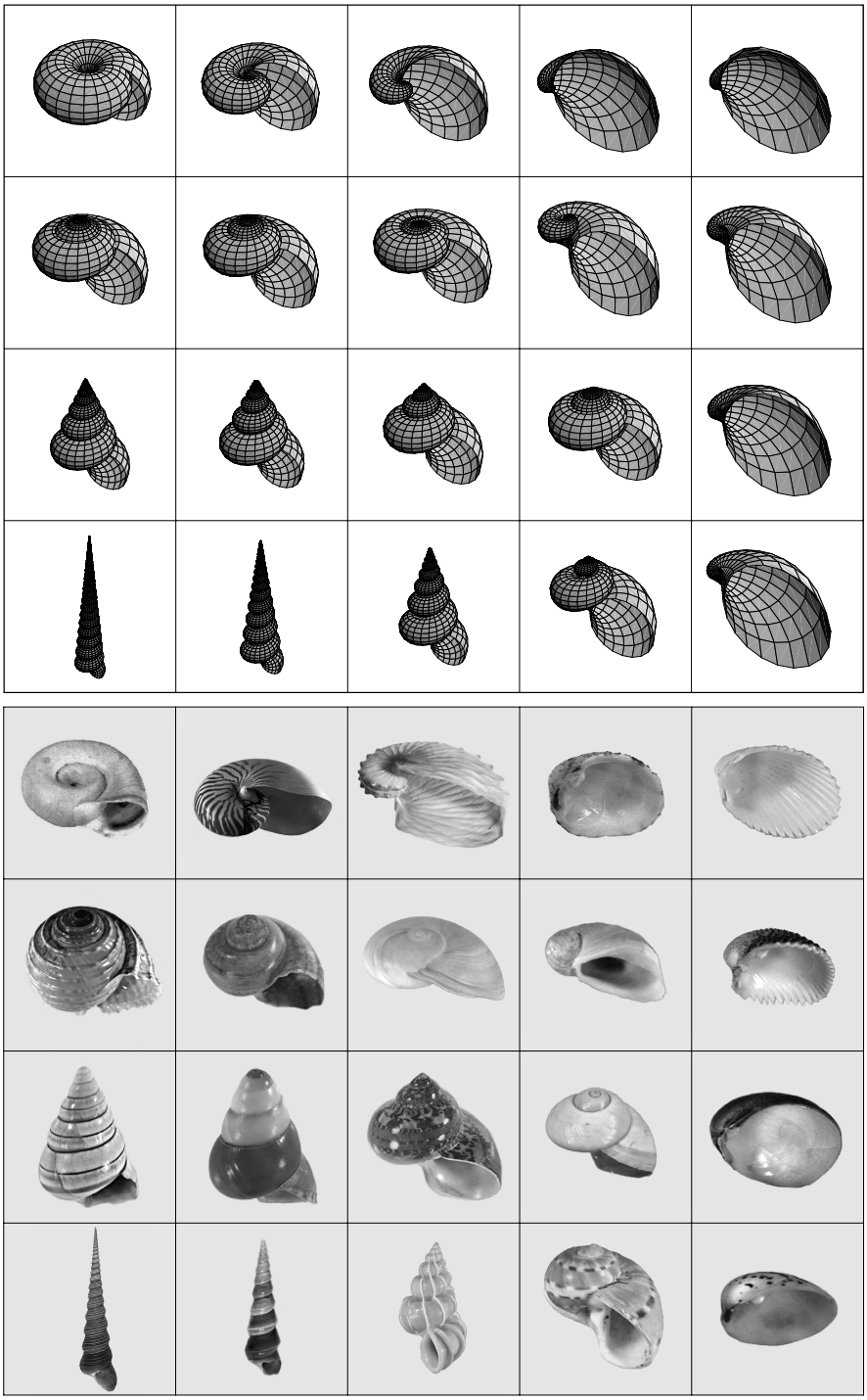
A simple model for the growth of mollusc shells. In each case new shell material is progressively added at the open end of the shell. The rule on the left shows the amount of material added at each stage at different points around the opening; the line from the center indicates the progressive lateral displacement of the opening. Case (a) is typical of a nautilus shell, (b) of a cone shell and (c) of one-half of a clam shell. All shells produced by adding material according to fixed rules of the kind shown here have the property that throughout their growth they maintain the same overall shape.



The effects of varying five simple features of the rule for the growth of a mollusc shell: (a) the overall factor by which the size increases in the course of each revolution; (b) the relative amount by which the opening is displaced downward at each revolution; (c) the size of the opening relative to the overall size of the shell; (d) the elongation of the opening; (e) the orientation of elongation in the opening. The pictures at the beginning and end of each row correspond roughly to the following: (a) pond snail shell, cockle shell; (b) pond snail shell, horn shell; (c) worm shell, bonnet shell; (d) periwinkle shell, cowrie shell; (e) olive shell, sundial shell.

So out of all the possible forms, which ones actually occur in real molluscs? The remarkable fact illustrated on the next page is that essentially all of them are found in some kind of mollusc or another.

If one just saw a single mollusc shell, one might well think that its elaborate form must have been carefully crafted by some long process of natural selection. But what we now see is that in fact all the different forms that are observed are in effect just consequences of the



Shell shapes generated by the simple model and found in nature. The array shows systematic variation of the first two parameters from the previous page. Similar arrays could be made for the other parameters.

application of three-dimensional geometry to very simple underlying rules of growth. And so once again therefore natural selection cannot reasonably be considered the source of the elaborate forms we see.

Away from mollusc shells, coiled structures—like branched ones—are not especially common in animals. Indeed, the vast majority of animals do not tend to have overall forms that are dominated by any single kind of structure. Rather, they are usually made up of a collection of separate identifiable parts, like heads, tails, legs, eyes and so on, all with their own specific structure.

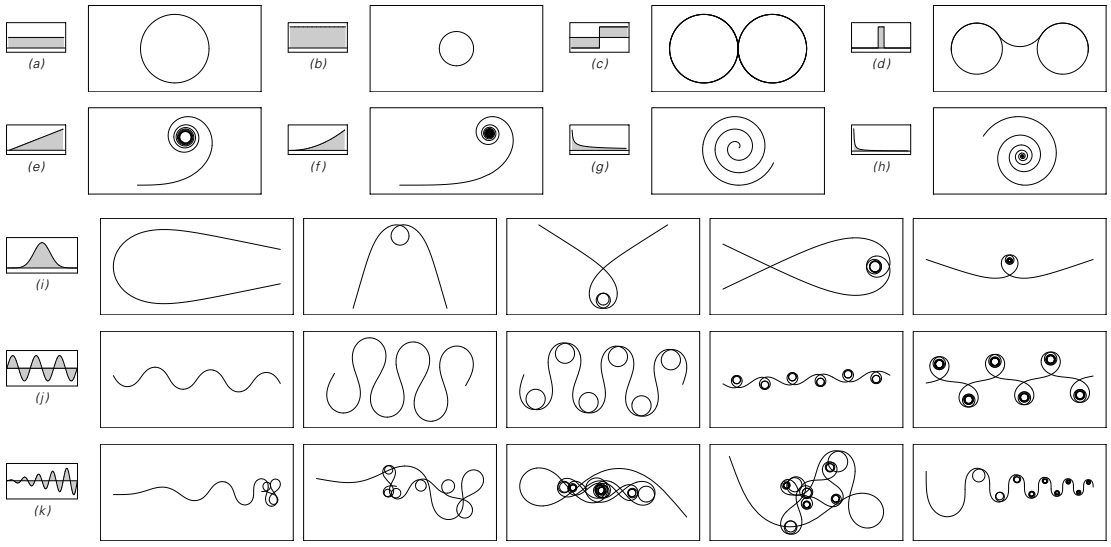
Sometimes some of these parts are repeated, perhaps in a sequence of segments, or perhaps in some kind of two-dimensional array. And very often the whole animal is covered by a fairly uniform outer skin. But the presence of many different kinds of parts is in the end one of the most obvious features of many animals.

So how do all these parts get produced? The basic mechanism seems to be that at different places and different times inside a developing animal different sections of its genetic program end up getting used—causing different kinds of growth to occur, and different structures to be produced. And part of what makes this possible is that particularly at the stage of the embryo most cells in an animal are not extremely rigid—so that even when different pieces of the animal grow quite differently they can still deform so as to fit together.

Usually there are some elements—such as bones—that eventually do become rigid. But the crucial point is that at the stage when the basic form of an animal is determined most of these elements are not yet rigid. And this allows various processes to occur that would otherwise be impossible.

Probably the most important of these is folding. For folding is not only involved in producing shapes such as teeth surfaces and human ear lobes, but is also critical in allowing flat sheets of tissue to form the kinds of pockets and tubes that are so common inside animals.

Folding seems to occur for a variety of reasons. Sometimes it is most likely the direct result of tugging by microscopic fibers. And in other cases it is probably a consequence of growth occurring at different rates in different places, as in the pictures on page 412.



Curves obtained by varying the local curvature according to definite rules as one goes from one end to the other. Each sequence of curves shows what happens when the local curvature is multiplied by a progressively larger factor. The local curvature at any particular point is defined to be the reciprocal of the radius of a circle that approximates the curve at that point. The formulas for local curvature as a function of arc length for each set of pictures are as follows: 1 (circle); s (Cornu spiral or clothoid); s^2; $1/\sqrt{s}$ (involute of circle); $1/s$ (logarithmic or equiangular spiral); $1/s^2$; e^{-s^2}; $Sin[s]$; $s\,Sin[s]$. The curvature functions $f[s]$ can be thought of as specifying how much to turn a vehicle at every moment in order to keep it driving along the curve. The curves have been rotated so as to fit into the frames provided.

But what kinds of shapes can folding produce? The pictures above show what happens when the local curvature—which is essentially the local rate of folding—is taken to vary according to several simple rules as one goes along a curve. In a few cases the shapes produced are rather simple. But in most cases they are fairly complicated. And it takes only very simple rules to generate shapes that look like the villi and other corrugated structures one often sees in animals.

In addition to folding, there are other kinds of processes that are made possible by the lack of rigidity in a developing animal. One is furrowing or tearing of tissue through a loss of adhesion between cells. And another is explicit migration of individual cells based on chemical or immunological affinities.

But how do all these various processes get organized to produce an actual animal? If one looks at the sequence of events that take place in a

typical animal embryo they at first seem remarkably haphazard. But presumably the main thing that is going on—as mentioned above—is that at different places and different times different sections of the underlying genetic program are being used, and these different sections can lead to very different kinds of behavior. Some may produce just uniform growth. Others may lead to various kinds of local folding. And still others may cause regions of tissue to die—thereby for example allowing separate fingers and toes to emerge from a single sheet of tissue.

But just how is it determined what section of the underlying genetic program should be used at what point in the development of the animal? At first, one might think that each individual cell that comes into existence might use a different section of the underlying genetic program. And in very simple animals with just a few hundred cells this is most likely what in effect happens.

But in general it seems to be not so much individual cells as regions of the developing animal that end up using different sections of the underlying program. Indeed, the typical pattern seems to be that whenever a part of an animal has grown to be a few tenths of a millimeter across, that part can break up into a handful of smaller regions which each use a different section of the underlying genetic program.

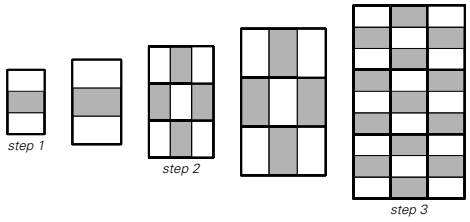
So how does this work? What appears to be the case is that there are cells which produce chemicals whose concentrations decrease over distances of a few tenths of a millimeter. And what has been discovered in the past decade or so is that in all animals—as well as plants—there are a handful of so-called homeobox genes which seem to become active or inactive at particular concentration levels and which control what section of the underlying genetic program will be used.

The existence of a fixed length scale at which such processes occur then almost inevitably implies that an embryo must develop in a somewhat hierarchical fashion. For at a sufficiently early stage, the whole embryo will be so small that it can contain only a handful of regions that use different sections of the genetic program. And at this stage there may, for example, be a leg region, but there will not yet be a distinct foot region.

As the embryo grows, however, the leg region will eventually become large enough that it can differentiate into several separate regions. And at this point, a distinct foot region can appear. Then, when the foot region becomes large enough, it too can break into separate regions that will, say, turn into bone or soft tissue. And when a region that will turn into bone becomes large enough, it can break into further regions that will, say, yield separate individual bones.

If at every stage the tissue in each region produced grows at the same rate, and all that differs is what final type of cells will exist in each region, then inevitably a simple and highly regular overall structure will emerge, as in the idealized picture below. With different substitution rules for each type of cell, the structure will in general be nested. And in fact there are, for example, some parts of the skeletons of animals that do seem to exhibit, at least roughly, a few levels of nesting of this kind.

A schematic illustration of the successive subdivisions which presumably occur in the growth of animals. Here the subdivisions are taken to occur in two directions, always giving three simple rectangles which all grow at the same rate. In practice, the geometry will usually be much more complex.



But in most cases there is no such obvious nesting of this kind. One reason for this is that a region may break not into a simple line of smaller regions, but into concentric circles or into some collection of regions in a much more complicated arrangement—say of the kind that I discuss in the next section. And perhaps even more important, a region may break into smaller regions that grow at different rates, and that potentially fold over or deform in other ways. And when this happens, the geometry that develops will in turn affect the way that subsequent regions break up.

The idea that the basic mechanism for producing different parts of animals is that regions a few tenths of a millimeter across break into separate smaller regions turns out in the end to be strangely similar to the idea that stems of plants whose tips are perhaps a millimeter across

grow by splitting off smaller stems. And indeed it is even known that some of the genetic phenomena involved are extremely similar.

But the point is that because of the comparative rigidity of plants during their most important period of growth, only structures that involve fairly explicit branching can be produced. In animals, however, the lack of rigidity allows a vastly wider range of structures to appear, since now tissue in different regions need not just grow uniformly, but can change shape in a whole variety of ways.

By the time an animal hatches or is born, its basic form is usually determined, and there are bones or other rigid elements in place to maintain this form. But in most animals there is still a significant further increase in size. So how does this work?

Some bones in effect just expand by adding material to their outer surface. But in many cases, bones are in effect divided into sections, and growth occurs between these sections. Thus, for example, the long bones in the arms and legs have regions of growth at each end of their main shafts. And the skull is divided into a collection of pieces that each grow around their edges.

Typically there are somewhat different rates of growth for different parts of an animal—leading, for example, to the decrease in relative head size usually seen from birth to adulthood. And this inevitably means that there will be at least some changes in the shapes of animals as they mature.

But what if one compares different breeds or species of animals? At first, their shapes may seem quite different. But it turns out that among animals of a particular family or even order, it is very common to find that their overall shapes are in fact related by fairly simple and smooth geometrical transformations.

And indeed it seems likely that—much like the leaves and shells that we discussed earlier in this section—differences between the shapes and forms of animals may often be due in large part merely to different patterns in the rates of growth for their different parts.

Needless to say, just like with leaves and shells, such differences can have effects that are quite dramatic both visually and mechanically—turning, say, an animal that walks on four legs into one that walks on

two. And, again just like with leaves and shells, it seems likely that among the animals we see are ones that correspond to a fair fraction of the possible choices for relative rates of growth.

We began this section by asking what underlying rules of growth would be needed to produce the kind of diversity and complexity that we see in the forms of plants and animals. And in each case that we have examined what we have found is that remarkably simple rules seem to suffice. Indeed, in most cases the basic rules actually seem to be somewhat simpler than those that operate in many non-biological systems. But what allows the striking diversity that we see in biological systems is that different organisms and different species of organisms are always based on at least slightly different rules.

In the previous section I argued that for the most part such rules will not be carefully chosen by natural selection, but instead will just be picked almost at random from among the possibilities. From experience with traditional mathematical models, however, one might then assume that this would inevitably imply that all plants and animals would have forms that look quite similar.

But what we have discovered in this book is that when one uses rules that correspond to simple programs, rather than, say, traditional mathematical equations, it is very common to find that different rules lead to quite different—and often highly complex—patterns of behavior. And it is this basic phenomenon that I suspect is responsible for most of the diversity and complexity that we see in the forms of plants and animals.

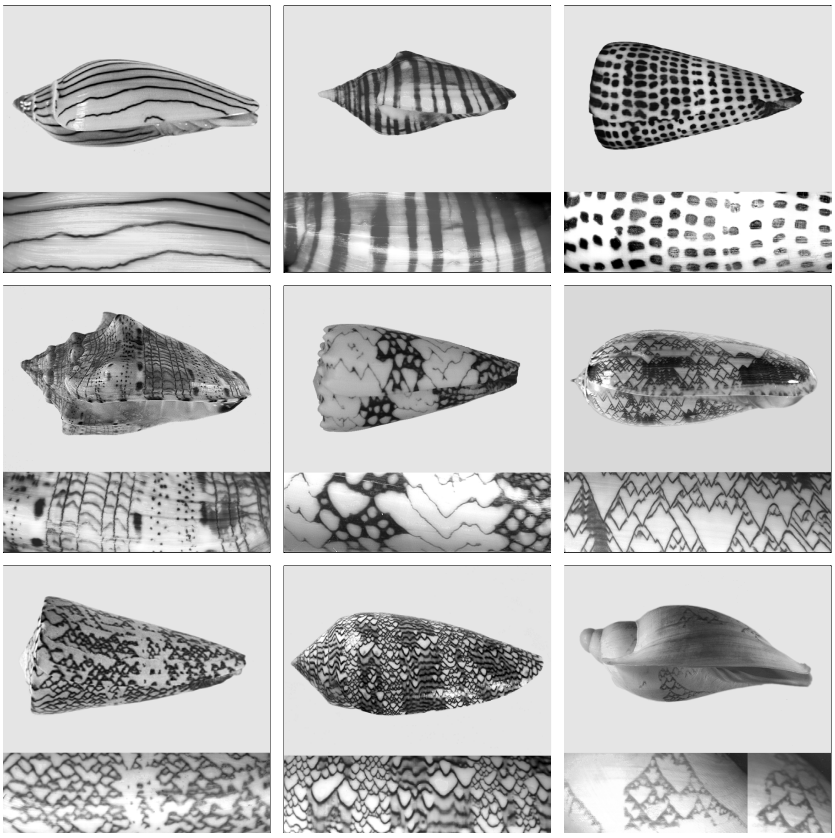
Biological Pigmentation Patterns

At a visual level, pigmentation patterns represent some of the most obvious examples of complexity in biological organisms. And in the past it has usually been assumed that to get the kind of complexity that one sees in such patterns there must be some highly complex underlying mechanism, presumably related to optimization through natural selection.

Following the discoveries in this book, however, what I strongly suspect is that in fact the vast majority of pigmentation patterns in

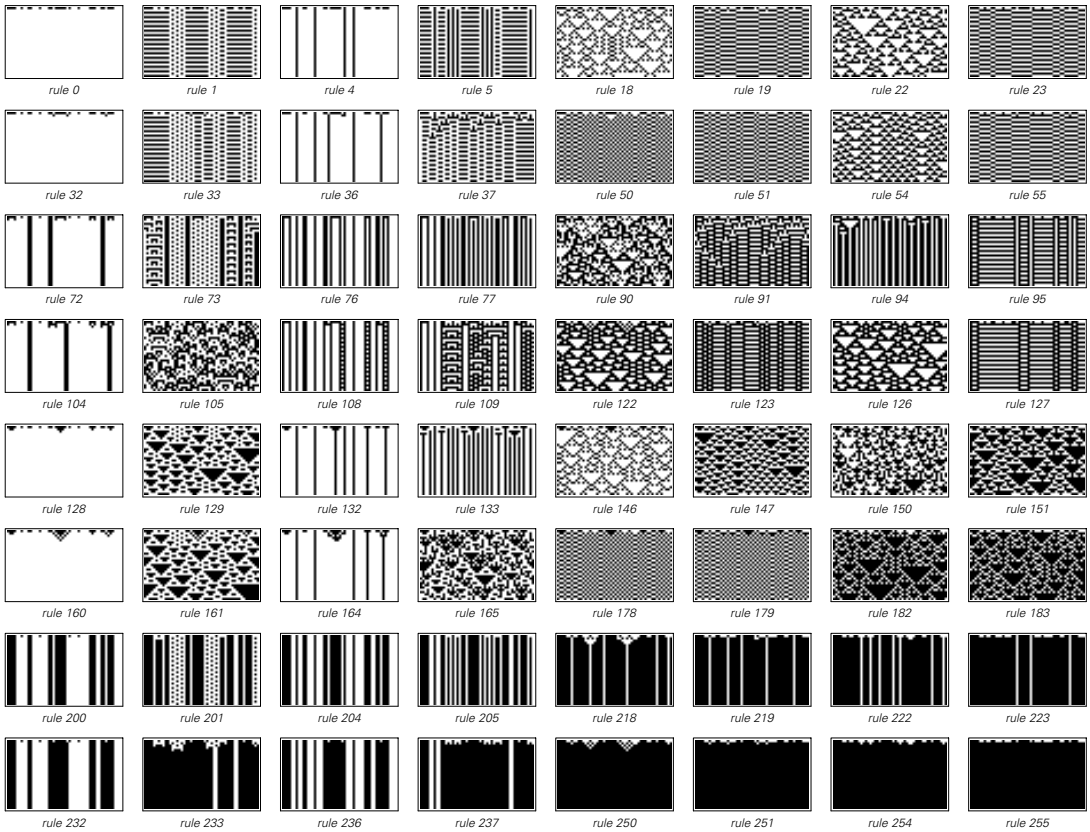
biological organisms are instead generated by processes whose basic rules are extremely simple—and are often chosen essentially at random.

The pictures below shows some typical examples of patterns found on mollusc shells. Many of these patterns are quite simple. But some are highly complex. Yet looking at these patterns one notices a remarkable similarity to patterns that we have seen many times before in this book—generated by simple one-dimensional cellular automata.



Typical examples of pigmentation patterns on mollusc shells. In each close-up the pattern grows from top to bottom, just like in a one-dimensional cellular automaton. Patterns with triangles are often said to have a "tent" or "divaricate" form. The shell on the bottom right is a slightly rare specimen where something close to an explicit nested pattern can be seen. Most of the shells are between one and four inches long; the one on the bottom right is nine inches long. The patterns are all various shades of brown on roughly white backgrounds. The shells are the following types: first row: Elliot's volute, vexillate volute, lettered cone; second row: music volute, banded marble cone, tent olive; third row: bough cone, textile cone, false melon volute (*Livonia mammilla*).

This similarity is, I believe, no coincidence. A mollusc shell, like a one-dimensional cellular automaton, in effect grows one line at a time, with new shell material being produced by a lip of soft tissue at the edge of the animal inside the shell. Quite how the pigment on the shell is laid down is not completely clear. There are undoubtedly elements in the soft tissue that at any point either will or will not secrete pigment. And presumably these elements have certain interactions with each other. And given this, the simplest hypothesis in a sense is that the new state of the element is determined from the previous state of its neighbors—just as in a one-dimensional cellular automaton.



Examples of patterns produced by the evolution of each of the simplest possible symmetrical one-dimensional cellular automaton rules, starting from a random initial condition. The types of patterns obtained show striking similarities to those seen on mollusc shells from the previous page.

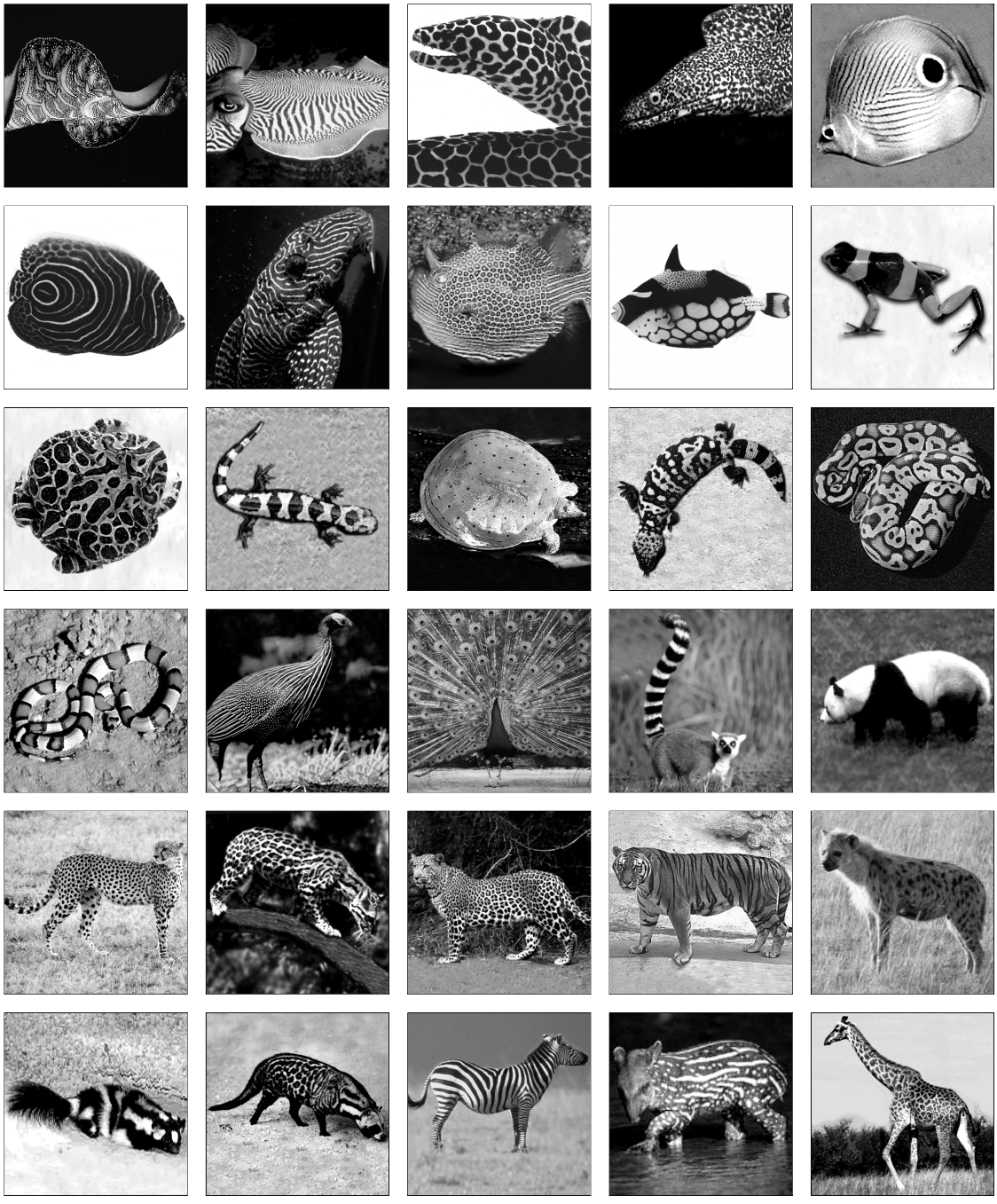
But which specific cellular automaton rule will any given mollusc use? The pictures at the bottom of the facing page show all the possible symmetrical rules that involve two colors and nearest neighbors. And comparing the patterns in these pictures with patterns on actual mollusc shells, one notices the remarkable fact that the range of patterns that occur in the two cases is extremely similar.

Traditional ideas might have suggested that each kind of mollusc would carefully optimize the pattern on its shell so as to avoid predators or to attract mates or prey. But what I think is much more likely is that these patterns are instead generated by rules that are in effect chosen at random from among a collection of the simplest possibilities. And what this means is that insofar as complexity occurs in such patterns it is in a sense a coincidence. It is not that some elaborate mechanism has specially developed to produce it. Rather, it just arises as an inevitable consequence of the basic phenomenon discovered in this book that simple rules will often yield complex behavior.

And indeed it turns out that in many species of molluscs the patterns on their shells—both simple and complex—are completely hidden by an opaque skin throughout the life of the animal, and so presumably cannot possibly have been determined by any careful process of optimization or natural selection.

So what about pigmentation patterns on other kinds of animals? Mollusc shells are almost unique in having patterns that are built up one line at a time; much more common is for patterns to develop all at once all over a surface.

Most often what seems to happen is that at some point in the growth of an embryo, precursors of pigment-producing cells appear on its surface, and groups of these cells associated with pigments of different colors then become arranged in a definite pattern. Typically each individual group of cells is initially some fraction of a tenth of a millimeter across. But since different parts of an animal usually grow at different rates, the final pattern that one sees on an adult animal ends up being scaled differently in different places—so that, for example, the pattern is smaller in scale on the head of an animal, since the head grows more slowly.

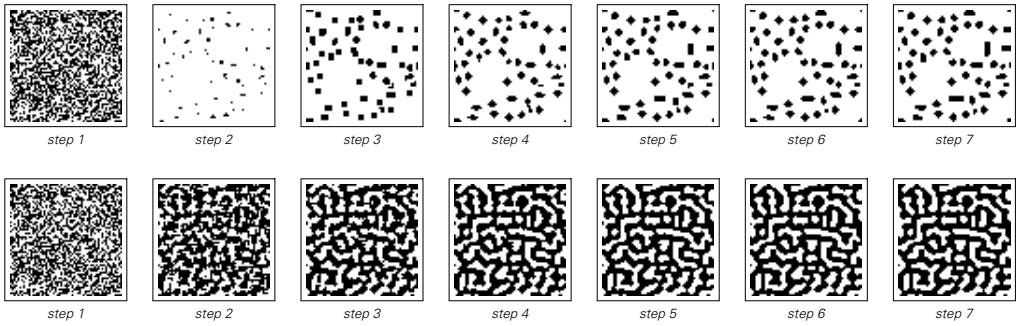


Typical examples of pigmentation patterns on animals. Note that many very different animals end up having remarkably similar patterns.

The pictures on the facing page show typical examples of pigmentation patterns in animals, and demonstrate that even across a vast range of different types of animals just a few kinds of patterns occur over and over again. So how are these patterns produced? Even though some of them seem quite complex, it turns out that once again there is a rather simple kind of rule that can account for them.

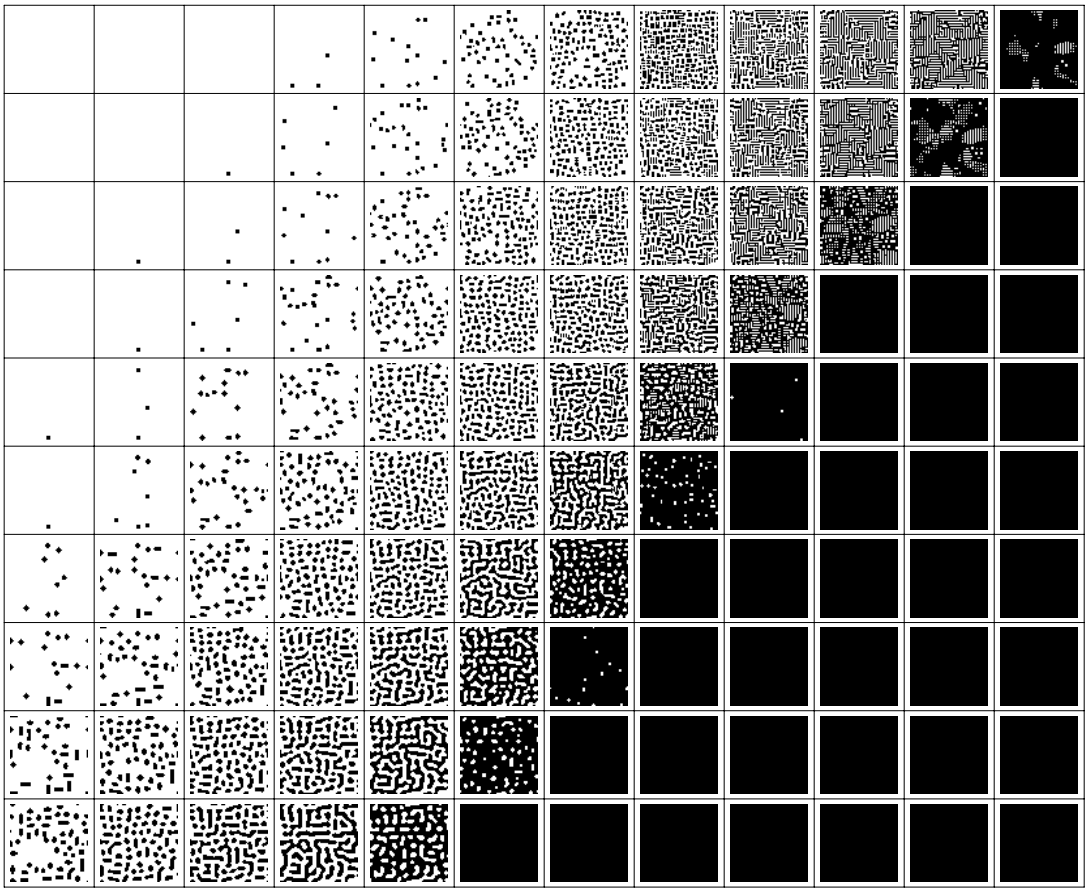
The idea is that when a pattern forms, the color of each element will tend to be the same as the average color of nearby elements, and opposite to the average color of elements further away. Such an effect could have its origin in the production and diffusion of activator and inhibitor chemicals, or, for example, in actual motion of different types of cells. But regardless of its origin, the effect itself can readily be captured just by setting up a two-dimensional cellular automaton with appropriate rules.

The pictures below show what happens with two slightly different choices for the relative importance of elements that are further away. In both cases, starting from a random distribution of black and white elements there quickly emerge definite patterns—in the first case a collection of spots, and in the second case a maze-like or labyrinthine structure.



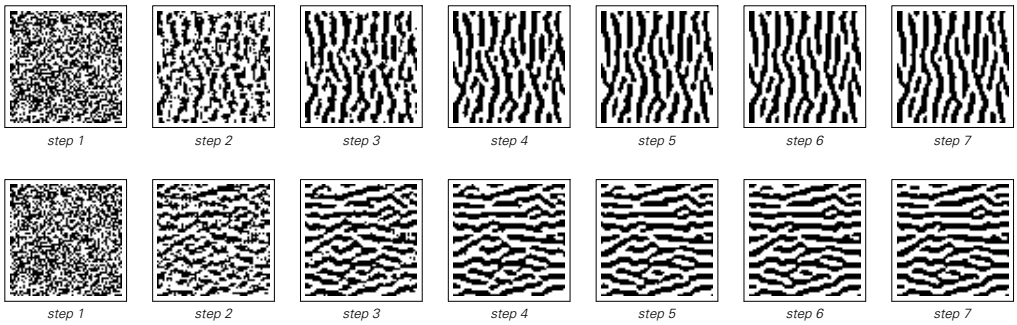
Evolution of simple two-dimensional cellular automata in which the color of each cell at each step is determined by looking at a weighted sum of the average colors of cells up to distance 3 away. In both rules shown the cell itself and its nearest neighbors enter with weight 1. Cells at distances 2 and 3 enter with negative weights— -0.4 per cell for the first rule, and -0.2 for the second. A cell becomes black if the weighted sum is positive, and white otherwise. Starting from random initial conditions, both rules quickly evolve to stationary states that look very much like pigmentation patterns seen in animals.

The next page shows the final patterns obtained with a whole array of different choices of weightings for elements at different distances. A certain range of patterns emerges—almost all of which turn out to be quite similar to patterns that one sees on actual animals.



Patterns generated by rules of the type shown on the previous page, with a range of choices for the weights of cells at distances 2 and 3. Weights vary from -0.9 to 0 down the page for distance 2, and from -0.7 to 0.4 across the page for distance 3. In all cases the evolution starts from the same random initial condition, and is continued until it stabilizes. Note that pigmentation patterns for actual animals may contain either larger or smaller numbers of elements than the patterns shown here.

But all of these patterns in a sense have the same basic form in every direction. Yet there are many animals whose pigmentation patterns exhibit stripes with a definite orientation. Sometimes these stripes are highly regular, and can potentially arise from any of the possible mechanisms that yield repetitive behavior. But in cases where the stripes are less regular they typically look very much like the patterns generated in the pictures at the top of the facing page using a version of the simple mechanism described above.



Examples of rules in which cells in the horizontal and vertical directions are weighted differently. In the first case, cells at distances 2 and 3 only have an effect in the vertical direction; in the second case, they only have an effect in the horizontal direction. The result is the formation of either vertical or horizontal stripes.

Financial Systems

During the development of the ideas in this book I have been asked many times whether they might apply to financial systems. There is no doubt that they do, and as one example I will briefly discuss here what is probably the most obvious feature of essentially all financial markets: the apparent randomness with which prices tend to fluctuate.

Whether one looks at stocks, bonds, commodities, currencies, derivatives or essentially any other kind of financial instrument, the sequences of prices that one sees at successive times show some overall trends, but also exhibit varying amounts of apparent randomness.

So what is the origin of this randomness?

In the most naive economic theory, price is a reflection of value, and the value of an asset is equal to the total of all future earnings—such as dividends—which will be obtained from it, discounted for the interest that will be lost from having to wait to get these earnings.

With this view, however, it seems hard to understand why there should be any significant fluctuations in prices at all. What is usually said is that prices are in fact determined not by true value, but rather by the best estimates of that value that can be obtained at any given time. And it is then assumed that these estimates are ultimately affected by all sorts of events that go on in the world, making random movements

in prices in a sense just reflections of random changes going on in the outside environment.

But while this may be a dominant effect on timescales of order weeks or months—and in some cases perhaps even hours or days—it is difficult to believe that it can account for the apparent randomness that is often seen on timescales as short as minutes or even seconds.

In addition, occasionally one can identify situations of seemingly pure speculation in which trading occurs without the possibility of any significant external input—and in such situations prices tend to show more, rather than less, seemingly random fluctuations.

And knowing this, one might then think that perhaps random fluctuations are just an inevitable feature of the way that prices adjust to their correct values. But in negotiations between two parties, it is common to see fairly smooth convergence to a final price. And certainly one can construct algorithms that operate between larger numbers of parties that would also lead to fairly smooth behavior.

So in actual markets there is presumably something else going on. And no doubt part of it is just that the sequence of trades whose prices are recorded are typically executed by a sequence of different entities—whether they be humans, organizations or programs—each of which has its own detailed ways of deciding on an appropriate price.

But just as in so many other systems that we have studied in this book, once there are sufficiently many separate elements in a system, it is reasonable to expect that the overall collective behavior that one sees will go beyond the details of individual elements.

It is sometimes claimed that it is somehow inevitable that markets must be random, since otherwise money could be made by predicting them. Yet many people believe that they make money in just this way every day. And beyond certain simple situations, it is difficult to see how feedback mechanisms could exist that would systematically remove predictable elements whenever they were used.

No doubt randomness helps in maintaining some degree of stability in markets—just as it helps in maintaining stability in many other kinds of systems that we have discussed in this book. Indeed, most markets are set up so that extreme instabilities associated with

certain kinds of loss of randomness are prevented—sometimes by explicit suspension of trading.

But why is there randomness in markets in the first place?

Practical experience suggests that particularly on short timescales much of the randomness that one sees is purely a consequence of internal dynamics in the market, and has little if anything to do with the nature or value of what is being traded.

So how can one understand what is going on? One needs a basic model for the operation and interaction of a large number of entities in a market. But traditional mathematics, with its emphasis on reducing everything to a small number of continuous numerical functions, has rather little to offer along these lines.

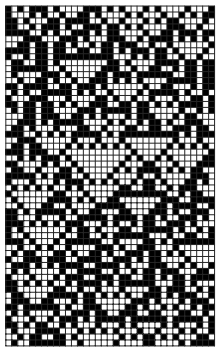
The idea of thinking in terms of programs seems, however, much more promising. Indeed, as a first approximation one can imagine that much as in a cellular automaton entities in a market could follow simple rules based on the behavior of other entities.

To be at all realistic one would have to set up an elaborate network to represent the flow of information between different entities. And one would have to assign fairly complicated rules to each entity—certainly as complicated as the rules in a typical programmed trading system. But from what we have learned in this book it seems likely that this kind of complexity in the underlying structure of the system will not have a crucial effect on its overall behavior.

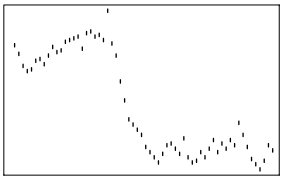
And so as a minimal idealization one can for example try viewing a market as being like a simple one-dimensional cellular automaton. Each cell then corresponds to a single trading entity, and the color of the cell at a particular step specifies whether that entity chooses to buy or sell at that step. One can imagine all sorts of schemes by which such colors could be updated. But as a very simple idealization of the way that information flows in a market, one can, for example, take each color to be given by a fixed rule that is based on each entity looking at the actions of its neighbors on the previous step.

With traditional intuition one would assume that such a simple model must have extremely simple behavior, and certainly nothing like what is seen in a real market. But as we have discovered in this book,

simple models do not necessarily have simple behavior. And indeed the picture below shows an example of the behavior that can occur.



An example of a very simple idealized model of a market. Each cell corresponds to an entity that either buys or sells on each step. The behavior of a given cell is determined by looking at the behavior of its two neighbors on the step before according to the rule shown. The plot below gives as a rough analog of a market price the running difference of the total numbers of black and white cells at successive steps. And although there are patches of predictability that can be seen in the complete behavior of the system the plot on the right looks in many respects random.



In real markets, it is usually impossible to see in detail what each entity is doing. Indeed, often all that one knows is the sequence of prices at which trades are executed. And in a simple cellular automaton the rough analog of this is the running difference of the total numbers of black and white cells obtained on successive steps.

And as soon as the underlying rule for the cellular automaton is such that information will eventually propagate from one entity to all others—in effect a minimal version of an efficient market hypothesis—it is essentially inevitable that running totals of numbers of cells will exhibit significant randomness.

One can always make the underlying system more complicated—say by having a network of cells, or by allowing different cells to have different and perhaps changing rules. But although this will make it more difficult to recognize definite rules even if one looks at the complete behavior of every element in the system, it does not affect the basic point that there is randomness that can intrinsically be generated by the evolution of the system.

Implications for Everyday Systems

Issues of Modelling

▪ **Page 363 · Uncertainties of this chapter.** In earlier chapters of this book what I have said can mostly be said with absolute certainty, since it is based on observations about the behavior of purely abstract systems that I have explicitly constructed. But in this chapter, I study actual systems that exist in nature, and as a result, most of what I say cannot be said with any absolute certainty, but instead must involve a significant component of hypothesis. For I no longer control the basic rules of the systems I am studying, and instead I must just try to deduce these rules from observation—with the potential that despite my best efforts my deductions could simply be incorrect.

▪ **Experiences of modelling.** Over the course of the past 25 years I have constructed an immense number of models for a wide range of scientific, technical and business purposes. But while these models have often proved extremely useful in practice, I have usually considered them intellectually quite unsatisfactory. For being models, they are inevitably incomplete, and it is never in any definitive sense possible to establish their validity.

▪ **Page 363 · Notes on this chapter.** Much of this book is concerned with topics that have never been discussed in any concrete form before, so that between the main text and these notes I have been able to include a large fraction of everything that is known about them. But in this chapter (as well as some of the ones that follow) the systems I consider have often had huge amounts written about them before, making any kind of complete summary quite impossible.

▪ **Material for this chapter.** Like the rest of this book, this chapter is strongly based on my personal work and observations. For almost all of the systems discussed I have personally collected extensive data and samples, often over the course of many years, and sometimes in quite unlikely and amusing circumstances. I have also tried to study the existing scientific literature, and indeed in working on this chapter I have looked at many thousands of papers and books—even though the vast majority of them tend to ignore overall issues, and instead concentrate on details of often excruciating specificity.

▪ **Page 365 · Models versus experiments.** In modern science it is usually said that the ultimate test of any model is its agreement with experiment. But this is often interpreted to mean that if an experiment ever disagrees with a model, then the model must be wrong. Particularly when the model is simple and the experiment is complex, however, my personal experience has been that it is quite common for it to be the experiment, rather than the model, that is wrong. When I started doing particle physics in the mid-1970s I assumed—like most theoretical scientists—that the results of experiments could somehow always be treated as rigid constraints on models. But in 1977 I worked on constructing the first model based on QCD for heavy particle production in high-energy proton-proton collisions. The model predicted a certain rate for the production of such particles. But an experiment which failed to see any of these particles implied that the rate must be much lower. And on the basis of this I spent great effort trying to see what might be wrong with the model—only to discover some time later that in fact the methodology of the experiment was flawed and its results were wrong. At first I thought that perhaps this was an isolated incident. But soon I had seen many examples where the stated results of physics experiments were incorrect, either through straightforward mistakes or through subtly prejudiced analysis. And outside of physics, I have tended to find still less reliability in the results of complex experiments.

▪ **Page 366 · Models versus reality.** Questions about the correspondence between models and reality have been much debated in the philosophy of science for many centuries, and were, for example, central to the disagreement between Galileo and the church in the early 1600s. Many successful

models are in practice first introduced as convenient calculational devices, but later turn out to have a direct correspondence to reality. Two examples are planets orbiting the Sun, and quarks being constituents of particles. It remains to be seen whether such models as the imaginary time statistical mechanics formalism for quantum mechanics (see page 1061) turn out to have any direct correspondence to reality.

■ **History of modelling.** Creation myths can in a sense be viewed as primitive models. Early examples of models with more extensive structure included epicycles. Traditional mathematical models of the modern type originated in the 1600s. The success of such models in physics led to attempts to imitate them in other fields, but for the most part these did not succeed. The idea of modelling intricate patterns using programs arose to some extent in the study of fractals in the late 1970s. And the notion of models based on simple programs such as cellular automata was central to my work in the early 1980s. But despite quite a number of fairly well-known successes, there is even now surprisingly little understanding among most scientists of the idea of models based on simple programs. Work in computer graphics—with its emphasis on producing pictures that look right—has made some contributions. And it seems likely that the possibility of computerized and especially image-based data taking will contribute further. (See also page 860.)

■ **Page 367 · Finding models.** Even though a model may have a simple form, it may not be at all easy to find. Indeed, many of the models in this chapter took me a very long time to find. By far my most common mistake was trying to build too much into the basic structure of the model. Often I was sure that some feature of the behavior of a system must be built into the underlying model—yet I could see no simple way to do it. But eventually what happened was that I tried a few other very simple models, and to my great surprise one of them ended up showing the behavior I wanted, even though I had in no way explicitly built it in.

■ **Page 369 · Consequences of models.** Given a program it is always possible to run the program to find out what it will do. But as I discuss in Chapter 12, when the behavior is complex it may take an irreducible amount of computational work to answer any given question about it. However, this is not a sign of imperfection in the model; it is merely a fundamental feature of complex behavior.

■ **Universality in models.** With traditional models based on equations, it is usually assumed that there is a unique correct version of any model. But in the previous chapter we saw that it is possible for quite different programs to yield essentially the same large-scale behavior, implying that with programs there can be many models that have the same consequences but different detailed underlying structure.

The Growth of Crystals

■ **Page 369 · Nucleation.** In the absence of container walls or of other objects that can act as seeds, liquids and gases can typically be supercooled quite far below their freezing points. It appears to be extremely unlikely for spontaneous microscopic fluctuations to initiate crystal growth, and natural snowflakes, for example, presumably nucleate around dust or other particles in the air. Snowflakes in man-made snow are typically nucleated by synthetic materials. In this case and in experiments on cloud seeding it has been observed that the details of seeds can affect the overall shapes of crystals that grow from them.

■ **Page 369 · Implementation.** One can treat hexagonal lattices as distorted square lattices, updated according to

```
CAStep[rule_List, a_] := Map[rule[[14 - #]] &,
    a + 2 ListConvolve[{{1, 1, 0}, {1, 0, 1}, {0, 1, 1}}, a, 2], {2}]
```

where `rule = IntegerDigits[code, 2, 14]`. On this page the rule used is code 16382; on page 371 it is code 10926. The centers of an array of regular hexagons are given by `Table[{i Sqrt[3], j}, {i, 1, m}, {j, Mod[i, 2], n, 2}]`.

■ **Page 372 · Identical snowflakes.** The widespread claim that no two snowflakes are alike is not in practice true. It is however the case that as a result of turbulent air currents a collection of snowflakes that fall to the ground in a particular region will often have come from very different regions of a cloud, and therefore will have grown in different environments. Note that the reason that the six arms of a single snowflake usually look the same is that all of them have grown in essentially the same environment. Deviations are usually the result of collisions between falling snowflakes.

■ **History of snowflake studies.** Rough sketches of snowflakes were published by Olaus Magnus of Uppsala around 1550. Johannes Kepler made more detailed pictures and identified hexagonal symmetry around 1611. Over the course of the next few centuries, following work by René Descartes, Robert Hooke and others, progressively more accurate pictures were made and correlations between weather conditions and snowflake forms were found. Thousands of photographs of snowflakes were taken by Wilson Bentley over the period 1884–1931. Beginning in 1932 an extensive study of snowflakes was made by Ukichiro Nakaya, who in 1936 also produced the first artificial snowflakes. Most of the fairly

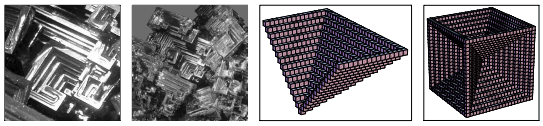
small amount of more recent work on snowflakes has been done as part of more general studies on dendritic crystal growth. Note that tree-like snowflakes are what make snow fluffy, while simple hexagons make it denser and more slippery. The proportion of different types of snowflakes is important in understanding phenomena such as avalanches.

▪ **History of crystal growth.** The vast majority of work done on crystal growth has been concerned with practical methods rather than with theoretical analyses. The first synthetic gemstones were made in the mid-1800s, and methods for making high-quality crystals of various materials have been developed over the course of the past century. Since the mid-1970s such crystals have been crucial to the semiconductor industry. Systematic studies of the symmetries of crystals with flat facets began in the 1700s, and the relationship to internal structure was confirmed by X-ray crystallography in the 1920s. The many different possible external forms of crystals have been noted in mineralogy since Greek times, but although classification schemes have been given, these forms have apparently still not been studied in a particularly systematic way.

▪ **Models of crystal growth.** There are two common types of models for crystal growth: ones based on the physics of individual atoms, and ones based on continuum descriptions of large collections of atoms. In the former category, it was recognized in the 1940s that a single atom is very unlikely to stick to a completely flat surface, so growth will always tend to occur at steps on a crystal surface, often associated with screw dislocations in the crystal structure. In practice, however, as scanning tunnelling microscopes have revealed, most crystal surfaces that are not grown at an extremely slow rate tend to be quite rough at an atomic scale—and so it seems that for example the aggregation model from page 331 may be more appropriate. In snowflakes and other crystals features such as the branches of tree-like structures are much larger than atomic dimensions, so a continuum description can potentially be used. It is possible to write down a nonlinear partial differential equation for the motion of the solidification front, taking into account basic thermodynamic effects. The first result (discovered by William Mullins and Robert Sekerka in 1963) is that if every part of the front is at the same temperature, then any deviations from planarity in the front will tend to grow. The shape of the front is presumably stabilized by the Gibbs-Thomson effect, which implies that the freezing temperature is lower when the front is more curved. The characteristic length for deformations of the front turns out to be the geometric mean of a microscopic length associated with surface energy and a macroscopic length associated with diffusion. It is this characteristic length that presumably determines the size of an individual cell in the cellular automaton model.

Dendritic crystals are commonly seen in ice formations on windows, and in pieces of aluminum of the kind found at typical hardware stores.

▪ **Hopper crystals.** When a pool of molten bismuth solidifies it tends to form crystals like those in the first two pictures below. What seems to give these crystals their characteristic "hoppered" shapes is that there is more rapid growth at the edges of each face than at the center. (Spirals are probably associated with underlying screw dislocations.) Hoppering has not been much studied for scientific purposes, but has been noticed in many substances, including galena, rose quartz, gold, calcite, salt and ice.



▪ **Page 373 · Other models.** There are many ways to extend the simple cellular automata shown here. One possibility is to allow dependence on next-nearest as well as nearest neighbors. In this case it turns out that non-convex as well as convex faceted shapes can be obtained. Another possibility is to allow cells that have become black to turn white again. In this case all the various kinds of patterns that we saw in Chapter 5 can occur. A general feature of cellular automaton rules is that they are fundamentally local. Some models of crystal growth, however, call for long-range effects such as a temperature field which changes throughout the crystal in an effectively instantaneous way. It turns out, however, that many seemingly long-range effects can actually be captured quite easily in cellular automata. In a typical case, this can be done by introducing a third possible color for each cell, and then having rapidly changing arrangements of this color.

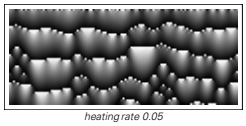
▪ **Polycrystalline materials.** When solids with complicated forms are seen, it has usually been assumed that they must be aggregates of many separate crystals, with each crystal having a simple faceted shape. But the results given here indicate that in fact individual crystals can yield highly complex shapes. There will nevertheless be cases however where multiple crystals are involved. These can be modelled by having a cellular automaton in which one starts from several separated seeds. Sometimes the regions associated with different seeds may have different characteristics; the boundaries between these regions then form a Voronoi diagram (see page 1038).

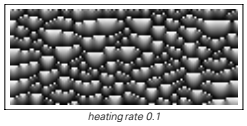
■ **Quasicrystals.** In some special materials it was discovered in 1984 that atoms are arranged not on a purely repetitive grid, but instead in a pattern with the nested type of structure discussed on page 932. A characteristic feature of such patterns is that they can have approximate pentagonal or icosahedral symmetry, which is impossible for purely repetitive patterns. It has usually been assumed that the arrangement of atoms in a quasicrystal is determined by satisfying a constraint analogous to minimization of energy. And as we saw on page 932 it is indeed possible to get nested patterns by requiring that certain constraints be satisfied. But another explanation for such patterns is that they are the result of growth processes that are some kind of cross between those on pages 373 and 659.

■ **Amorphous materials.** When solidification occurs fairly slowly, atoms have time to arrange themselves in a regular crystalline way. But if the cooling is sufficiently rapid, amorphous solids such as glasses are often formed. And in such cases, the packing of atoms is quite random—except that locally there is often approximate icosahedral structure, analogous to that discussed on page 943. (See also page 986.)

■ **Diffusion-limited aggregation (DLA).** DLA is a model for a variety of natural growth processes that was invented by Thomas Witten and Leonard Sander in 1981, and which at first seems quite different from a cellular automaton. The basic idea of DLA is to build up a cluster of black cells by starting with a single black cell and then successively introducing new black cells far away that undergo random walks and stick to the cluster as soon as they come into contact with it. The patterns that are obtained by this procedure turn out for reasons that are still not particularly clear to have a random but on average nested form. (Depending on precise details of the underlying model, very large clusters may sometimes not have nested forms, at least in 2D.) The basic reason that DLA patterns are not very dense is that once arms have formed on the outside of the cluster, they tend to catch new cells before these cells have had a chance to go inside. It turns out that at a mathematical level DLA can be reproduced by solving the Laplace equation at each step with a constant boundary condition on the cluster, and then using the result to give the probability for adding a new cell at each point on the cluster. To construct a cellular automaton analog of DLA one can introduce gray as well as black and white cells, and then have the gray cells represent pieces of solid that have not yet become permanently attached to the main cluster. Rapid rearrangement of gray cells on successive steps can then have a similar effect to the random walks that occur in the usual DLA model. Whether a pattern with all the properties expected in DLA is produced seems to depend in some detail on the rules for the gray cells. But so long as there is effective randomness in the successive positions of these cells, and so long as the total number of them is conserved, then it appears that DLA-like results are usually obtained. No doubt there are also simpler cellular automaton rules that yield similar results. (See also page 979.)

■ **Boiling.** The boiling of a liquid such as water involves a kind of growth inhibition that is in some ways analogous to that seen in dendritic crystal growth. When a particular piece of liquid boils—forming a bubble of gas—a certain latent heat is consumed, reducing the local temperature, and inhibiting further boiling. In the pictures below the liquid is divided into cells, with each cell having a temperature from 0 to 1, corresponding exactly to a continuous cellular automaton of the kind discussed on page 155. At each step, the temperature of every cell is given by the average of its temperature and the temperatures of its neighbors, representing the process of heat diffusion, with a constant amount added to represent external heating. If the temperature of any cell exceeds 1, then only the fractional part is kept, as in the systems on page 158, representing the consumption of latent heat in the boiling process. The pictures below illustrate the kind of seemingly random pattern of bubble formation that can be heard in the noise produced by boiling water.


heating rate 0.05


heating rate 0.1

The Breaking of Materials

■ **Phenomenology of microscopic fracture.** Different materials show rather different characteristics depending on how ductile or brittle they are. Ductile materials—such as taffy or mild steel—bend and smoothly neck before breaking. Brittle materials—such as chalk or glass—do not deform significantly before catastrophic failure. Ductile materials in effect flow slightly before breaking, and as a result their fracture surfaces tend to be less jagged. In addition, in response to stresses in the material, small voids often form—perhaps nucleating around imperfections—yielding a pock-marked surface. In brittle materials, the beginning of the fracture surface typically looks quite mirror-like, then it starts to look misty, and finally, often at a sharply defined point, it begins to look complex and hackled. (This sequence is qualitatively not unlike the initiation of randomness in turbulent fluid flow and many other systems.) Cracks in

brittle materials typically seem to start slowly, then accelerate to about half the Rayleigh speed at which small deformation waves on the surface would propagate. Brittle fracture involves violent breaking of atomic bonds; it usually leaves a jagged surface, and can lead to emission of both high-frequency sound as well as light. Directly around a crack complex patterns of stress are typically produced, though away from the crack they resolve quickly to a fairly smooth and simple form. It is known that ultrasound can affect the course of cracks, suggesting that crack propagation is affected by local stresses. There are many different detailed geometries for fracture, associated with snapping, tearing, shattering, pulling apart, and so on. In many situations, individual cracks will split into multiple cracks as they propagate, sometimes producing elaborate tree-like structures. The statistical properties of fracture surfaces have been studied fairly extensively. There is reasonable evidence of self-similarity, typically associated with a fractal dimension around 0.8 or slightly smaller.

▪ **Models of microscopic fracture.** Two kinds of models have traditionally been studied: ones based on looking at arrays of atoms, and ones based on continuum descriptions of materials. At the atomic level, a simple model suggested fairly recently is that atoms are connected by bonds with a random distribution of strengths, and that cracks follow paths that minimize the total strength of bonds to be broken. It is not clear why in a crystal bonds should be of different strengths, and there is some evidence that this model yields incorrect predictions for the statistical properties of actual cracks. A slightly better model, related to the one in the main text, is that the bonds between atoms are identical, and act like springs which break when they are stretched too far. In recent years, computer simulations with millions of atoms have been carried out—usually with realistic but complicated interatomic force laws—and some randomness has been observed, but its origins have not been isolated. A set of nonlinear partial differential equations known as the Lamé equations are commonly used as a continuum description of elastic materials. Various instabilities have been found in these equations, but the equations are based on small deformations, and presumably cannot be relied upon to provide information about fracture.

▪ **History.** Fracture has been a critical issue throughout the history of engineering. Its scientific study was particularly stimulated by failures of various types of ships and aircraft in the 1940s and 1950s, and many quantitative empirical results were obtained, so that by the 1960s ductile fracture as an engineering issue became fairly well understood. In the 1980s, ideas about fractals suggested new interpretations of fracture surfaces, and in the past few years, various models of fracture based on ideas from statistical physics have been tried. Atomic-level computer experiments on fracture began in earnest in the late 1980s, but only very recently has it been possible to include enough atoms to even begin addressing questions about the structure of cracks.

▪ **Page 375 · Experimental data.** To investigate the model in the main text requires looking not only at the path of a crack, but also at dislocations of atoms near it. To do this dynamically is difficult, but in a perfect crystal final patterns of dislocations that remain at the edge of a region affected by fracture can be seen for example by electron diffraction. And it turns out that these often look remarkably like patterns made by 1D class 3 cellular automata. (Similar patterns may perhaps also be seen in recent detailed simulations of fracture processes in arrays of idealized atoms.)

▪ **Large-scale fractures.** It is remarkable to what extent very large-scale fractures can look like small-scale ones. If the path of a crack were, say, a perfect random walk, then one might imagine that large-scale cracks could simply be combinations of many small-scale segments. But when one looks at geological systems, for example, the smallest relevant scales for the cracks one sees are certainly no smaller than particles of soil. And as a result, one needs a more general mechanism, not just one that just relates to atoms and molecules.

▪ **Alternate models.** It is straightforward to set up 3-color cellular automata with the same basic idea as in the main text, but in which there is no need for a special cell to represent the crack. In addition, instead of modelling the displacement of atoms, one can try to model directly the presence or absence of atoms at particular positions. And then one can start from a repetitive array of cells, with a perturbation to represent the beginning of the crack.

▪ **Electric breakdown.** Somewhat related to fracture is the process of electric breakdown, visible for example in lightning, Lichtenberg figures or plasma-filled glass globes used as executive toys. At least in the case of lightning, there is some evidence that small inhomogeneities in the atmosphere can be important in producing at least some aspects of the apparent randomness that is seen. (With electric potential thought of like a diffusion field, models based on diffusion-limited aggregation are sometimes used.)

▪ **Crushing.** For a rather wide range of cases it appears that in crushed solids such as rocks the probability of a particular fragment having a diameter larger than r is given approximately by $r^{-2.5}$. It seems likely that the origin of this is

that each rock has a certain probability to break into, say, two smaller rocks at each stage in the crushing process, much as in a substitution system.

▪ **Effects of microscopic roughness.** The two most obvious features that are affected by the microscopic roughness of materials are visual appearance and sliding friction. A perfectly flat surface will reflect light like a mirror. Roughness will lead to more diffuse reflection, although the connection between observed properties of rough surfaces and typical parametrizations used in computer graphics is not clear.

The friction force that opposes sliding is usually assumed to be proportional purely to the force with which surfaces are pressed together. Presumably at least the beginning of the explanation for this slightly bizarre fact is that most of the friction force is associated with microscopic peaks in rough surfaces, and that the number of these peaks that come into close contact increases as surfaces are pushed together.

▪ **Crinkling.** A question somewhat related to fracture concerns the generation of definite creases in crumpled or wrinkled objects such as pieces of paper or fabric. It is not too difficult to make various statements about details of the particular arrangements of creases that can occur, but nothing seems to be known about the origin of the overall randomness that is almost universally seen.

Fluid Flow

▪ **Page 376 · Reynolds numbers.** If a system is to act like a continuum fluid, then almost by definition its behavior can involve only a limited number of macroscopic quantities, such as density and velocity. And from this it follows that patterns of flow should not depend separately on absolute speeds and sizes. Instead, the character of a flow should typically be determined by a single Reynolds number, $Re = U L/\nu$, where U is the characteristic speed of the flow (measured say in *cm/sec*), L is a characteristic size (measured say in *cm*), and ν is the kinematic viscosity of the fluid. For water, $\nu = 0.01$, for air $\nu = 0.15$, and for glycerine $\nu = 10000$, all in units of cm^2/sec. In flow past a cylinder it is conventional to take L to be the diameter of the cylinder. But the fact that the form of flow should depend only on Reynolds number means that in the pictures in the main text for example it is not necessary to specify absolute sizes or speeds: one need only know the product $U L$ that appears in the Reynolds number. In practice, moving one's finger slowly through water gives a Reynolds number of about 100 (so that a regular array of dimples corresponding to eddies are visible behind one's finger), walking in air about 10,000, a boat in the millions, and a large airplane in the billions.

The Reynolds number roughly measures the ratio of inertial to viscous effects. When the Reynolds number is small the viscous damping dominates, and the flow is laminar and smooth. When the Reynolds number is large, inertia associated with fluid motions dominates, and the flow is turbulent and complicated.

In different systems, the characteristic length used typically in the definition of Reynolds number is different. In most cases, however, the transition from laminar to turbulent flow occurs at Reynolds numbers around a hundred.

In some situations, however, Reynolds number alone does not appear to be sufficient to determine when a flow will become turbulent. Indeed, modern experiments on streams of dye in water (or rising columns of smoke) typically show a transition to turbulence at a significantly lower Reynolds number than the original experiments on these systems done by Osborne Reynolds in the 1880s. Presumably the reason for this is that the transition point can be lowered by perturbations from the environment, and such perturbations are more common in the modern mechanized world. If perturbations are indeed important, it implies that a traditional fluid description is not adequate. I suspect, however, that even though perturbations may determine the precise point at which turbulence begins, intrinsic randomness generation will dominate once turbulence has been initiated.

▪ **Navier-Stokes equations.** The traditional model of fluids used in physics is based on a set of partial differential equations known as the Navier-Stokes equations. These equations were originally derived in the 1840s on the basis of conservation laws and first-order approximations. But if one assumes sufficient randomness in microscopic molecular processes they can also be derived from molecular dynamics, as done in the early 1900s, as well as from cellular automata of the kind shown on page 378, as I did in 1985 (see below). For very low Reynolds numbers and simple geometries, it is often possible to find explicit formulas for solutions to the Navier-Stokes equations. But even in the regime of flow where regular arrays of eddies are produced, analytical methods have never yielded complete explicit solutions. In this regime, however, numerical approximations are fairly easy to find. Since about the 1960s computers have been powerful enough to allow computations at least nominally to be extended to considerably higher Reynolds numbers. And indeed it has become increasingly common to see numerical results given far into the turbulent regime—leading

sometimes to the assumption that turbulence has somehow been derived from the Navier-Stokes equations. But just what such numerical results actually have to do with detailed solutions to the Navier-Stokes equations is not clear. For in particular it ends up being almost impossible to distinguish whatever genuine instability and apparent randomness may be implied by the Navier-Stokes equations from artifacts that get introduced through the discretization procedure used in solving the equations on a computer. One of the key advantages of my cellular automaton approach to fluids is precisely that it does not require any such approximations.

At a mathematical level analysis of the Navier-Stokes has never established the formal uniqueness and existence of solutions. Indeed, there is even some evidence that singularities might almost inevitably form, which would imply a breakdown of the equations, and perhaps a need to account for underlying molecular processes.

In turbulent flow at higher Reynolds numbers there begin to be eddies with a wide range of sizes. And to capture all these eddies in a computation eventually involves prohibitively large amounts of information. In practice, therefore, semi-empirical models of turbulence tend to be used—often "eddy viscosities"—with no direct relation to the Navier-Stokes equations. In airflow past an airplane there is however typically only a one-inch layer on each surface where such issues are important; the large-scale features of the remainder of the flow, which nevertheless accounts for only about half the drag on the airplane, can usually be studied without reference to turbulence.

The Navier-Stokes equations assume that all speeds are small compared to the speed of sound—and thus that the Mach number giving the ratio of these speeds is much less than one. In essentially all practical situations, Mach numbers close to one occur only at extremely high Reynolds numbers—where turbulence in any case would make it impossible to work out the detailed consequences of the Navier-Stokes equations. Nevertheless, in the case of cellular automaton fluids, I was able in 1985 to work out the rather complicated next order corrections to the Navier-Stokes equations.

Above the speed of sound, fluids form shocks where density or velocity change over very small distances (see below). And by Mach 4 or so, shocks are typically so sharp that changes occur in less than the distance between molecular collisions—making it essential to go beyond the continuum fluid approximation, and account for molecular effects.

▪ **Models of turbulence.** Traditional models typically view turbulence as consisting of some form of cascade of eddies. This notion was already suggested in pictures by Leonardo da Vinci from around 1510, and in Japanese pictures (notably by Katsushika Hokusai) from around 1800 showing ocean waves breaking into precisely nested tongues of water. The theoretical study of turbulence began in earnest in the early 1900s, with emphasis on issues such as energy transfer among eddies and statistical correlations between velocities. Most published work became increasingly mathematical, but particularly following the ideas of Lewis Richardson in the 1920s, the underlying physical notion was that a large eddy, formed say by fluid flowing around an object, would be unstable, and would break up into smaller eddies, which in turn would break up into still smaller eddies, until eventually the eddies would be of such a size as to be readily damped by viscosity. An important step was taken in 1941 by Andrei Kolmogorov who argued that if the eddies in such a cascade were in a statistical equilibrium, then dimensional analysis would effectively imply that the spectrum of velocity fluctuations associated with the eddies must have a $k^{-5/3}$ distribution, with k being wavenumber. This result has turned out to be in respectable agreement with a range of experimental data, but its physical significance has remained somewhat unclear. For there appear to be no explicit entities in fluids that can be directly identified as cascades of eddies. One possibility might be that an eddy could correspond to a local patch of vorticity or rotation in the fluid. And it is a general feature of fluids that interfaces between regions of different velocity are unstable, typically first becoming wavy and then breaking into separate pieces. But physical experiments and simulations in the past few years have suggested that vorticity in turbulent fluids in practice tends to become concentrated on a complicated network of lines that stretch and twist. Perhaps some interpretation can be made involving eddies existing only in a fractal region, or interacting with each other as well as branching. And perhaps new forms of definite localized structures can be identified. But no clear understanding has yet emerged, and indeed most of the analysis that is done—which tends to be largely statistical in nature—is not likely to shed much light on the general question of why there is so much apparent randomness in turbulence.

▪ **Chaos theory and turbulence.** The full Navier-Stokes equations for fluid flow are far from being amenable to traditional mathematical analysis. But some simplified ordinary differential equations which potentially approximate various situations in fluid flow can be more amenable to analysis—and can exhibit the chaos phenomenon. Work in the 1950s by Lev Landau, Andrei Kolmogorov and others focused on equations with periodic

and quasiperiodic behavior. But in 1962 Edward Lorenz discovered more complicated behavior in computer experiments on equations related to fluid flow (see page 971). Analysis of this behavior was closely linked to the chaos phenomenon of sensitive dependence on initial conditions. And by the late 1970s it had become popular to believe that the randomness in fluid turbulence was somehow associated with this phenomenon.

Experiments in very restricted situations showed correspondence with iterated maps in which the chaos phenomenon is seen. But the details of the connection with true turbulence remained unclear. And as I argue in the main text, the chaos phenomenon in the end seems quite unlikely to explain most of the randomness we see in turbulence. The basic problem is that a complex pattern of flow in effect involves a huge amount of information—and to extract this information purely from initial conditions would require for example going to a submolecular level, far below where traditional models of fluids could possibly apply.

Even within the context of the Lorenz equations there are already indications of difficulties with the chaos explanation. The Lorenz equations represent a first-order approximation to certain Navier-Stokes-like equations, in which viscosity is ignored. And when one goes to higher orders progressively more account is taken of viscosity, but the chaos phenomenon becomes progressively weaker. I suspect that in the limit where viscosity is fully included most details of initial conditions will simply be damped out, as physical intuition suggests. Even within the Lorenz equations, however, one can see evidence of intrinsic randomness generation, in which randomness is produced without any need for randomness in initial conditions. And as it turns out I suspect that despite subsequent developments the original ideas of Andrei Kolmogorov about complicated behavior in ordinary differential equations were probably more in line with my notion of intrinsic randomness generation than with the chaos phenomenon.

▪ **Flows past objects.** By far the most experimental data has been collected for flows past cylinders. The few comparisons that have been done indicate that most results are extremely similar for plates and other non-streamlined or "bluff" objects. For spheres at infinitesimal Reynolds numbers a fairly simple exact analytical solution to the Navier-Stokes equations was found by George Stokes in 1851, giving a drag coefficient of $6\pi/R$. For a cylinder, there are difficulties with boundary conditions at infinity, but the drag coefficient was nevertheless calculated by William Oseen in 1915 to be $8\pi/(R(1/2 + \text{Log}[8/R] - \text{EulerGamma}))$. At infinitesimal Reynolds number the flow around a symmetrical object is always symmetrical. As the Reynolds number increases, it becomes progressively more asymmetrical, and at $R \simeq 6$ for a cylinder, closed eddies begin to appear behind the object. The length of the region associated with these eddies is found to grow almost perfectly linearly with Reynolds number. At $R \simeq 30-40$ for a cylinder, oscillations are often seen in the eddies, and at $R \simeq 46-49$, a vortex street forms. Increasingly accurate numerical calculations based on direct approximations to the Navier-Stokes equations have been done in the regime of attached eddies since the 1930s. For a vortex street no analytical solution has ever been found, and indeed it is only recently that the general paths of fluid elements have even been accurately deduced. A simple model due to Theodore von Kármán from 1911 predicts a relative spacing of $\pi/\text{Log}[1+\sqrt{2}]$ between vortices, and bifurcation theory analyses have provided some justification for some such result. Over the range $50 \lesssim R \lesssim 150$ vortices are found to be generated at a cylinder with almost perfect periodicity at a dimensionless frequency (Strouhal number) that increases smoothly from about 0.12 to 0.19. But even though successive vortices are formed at fixed intervals, irregularities can develop as the array of vortices goes downstream, and such irregularities seem to occur at lower Reynolds numbers for flows past plates than cylinders. Some direct calculations of interactions between vortices have been done in the context of the Navier-Stokes equations, but the cellular automaton approach of page 378 seems to provide essentially the first reliable global results. In both calculations and experiments, there is often sensitivity to details of whatever boundary conditions are imposed on the fluid, even if they are far from the object. Results can also be affected by the history of the flow. In general, the early way the flow develops over time typically mirrors quite precisely the long-time behavior seen at successively greater Reynolds numbers. In experiments, the process of vortex generation at a cylinder first becomes irregular somewhere between $R = 140$ and $R = 194$. After this surprisingly few qualitative changes are seen even up to Reynolds numbers as high as 100,000. There is overall periodicity much like in a vortex street, but the detailed motion of the fluid is increasingly random. Typically the scale of the smallest eddies gets smaller in rough correspondence with the $R^{-3/4}$ prediction of Kolmogorov's general arguments about turbulence. In flow past a cylinder, there are various quite sudden changes in the periodicity, apparently associated with 3D phenomena in which the flow is not uniform along the axis of the cylinder. The drag coefficient remains almost constant at a value around 1 until $R \simeq 3 \times 10^5$, at which point it drops precipitously for a while.

This phenomenon is associated with details of flow close to the cylinder. At lower Reynolds numbers, the flow is still laminar when it first comes around the cylinder; but there is a transition to turbulence in this boundary layer after which the fluid can in effect slide more easily around the cylinder. When the speed of the flow passes the speed of sound in the fluid, shocks appear. Usually they form simple geometrical patterns (see below), and have the effect of forcing the turbulent wake behind the cylinder to become narrower.

■ **2D fluids.** The cellular automaton shown in the main text is purely two-dimensional. Experiments done on soap films since the 1980s indicate, however, that at least up to Reynolds numbers of several hundred, the patterns of flow around objects such as cylinders are almost identical to those seen in ordinary 3D fluids. The basic argument for Kolmogorov's $k^{-5/3}$ result for the spectrum of turbulence is independent of dimension, but there are reasons to believe that in 2D eddies will tend to combine, so that after sufficiently long times only a small number of large eddies will be left. There is some evidence for this kind of process in the Earth's atmosphere, as well as in such phenomena as the Red Spot on Jupiter. At a microscopic level, there are some not completely unrelated issues in 2D about whether perturbations in a fluid made up of discrete molecules damp quickly enough to lead to ordinary viscosity. Formally, there is evidence that the Navier-Stokes equations in 2D might have a $\nabla^2 \text{Log}[\nabla^2]$ viscosity term, rather than a ∇^2 one. But this effect, even if it is in fact present in principle, is almost certainly irrelevant on the scales of practical experiments.

■ **Cellular automaton fluids.** A large number of technical issues can be studied in connection with cellular automaton fluids. Many were already discussed in my original 1985 paper. Others have been covered in some of the many papers that have appeared since then. Of particular concern are issues about how rotation and translation invariance emerge at the level of fluid processes even though they are absent in the underlying cellular automaton structure. The very simplest rules turn out to have difficulties in these regards (see page 1024), which is why the model shown in the main text, for example, is on a hexagonal rather than a square grid (compare page 980). The model can be viewed as a block cellular automaton of the type discussed on page 460, but on a 2D hexagonal grid. In general a block cellular automaton works by making replacements for overlapping blocks of cells on alternating steps. In the 1D case of page 460, the blocks that are replaced consist of pairs of adjacent cells with two different alignments. On a 2D square grid, one can use overlapping 2×2 square blocks. But on a 2D hexagonal grid, one must instead alternate on successive steps between hexagons and their dual triangles.

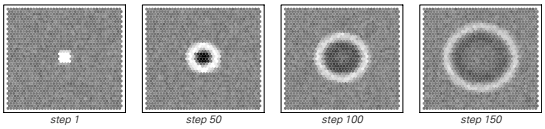
■ **Vorticity-based models.** As an alternative to models of fluids based on elements with discrete velocities, one can consider using elements with discrete vorticities.

■ **History of cellular automaton fluids.** Following the development of the molecular model for gases in the late 1800s (see page 1019), early mathematical derivations of continuum fluid behavior from underlying molecular dynamics were already complete by the 1920s. More streamlined approaches with the same basic assumptions continued to be developed over the next several decades. In the late 1950s Berni Alder and Thomas Wainwright began to do computer simulations of idealized molecular dynamics of 2D hard spheres—mainly to investigate transitions between solids, liquids and gases. In 1967 they observed so-called long-time tails not expected from existing calculations, and although it was realized that these were a consequence of fluid-like behavior not readily accounted for in purely microscopic approximations, it did not seem plausible that large-scale fluid phenomena could be investigated with molecular dynamics. The idea of setting up models with discrete approximations to the velocities of molecules appears to have arisen first in the work of James Broadwell in 1964 on the dynamics of rarefied gases. In the 1960s there was also interest in so-called lattice gases in which—by analogy with spin systems like the Ising model—discrete particles were placed in all possible configurations on a lattice subject to certain local constraints, and average equilibrium properties were computed. By the early 1970s more dynamic models were sometimes being considered, and for example Yves Pomeau and collaborators constructed idealized models of gases in which both positions and velocities of molecules were discrete. As it happens, in 1973, as one of my earliest computer programs, I created a simulation of essentially the same kind of system (see page 17). But it turned out that this particular kind of system, set up as it was on a square grid, was almost uniquely unable to generate the kind of randomness that we have seen so often in this book, and that is needed to obtain standard large-scale fluid behavior. And as a result, essentially no further development on discrete models of fluids was then done until after my work on cellular automata in the early 1980s. I had always viewed turbulent fluids as an important potential application for cellular automata. And in 1984, as part of work I was doing on massively parallel computing, I resolved to develop a practical approach to fluid mechanics based on cellular automata. I initiated discussions with

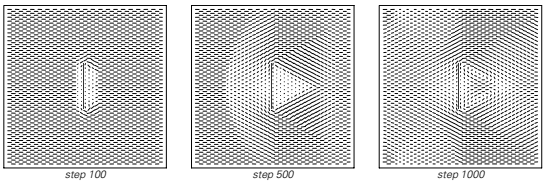
various members of the fluid dynamics community, who strongly discouraged me from pursuing my ideas. But I persisted, and by the summer of 1985 I had managed to produce pictures like those on page 378. Meanwhile, however, some of the very same individuals who had discouraged me had in fact themselves pursued exactly the line of research I had discussed. And by late 1985, cellular automaton fluids were generating considerable interest throughout the fluid mechanics community. Many claims were made that existing computational methods were necessarily far superior. But in practice over the years since 1985, cellular automaton methods have grown steadily in popularity, and are now widely used in physics and engineering. Yet despite all the work that has been done, the fundamental issues about the origins of turbulence that I had originally planned to investigate in cellular automaton fluids have remained largely untouched.

▪ **Computational fluid dynamics.** From its inception in the mid-1940s until the invention of cellular automaton fluids in the 1980s, essentially all computational fluid dynamics involved taking the continuum Navier-Stokes equations and then approximating these equations using some form of discrete mesh in space and time, and arguing that when the mesh becomes small enough, correct results would be obtained. Cellular automaton fluids start from a fundamentally discrete system which can be simulated precisely, and thus avoid the need for any such arguments. One issue however is that in the simplest cellular automaton fluids molecules are in effect counted in unary: each molecule is traced separately, rather than just being included as part of a total number that can be manipulated using standard arithmetic operations. A variety of tricks, however, maintain precision while in effect allowing a large number of molecules to be handled at the same time.

▪ **Sound waves and shocks.** Sound waves in a fluid correspond to periodic variations in density. The pictures below show how a density perturbation leads to a sound wave in a cellular automaton fluid. The sound wave turns out to travel at a fraction $1/\sqrt{2}$ of the microscopic particle speed.



When the speed of a fluid relative to an object becomes comparable to the speed of sound, the fluid will inevitably show variations in density. Typically shocks develop at the front and back of an object, as illustrated below.



It turns out that when two shocks meet, they usually have little effect on each other, and when there are boundaries, shocks are usually reflected in simple ways. The result of this is that in most situations patterns of shocks generated have a fairly simple geometrical structure, with none of the randomness of turbulence.

▪ **Splashes.** Particularly familiar everyday examples of complex fluid behavior are splashes made by objects falling into water. When a water drop hits a water surface, at first a symmetrical crater forms. But soon its rim becomes unstable, and several peaks (often with small drops at the top) appear in a characteristic coronet pattern. If the original drop was moving quickly, a whole hemisphere of water then closes in above. But in any case a peak appears at the center, sometimes with a spherical drop at the top. If a solid object is dropped into water, the overall structure of the splash made can depend in great detail on its shape and surface roughness. Splashes were studied using flash photography by Arthur Worthington around 1900 (as well as Harold Edgerton in the 1950s), but remarkably little theoretical investigation of them has ever been made.

▪ **Generalizations of fluid flow.** In the simplest case the local state of a fluid is characterized by its velocity and perhaps density. But there are many situations where there are also other quantities relevant, notably temperature and chemical composition. And it turns out to be rather straightforward to generalize cellular automaton fluids to handle these.

▪ **Convection.** When there is a temperature difference between the top and bottom of a fluid, hot fluid tends to rise, and cold fluid then comes down again. At low temperature differences (characterized by a low dimensionless Rayleigh number) a regular pattern of hexagonal Bénard convection cells is formed (see page 377). But as the temperature difference increases, a transition to turbulence is seen, with most of the same characteristics as in flow past an object. A cellular automaton model can be made by allowing particles with more than one possible energy: the average particle energy in a region corresponds to fluid temperature.

▪ **Atmospheric turbulence.** Convection occurs because air near the ground is warmer than air at higher altitudes. On a clear night over flat terrain, air flow can be laminar near the ground. Usually, however, it is turbulent near the ground—producing, for example, random gusting in wind—but becomes laminar at higher altitudes. Turbulent convection nevertheless occurs in most clouds, leading to random billowing shapes. The "turbulence" that causes bumps in airplanes is often associated with clouds, though sometimes with larger-scale wave-like fluid motions such as the jet stream.

▪ **Ocean surfaces.** At low wind speeds, regular ripples are seen; at higher wind speeds, a random pattern of creases occurs. It seems likely that randomness in the wind has little to do with the behavior of the ocean surface; instead it is the intrinsic dynamics of the water that is most important.

▪ **Granular materials.** Sand and other granular materials show many phenomena seen in fluids. (Sand dunes are the rough analog of ocean waves.) Vortices have recently been seen, and presumably under appropriate conditions turbulence will eventually also be seen.

▪ **Geological structures.** Typical landscapes on Earth are to a first approximation formed by regions of crust being uplifted through tectonic activity, then being sculpted by progressive erosion (and redeposition of sediment) associated with the flow of water. (Visually very different special cases include volcanos, impact craters and wind-sculpted deserts.) Eventually erosion and deposition will in effect completely smooth out a landscape. But at intermediate times one will see all sorts of potentially dramatic gullies that reflect the pattern of drainage, and the formation of a whole tree of streams and rivers. (Such trees have been studied since at least the early 1900s, with typical examples of concepts being Horton stream order, equal to *Depth* for trees given as *Mathematica* expressions.) If one imagines a uniform slope with discrete streams of water going randomly in each direction at the top, and then merging whenever they meet, one immediately gets a simple tree structure a little like in the pictures at the top of page 359. (More complicated models based for example on aggregation, percolation and energy minimization have been proposed in recent years—and perhaps because most random spanning trees are similar, they tend to give similar results.) As emphasized by Benoit Mandelbrot in the 1970s and 1980s, topography and contour lines (notably coastlines) seem to show apparently random structure on a wide range of scales—with definite power laws being measured in quite a few cases. And presumably at some level this is the result of the nested patterns in which erosion occurs. (An unrelated effect is that as a result of the dynamics of flow in it, even a single river on a featureless landscape will typically tend to increase the curvature of its meanders, until they break off and form oxbow lakes.)

Fundamental Issues in Biology

▪ **Page 383 · History.** The origins of biological complexity have been debated since antiquity. For a long time it was assumed that the magnitude of the complexity was so great that it could never have arisen from any ordinary natural process, and therefore must have been inserted from outside through some kind of divine plan. However, with the publication of Charles Darwin's *Origin of Species* in 1859 it became clear that there were natural processes that could in fact shape features of biological organisms. There was no specific argument for why natural selection should lead to the development of complexity, although Darwin appears to have believed that this would emerge somewhat like a principle in physics. In the century or so after the publication of *Origin of Species* many detailed aspects of natural selection were elucidated, but the increasing use of traditional mathematical methods largely precluded serious analysis of complexity. Continuing controversy about contradictions with religious accounts of creation caused most scientists to be adamant in assuming that every aspect of biological systems must be shaped purely by natural selection. And by the 1980s natural selection had become firmly enshrined as a force of practically unbounded power, assumed—though without specific evidence—to be capable of solving almost any problem and producing almost any degree of complexity.

My own work on cellular automata in the early 1980s showed that great complexity could be generated just from simple programs, without any process like natural selection. But although I and others believed that my results should be relevant to biological systems there was still a pervasive belief that the level of complexity seen in biology must somehow be uniquely associated with natural selection. In the late 1980s the study of artificial life caused several detailed computer simulations of natural selection to be done, and these simulations reproduced various known features of biological evolution. But from looking at such simulations, as well as from my own experiments done from 1980 onwards, I increasingly came to believe that almost any complexity being generated had its origin in phenomena similar to those I had seen in cellular automata—and had essentially nothing to do with natural selection.

▪ **Attitudes of biologists.** Over the years, I have discussed versions of the ideas in this section with many biologists of different kinds. Most are quick to point out at least anecdotal

cases in which features of organisms do not seem to have been shaped by natural selection. But if asked about complexity—either in specific examples or in general—the vast majority soon end up trying to give explanations based on natural selection. Those with a historical bent often recognize that the origins of complexity have always been somewhat mysterious in biology, and indeed sometimes state that this has laid the field open to many attacks. But generally my experience has been that the further one goes from those involved with specific molecular or other details of biological systems the more one encounters a fundamental conviction that natural selection must be the ultimate origin of any important feature of biological systems.

■ **Page 383 · Genetic programs.** Genetic programs are encoded as sequences of four possible nucleotide bases on strands of DNA or RNA. The simplest known viruses have programs that are a few thousand elements in length; bacteria typically have programs that are a few million elements; fruit flies a few hundred million; and humans around four billion. There is not a uniform correspondence between apparent sophistication of organisms and lengths of genetic programs: different species of amphibians, for example, have programs that can differ in length by a factor of a hundred, and can be as many as tens of billions of elements long. Genetic programs are normally broken into sections, many of which are genes that provide templates for making particular proteins. In humans, there are perhaps around 40,000 genes, specifying proteins for about 200 distinct cell types. Many of the low-level details of how proteins are produced is now known, but higher-level issues about organization into different cell types remain somewhat mysterious. Note that although most of the information necessary to construct an organism is encoded in its genetic program, other material in the original egg cell or the environment before birth can probably also sometimes be relevant.

■ **Page 386 · Tricks in evolution.** Among the tricks used are: sexual reproduction, causing large-scale mixing of similar programs; organs, suborganisms, symbiosis and parasitism, allowing different parts of programs to be optimized separately; mutation rate enzymes, allowing parts that need change to be searched more quickly; learnability in individual organisms, allowing larger local deviations from optimality to be tried.

■ **Page 387 · Belief in optimality.** The notion that features of biological organisms are always somehow optimized for a particular purpose has become extremely deep seated—and indeed it has been discussed since antiquity. Most modern biologists at least pay lip service to historical accidents and developmental constraints, but if pressed revert surprisingly quickly to the notion of optimization for a purpose.

■ **Page 390 · Studying natural selection.** From the basic description of natural selection one might have thought that it would correspond to a unique simple program. But in fact there are always many somewhat arbitrary details, particularly centering around exactly how to prune less fit organisms. And the consequence of this is that in my experience it is essentially impossible to come up with precise definitive conclusions about natural selection on the basis of specific simple computer experiments. Using the Principle of Computational Equivalence discussed in Chapter 12, however, I suspect that it will nevertheless be possible to develop a general theory of what natural selection typically can and cannot do.

■ **Page 391 · Other models.** Sequential substitution systems are probably more realistic than cellular automata as models of genetic programs, since elements can explicitly be added to their rules at will. As a rather different approach, one can consider a fixed underlying rule—say a class 4 cellular automaton—with modifications in initial conditions. The notion of universality in Chapter 11 implies that under suitable conditions this should be equivalent to modifications in rules. As an alternative to modelling individual organisms, one can also consider substitution systems which directly generate genealogical trees for populations of organisms, somewhat like Leonardo Fibonacci's original model of a rabbit population.

■ **Page 391 · Adaptive value of complexity.** One might think that the reason complexity is not more widespread in biology is that somehow it is too sensitive to perturbations. But in fact, as discussed in Chapter 7, randomness and complexity tend to lead to more, rather than less, robustness in overall behavior. Indeed, many even seemingly simple biological processes appear to be stabilized by randomness—leading, for example, to random fluctuations in interbeat intervals for healthy hearts. And some biological processes rely directly on complex or random phenomena—for example, finding good paths for foraging for food, avoiding predators or mounting suitable immune responses. (Compare page 1192.)

■ **Page 393 · Genetic algorithms.** As mentioned on page 985, it is straightforward to apply natural selection to computer programs, and for certain kinds of practical tasks with appropriate continuity properties this may be a useful approach.

■ **Page 394 · Smooth variables.** Despite their importance in understanding natural selection both in biology and in potential computational applications, the fundamental

origins of smooth variables or so-called quantitative traits seem to have been investigated rather little. Within populations of organisms such traits are often found to have Gaussian distributions (as, for example, in heights of humans), but this gives little clue as to their origin. (Weights of humans nevertheless have closer to a lognormal distribution.) It is generally assumed that smooth variables must be associated with so-called polygenes that effectively include a large number of individual discrete genes. In pre-Mendelian genetics, observations on smooth variables are presumably what led to the theory that traits of offspring are determined by smoothly mixing the blood of their parents.

▪ **Page 395 · Species.** One feature of biological organisms is that they normally occur in discrete species, with distinct differences between different species. It seems likely that the existence of such discreteness is related to the discreteness of underlying genetic programs. Currently there are a few million species known. Most are distinguished just by their habitats, visual appearance or various simple numerical characteristics. Sometimes, however, it is known that members of different species have the traditional defining characteristic that they cannot normally mate, though this may well be more a matter of the mechanics of mating and development than a fundamental feature.

▪ **Defining life.** See pages 823 and 1178.

▪ **Page 397 · Analogies with thermodynamics.** Over the past century there have been a number of attempts to connect the development of complexity in biological systems with the increase of entropy in thermodynamic systems. In fact, when it was first introduced the very term "entropy" was supposed to suggest an analogy with biological evolution. But despite this, no detailed correspondence between thermodynamics and evolution has ever been forthcoming. However, my statement here that complexity in biology can occur because natural selection cannot control complex behavior is rather similar to my statement in Chapter 9 that entropy can increase because physical experiments in a sense also cannot control complex behavior.

▪ **Page 398 · Major new features.** Traditional groupings of living organisms into kingdoms and phyla are typically defined by the presence of major new features. Standard examples from higher animals include regulation of body temperature and internal gestation of young. Important examples from earlier in the history of life include nuclear membranes, sexual reproduction, multicellularity, protective shells and photosynthesis.

Trilobites are a fairly clear example of organisms where over the course of a few hundred million years the fossil record shows increases in apparent morphological complexity, followed by decreases. Something similar can be seen in the historical evolution of technological systems such as cars.

▪ **Software statistics.** Empirical analysis of the million or so lines of source code that make up *Mathematica* suggests that different functions—which are roughly analogous to different genes—rather accurately follow an exponential distribution of sizes, with a slightly elevated tail.

▪ **Proteins.** At a molecular level much of any living cell is made up of proteins formed from chains of tens to thousands of amino acids. Of the thousands of proteins now known some (like keratin and collagen) are fibrous, and have a simple repetitive underlying structure. But many are globular, and have at least a core in which the 3D packing of amino acids seems quite random. Usually there are some sections that consist of simple α helices, β sheets, or combinations of these. But other parts—often including sites important for function—seem more like random walks. At some level the 3D shapes of proteins (tertiary structure) are presumably determined by energy minimization. But in practice very different shapes can probably have almost identical energies, so that in as much as a given protein always takes on the same shape this must be associated with the dynamics of the process by which the protein folds when it is assembled. (Compare page 988.) One might expect that biological evolution would have had obvious effects on proteins. But as mentioned on page 1184 the actual sequences of amino acids in proteins typically appear quite random. And at some level this is presumably why there seems to be so much randomness in their shapes. (Biological evolution may conceivably have selected for proteins that fold reliably or are more robust with respect to changes in single amino acids, but there is currently no clear evidence for this.)

Growth of Plants and Animals

▪ **History.** The first steps towards a theory of biological form were already taken in Greek times with attempts—notably by Aristotle—to classify biological organisms and to understand their growth. By the 1600s extensive classification had been done, and many structural features had been identified as in common between different organisms. But despite hopes on the part of René Descartes, Galileo and others that biological processes might follow the same kind of rigid clockwork rules that were beginning to emerge in physics, no general principles were forthcoming. Rough analogies between the forms and functions of biological and non-biological systems were fairly common among both artists and scientists, but were rarely thought to

have much scientific significance. In the 1800s more detailed analogies began to emerge, sometimes as offshoots of the field of morphology named by Johann Goethe, and sometimes with mathematical interpretations, and in 1917 D'Arcy Thompson published the first edition of his book *On Growth and Form* which used mathematical methods—mostly from analytical geometry—to discuss a variety of biological processes, usually in analogy with ones in physics. But emphasis on evolutionary rather than mechanistic explanations for a long time caused little further work to be done along these lines. Much additional data was obtained, particularly in embryology, and by the 1930s it seemed fairly clear that at least some aspects of growth in the embryo were controlled by chemical messengers. In 1951 Alan Turing worked out a general mathematical model of this based on reaction-diffusion equations, and suggested that such a model might account for many pigmentation and structural patterns in biological systems (see page 1012). For nearly twenty years, however, no significant follow-up was done on this idea. There were quite a few attempts—often misguided in my opinion—to use traditional ideas from physics and engineering to derive forms of biological organisms from constraints of mechanical or other optimality. And in the late 1960s, René Thom made an important attempt to use sophisticated methods from topology to develop a general theory of biological form. But the mathematics of his work was inaccessible to most natural scientists, and its popularized version, known as catastrophe theory, largely fell into disrepute.

The idea of comparing systems in biology and engineering dates back to antiquity, but for a long time it was mainly thought of just as an inspiration for engineering. In the mid-1940s, however, mostly under the banner of cybernetics, tools from the analysis of electrical systems began to be used for studying biological systems. And partly from this—with much reinforcement from the discovery of the genetic code—there emerged the idea of thinking about biological systems in purely abstract logical or computational terms. This led to an early introduction of 2D cellular automata (see page 876), but the emphasis was on ambitious general questions rather than specific models. Little progress was made, and by the 1960s most work along these lines had petered out. In the late 1970s, however, fractals and L systems (see below) began to provide examples where simple rules could be seen to yield biological-like branching behavior. And in the 1980s, interest in non-equilibrium physical processes, and in phenomena such as diffusion-limited aggregation, led to renewed interest in reaction-diffusion equations, and to somewhat more explicit models for various biological processes. My own work on cellular automata in the early 1980s started a number of new lines of computational modelling, some of which became involved in the rise and fall of the artificial life movement in the late 1980s and early 1990s.

■ **Page 400 · Growth in plants.** At the lowest level, the growth of any organism proceeds by either division or expansion of cells, together with occasional formation of cavities between cells. In plants, cells typically expand—normally through intake of water—only for a limited period, after which the cellulose in their walls crystallizes to make them quite rigid. In most plants—at least after the embryonic stage—cells typically divide only in localized regions known as meristems, and each division yields one cell that can divide again, and one that cannot. Often the very tip of a stem consists of a single cell in the shape of an inverted tetrahedron, and in lower plants such as mosses this is essentially the only cell that divides. In flowering plants, cell division normally occurs around the edge of a region of size 0.2–1 mm containing many tens of cells. (Hearts of palm in palm trees can however be much larger.) The details of how cell division works in plants remain largely unknown. There is some evidence that orientation of new cells is in part controlled by microscopic fibers. Various small molecules that can diffuse between cells (such as so-called auxins) are known to affect growth and production of new stems (see below).

■ **Page 401 · Branching in plants.** Almost all kinds of plants exhibit some form of branching, and particularly in smaller plants the branching is often extremely regular. In a plant as large as a typical tree—particularly one that grows slowly—different conditions associated with the growth of different branches may however destroy some of the regularity of branching. Among algae and more primitive plants such as whisk ferns, repeated splitting of a single branch into two is particularly common. Ferns and conifers both typically exhibit three-way branching. Among flowering plants so-called dicotyledons exhibit branching throughout the plant. Monocotyledons—of which palms and grasses are two examples—typically have only one primary site of growth, and thus do not exhibit repeated branching. (In grasses the growth site is at the bottom of the stem, and in bamboos there are multiple growth sites up the stem.)

The forms of branching in plants have been used as means of classification since antiquity. Alexander von Humboldt in 1808 identified 19 overall types of branching which have been used, with some modifications, by plant geographers and botanists ever since. Note that in the vast majority of

cases, branches do not lie in a plane; often they are instead arranged in a spiral, as discussed on page 408. But when projected into two dimensions, the patterns obtained still look similar to those in the main text.

▪ **Page 402 · Implementation.** It is convenient to represent the positions of all tips by complex numbers. One can take the original stem to extend from the point -1 to 0; the rule is then specified by the list *b* of complex numbers corresponding to the positions of the new tip obtained after one step. And after *n* steps the positions of all tips generated are given simply by

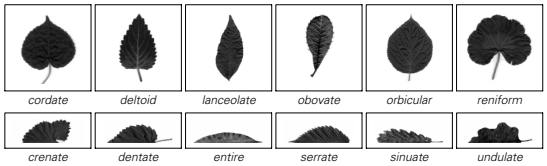
```
Nest[Flatten[Outer[Times, 1 + #, b]] &, {0}, n]
```

▪ **Mathematical properties.** If an element *c* of the list *b* is real, so that there is a stem that goes straight up, then the limiting height of the center of the pattern is obtained by summing a geometric series, and is given by *1/(1 - c)*. The overall limiting pattern will be finite so long as *Abs[c] < 1* for all elements of *b*. After *n* steps the total length of all stems is given by $Apply[Plus, Abs[b]]^n$. (See page 1006 for other properties.)

▪ **Page 402 · Simple geometries.** Page 357 shows how some of the nested patterns commonly seen in this book can be produced by the growth processes shown here.

▪ **History of branching models.** The concept of systematic rules for the way that stems—particularly those carrying flowers—are connected in a plant seems to have been clearly understood among botanists by the 1800s. Only with the advent of computer graphics in the 1970s, however, does the idea appear to have arisen of varying angles to get different forms. An early example was the work of Hisao Honda in 1970 on the structure of trees. Pictures analogous to the bottom row on page 402 were also generated by Benoit Mandelbrot in connection with his development of fractals. Starting in 1967 Aristid Lindenmayer emphasized the use of substitution or L systems (see page 893) as a way of modelling patterns of connections in plants. And beginning in the early 1980s—particularly through work by Alvy Ray Smith and later Przemyslaw Prusinkiewicz—models based on L systems and fractals became routinely used for producing images of plants in practical computer graphics. Around the same time Michael Barnsley also used so-called iterated function systems to make pictures of ferns—but he appears to have viewed these more as a curiosity than a contribution to botany. Over the past decade or so, a few mentions have been made of using complicated models based on L systems to reproduce shapes of specific types of leaves, but so far as I can tell, nothing like the simple model that I describe in the main text has ever been considered before.

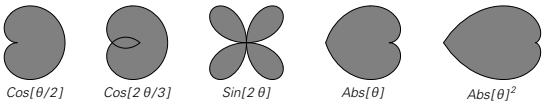
▪ **Page 404 · Leaf shapes.** Leaves are usually put into categories like the ones below, with names mostly derived from Latin words for similar-looking objects.



Some classification of leaf shapes was done by Theophrastus as early as 300 BC, and classifications similar to those above were in use by the early Renaissance period. (They appear for example in the first edition of the *Encyclopedia Britannica* from 1768.) Leaf shapes have been widely used since antiquity as a way of identifying plants—initially particularly for medicinal purposes. But there has been very little general scientific investigation of leaf shapes, and most of what has been done has concentrated on the expansion of leaves once they are out of their buds. Already in 1724 Stephen Hales looked at the motion of grids of marks on fig leaves, and noted that growth seemed to occur more or less uniformly throughout the leaf. Similar but increasingly quantitative studies have been made ever since, and have reported a variety of non-uniformities in growth. For a long time it was believed that after leaves came out of their buds growth was due mainly to cell expansion, but in the 1980s it became clear that many cell divisions in fact occur, both on the boundary and the interior. At the earliest stages, buds that will turn into leaves start as bumps on a plant stem, with a structure that is essentially impossible to discern. Surgically modifying such buds when they are as small as 0.1 mm can have dramatic effects on final leaf shape, suggesting that at least some aspects of the shape are already determined at that point. On a single plant different leaves can have somewhat different shapes—sometimes for example those lower on a tree are smoother, while those higher are pointier. It may nevertheless be that leaves on a single plant initially have a discrete set of possible shapes, with variations in final shape arising from differences in environmental conditions during expansion. My model for leaf shapes is presumably most relevant for initial shapes.

Traditional evolutionary explanations have not had much to say about detailed questions of leaf shape; one minor claim is that the pointed tips at the ends of many tropical leaves exist to allow moisture to drip off the leaves. The fossil record suggests that leaves first arose roughly 400 million years ago, probably when collections of branches which lay in a plane became joined by webbing. Early plants such as ferns have compound leaves in which explicit branching structure is still

seen. Extremely few models for shapes of individual leaves appear to have ever been proposed. In 1917 D'Arcy Thompson mentioned that leaves might have growth rates that are simple functions of angle, and drew the first of the pictures shown below.



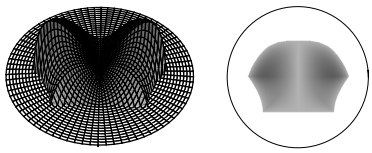
With new tip positions as on page 400 given by *{p Exp[i θ], p Exp[-i θ], q}*, rough *{p, q, θ}* for at least some versions of some common plants include: wild carrot (Queen Anne's lace) *{0.4, 0.7, 30°}*, cypress *{0.4, 0.7, 45°}*, coralbells *{0.5, 0.4, 0°}*, ivy *{0.5, 0.6, 0°}*, grape *{0.5, 0.6, 15°}*, sycamore *{0.5, 0.6, 15°}*, mallow *{0.5, 0.6, 30°}*, goosefoot *{0.55, 0.8, 30°}*, willow *{0.55, 0.8, 80°}*, morning glory *{0.7, 0.8, 0°}*, cucumber *{0.7, 0.8, 15°}*, ginger *{0.65, 0.6, 15°}*.

■ **Page 404 · Self-limiting growth.** It is often said that in plants, unlike animals, there is no global control of growth. And one feature of the simple branching processes I describe is that for purely mathematical reasons, their rules always produce structures that are of limited size. Note that in fact it is known that there is some global control of growth even in plants: for example hormones produced by leaves can affect growth of roots.

■ **Page 407 · Parameter space sets.** Points in the space of parameters can conveniently be labelled by a complex number *c*, where the imaginary direction is taken to increase to the right. The pattern corresponding to each point is the limit of *Nest[Flatten[1 + {c #, Conjugate[c] #}] &, {1}, n]* when $n \to \infty$. Such a limiting pattern exists only within the unit circle *Abs[c] < 1*. It then turns out that the limiting pattern is either completely connected or completely disconnected; which it is depends on whether it contains any points on the vertical axis *Im[c] == 0*. Every point in the pattern must correspond to some list of left and right branchings, represented by 0's and 1's respectively; in terms of this list the position of the point is given by *Fold[1 + {c, Conjugate[c]}[[1 + #2]] #1 &, 1, Reverse[list]]*. Patterns are disconnected if there is a gap between the parts obtained from lists starting with 0 and with 1. The magnitude of this gap turns out to be given by

```
With[{d = Conjugate[c], r = 1 - Abs[c]^2},
  Which[Im[c] < 0, d, Im[c] == 0, 0,
    Re[c] > 0, With[{n = Ceiling[π/2/Arg[c]]},
      Im[c (1 - d^n)/(1 - d)] + Im[c d^n (1 + d)]/r], Arg[c] >
      3 π/4, Im[c + c^2]/r, True, Im[c] + Im[c^2 + c^3]/r]]
```

The picture below shows the region for which the gap is positive, corresponding to trees which are not connected. (This region was found by Michael Barnsley and others in the late 1980s.) The overall maximum gap occurs at *c = 1/2 Sqrt[5 - √17] i*. The bottom boundary of the region lies along *Re[c] = -1/2*; the extremal point on the edge of the gap in this case corresponds to *{0, 0, 1, 0, 1, 0, 1, …}* where the last two elements repeat forever. The rest of the boundary consists of a sequence of algebraic curves, with almost imperceptible changes in slope in between; the first corresponds to *{0, 0, 0, 1, 0, 1, 0, 1, …}*, while subsequent ones correspond to *{0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, …}*, *{0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, …}*, etc.



In the pictures in the main text, the black region is connected wherever it does not protrude into the shaded region, which corresponds to disconnected patterns, in the pictures above. And in general it turns out that near any particular value of *c* the sets shown in black in the main text always look at sufficient magnification like the pattern that would be obtained for that value of *c*. The reason for this is that if *c* changes only slightly, then the pattern to a first approximation deforms only slightly, so that the part seen through the peephole just shifts, and in a small region of *c* values the peephole in effect simply scans over different parts of the pattern.

A simple way to approximate the pictures in the main text would be to generate patterns by iterating the substitution system a fixed number of times. In practice, however, it is essential to prune the tree of points at each stage. And at least for *Abs[c]* not too close to 1, this can be done by discarding points that are so far away from the peephole that their descendents could not possibly return to it.

The parameter space sets discussed here are somewhat analogous to the Mandelbrot set discussed on page 934, though in many ways easier to understand.

(See also the discussion of universal objects on page 1127.)

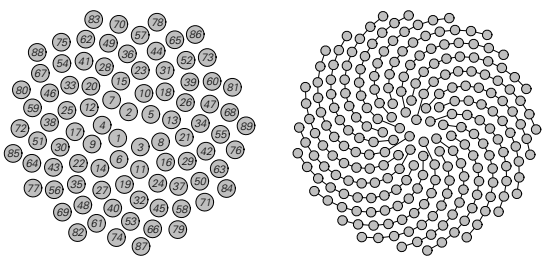
■ **Page 409 · Mathematics of phyllotaxis.** A rotation by *GoldenRatio == (1 + √5)/2* turns is equivalent to a rotation by *2 - GoldenRatio == GoldenRatio^-2* ≃ *0.38* turns, or 137.5°. Successive approximations to this number are given by *Fibonacci[n - 2]/Fibonacci[n]*, so that elements numbered *Fibonacci[n]* (i.e. 1, 2, 3, 5, 8, 13, …) will be the ones that come closest to being a whole number of turns apart, and thus to being lined up on the stem. As mentioned on page 891, having *GoldenRatio* turns between elements makes them in a

sense as evenly distributed as possible, given that they are added sequentially.

■ **History of phyllotaxis.** The regularities of phyllotaxis were presumably noticed in antiquity, and were certainly recognized in the 1400s, notably by Leonardo da Vinci. By the 1800s various mathematical features of phyllotaxis were known, and in 1837 Louis and Auguste Bravais identified the presence of a golden ratio angle. In 1868 Wilhelm Hofmeister proposed that new elements form in the largest gap left by previous elements. And in 1913 Johannes Schoute argued that diffusion of a chemical creates fields of inhibition around new elements—a model in outline equivalent to mine. In the past century features of phyllotaxis have been rediscovered surprisingly many times, with work being done quite independently both in abstract mathematical settings, and in the context of specific models (most of which are ultimately very similar). One development in the 1990s is the generation of phyllotaxis-like patterns in superconductors, ferrofluids and other physical systems.

■ **Observed phyllotaxis.** Many spiral patterns in actual plants converge to within a degree or less of 137.5 °, though just as in the model in the main text, there are usually deviations for the first few elements produced. The angles are particularly accurate in, for example, flower heads—where it is likely the positions of elements are adjusted by mechanical forces after they are originally generated. Other examples of phyllotaxis-like patterns in biology include the scales of pangolins and surfaces of tooth-like structures in certain kinds of rays and sharks.

■ **Projections of patterns.** The literature of phyllotaxis is full of baroque descriptions of the features of projections of patterns with golden ratio angles. In the pictures below, the n^{th} point has position *(√n {Sin[#], Cos[#]} &)[2 π n GoldenRatio]*, and in such pictures regular spirals or parastichies emanating from the center are seen whenever points whose numbers differ by *Fibonacci[m]* are joined. Note that the tips of many growing stems seem to be approximately paraboloidal, making the n^{th} point a distance $\sqrt{n}$ from the center.



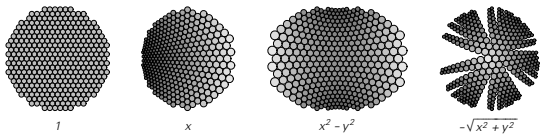
■ **Page 410 · Implementation.** It is convenient to consider a line of discrete cells, much as in a continuous cellular automaton. With a concentrations list *c*, the position *p* of a new element is given by *Position[c, Max[c], 1, 1][[1, 1]]*, while the new list of concentrations is *λ c + RotateRight[f, p]* where *f* is a list of depletions associated with addition of a new element at position 1. In the main text a Gaussian form is used for *f*. Other smooth functions typically nevertheless yield identical results. Note that in order to get an accurate approximation to a golden ratio angle there must be a fairly large number of cells.

■ **Shapes of cells.** Many types of cells are arranged like typical 3D packings of deformable objects (see page 988)—with considerable apparent randomness in individual shapes and positions, but definite overall statistical properties. Cells arranged on a surface—as in the retina or in skin—or that are intrinsically elongated—as in muscle—tend again to be arranged like typical packings, but now in 2D, where a regular hexagonal grid is formed.

■ **Page 412 · Symmetries.** Biological systems often show definite discrete symmetry. (In monocotyledon plants there is usually 3-fold symmetry; in dicotyledons 4- or 5-fold. Animals like starfish often have 5-fold symmetry; higher animals usually only 2-fold symmetry. There are fossils with 7- and 9-fold symmetry. At microscopic levels there are sometimes other symmetries: cilia of eukaryotic cells can for example show 9- and 13-fold symmetry. In the phyllotaxis process discussed in the main text one new element is produced at a time. But if several elements are produced together the same basic mechanism will tend to make these elements be equally spaced in angle—leading to overall discrete symmetry. (Individual proteins sometimes also arrange themselves into overall structures that have discrete symmetries—which can then be reflected in shapes of cells or larger objects.) (See also page 1011.)

■ **Page 412 · Locally isotropic growth.** A convenient way to see what happens if elements of a surface grow isotropically is to divide the surface into a collection of very small circles, and then to expand the circle at each point by a factor *h[x, y]*. If the local curvature of the surface is originally *c[x, y]*, then after such growth, the curvature turns out to be *(c[x, y] + Laplacian[Log[h[x, y]]])/h[x, y]* where *Laplacian[f_] := ∂xx f + ∂yy f*. In order for the surface to stay flat its growth rate *Log[h[x, y]]* must therefore solve Laplace's equation, and hence must be a harmonic function *Re[f[x + i y]]*. This is equivalent to saying that the growth must correspond to a conformal mapping which locally preserves angles. The pictures below show results for several growth rate functions; in the last case, the function

is not harmonic, and the surface cannot be drawn in the plane without tearing. Note that if the elements of a surface are allowed to change shape, then the surface can always remain flat, as in the top row of pictures on page 412. Harmonic growth rate functions can potentially be obtained from the large-time effects of a chemical subject to diffusion. And this may perhaps be related to the flatness observed in the growth of leaves. (See also page 1010.)



■ **Page 413 · Branching in animals.** Capillaries, bronchioles and kidney ducts in higher animals typically seem to form trees in which each tip as it extends repeatedly splits into two branches. (In human lungs, for example, there are about 20 levels of branching.) The same kind of structure is seen in the digestive systems of lower animals—as visible externally, for example, in the arms of a basket star.

■ **Page 413 · Antlers.** Like stems of plants antlers grow at their tips, and can thus exhibit branching. This is made possible by the fact that antlers, unlike horns, have a layer of soft tissue on the outside—which delivers the nutrients needed for growth to occur on the outer surface of the bone at their tips.

■ **Page 414 · Shells.** Shells grow through the secretion of rigid material from the soft lip or mantle of the animal inside, and over periods of months to years they form coiled structures that normally follow rather accurate equiangular spirals, typically right-handed. The number of turns or whorls varies widely, from less than one in a typical bivalve, to more than thirty in a highly pointed univalve such as a screw shell. Usually the coiled structure is obvious from looking at the apex on the outside of the shell, but in cowries, for example, it is made less obvious by the fact that later whorls completely cover earlier ones, and at the opening of the shell some dissolving and resculpting of material occurs. In addition to smooth coiled overall structures, some shells exhibit spines. These are associated with tentacles of tissue which secrete shell material at their tips as they grow. Inside shells such as nautiluses, there are a sequence of sealed chambers, with septa between them laid down perhaps once a month. These septa in present-day species are smooth, but in fossil ammonites they can be highly corrugated. Typically the corrugations are accurately symmetrical, and I suspect that they in effect represent slices through a lettuce-leaf-like structure formed from a surface with tree-like internal growth.

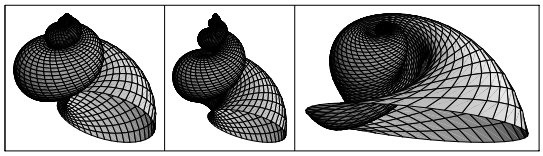
■ **Shell model.** The center of the opening of a shell is taken to trace out a helix whose $\{x, y, z\}$ coordinates are given as a function of the total angle of revolution t by $a^t \{\text{Cos}[t], \text{Sin}[t], b\}$. On row (a) of page 415 the parameter a varies from 1.05 to 1.65, while on row (b) b varies from 0 to 6. The complete surface of the shell is obtained by varying both t and θ in

$$a^t \{\text{Cos}[t] (1 + c (\text{Cos}[e] \text{Cos}[\theta] + d \text{Sin}[e] \text{Sin}[\theta])),\\ \text{Sin}[t] (1 + c (\text{Cos}[e] \text{Cos}[\theta] + d \text{Sin}[e] \text{Sin}[\theta])),\\ b + c (\text{Cos}[\theta] \text{Sin}[e] - d \text{Cos}[e] \text{Sin}[\theta])\}$$

where c varies from 0.4 to 1.6 on row (c), d from 1 to 4 on row (d) and e from 0 to 1.2 on row (e). For many values of parameters the surface defined by this formula intersects itself. However, in an actual shell material can only be added on the outside of what already exists, and this can be represented by restricting θ to run over only part of the range $-\pi$ to π. The effect of this on internal structure can be seen in the slice of the cone shell on row (b) of page 414. Most real shells follow the model described here with remarkable accuracy. There are, however, deviations in some species, most often as a result of gradual changes in parameters during the life of the organism. As the pictures in the main text show, shells of actual molluscs (both current and fossil) exist throughout a large region of parameter space. And in fact it appears that the only parameter values that are not covered are ones where the shell could not easily have been secreted by an animal because its shape is degenerate and leaves little useful room for the animal. Some regions of parameter space are more common than others, and this may be a consequence either of natural selection or of the detailed molecular biology of mollusc growth. Shells where successive whorls do not touch (as in the first picture on row (c) of page 415) appear to be significantly less common than others, perhaps because they have lower mechanical rigidity. They do however occur, though sometimes as internal rather than external shells.

■ **History.** Following Aristotle's notion of gnomon figures that keep the same shape when they grow, equiangular spirals were discussed by René Descartes in 1638, and soon thereafter Christopher Wren noted their relation to shells. A clear mathematical model of shell growth based on equiangular spirals was given by Henry Moseley in 1838, and the model used here is a direct extension of his. Careful studies from the mid-1800s to mid-1900s validated Moseley's basic model for a wide variety of shells, though an increasing emphasis was placed on shells that showed deviations from the model. In the mid-1960s David Raup used early computer graphics to generate pictures for various ranges of parameters, but perhaps because he considered only specific classes of molluscs there emerged from his work the belief

that parameters of shells are greatly constrained—with explanations being proposed based on optimization of such features as strength, relative volume, and stability when falling through water. But as discussed in the main text I strongly suspect that in fact there are no such global constraints, and instead almost all reasonable values of parameters from the simple model used do actually occur in real molluscs. In the past few decades, increasingly complex models for shells have been constructed, typically focusing on fairly specific or unusual cases. Most of these models have far more parameters than the simple one used here, and by varying these parameters it is almost always possible to get forms that probably do not correspond to real shells. And presumably the reason for this is just that such models represent processes that do not occur in the growth of actual molluscs. One widespread issue concerns the orientation of the opening to a shell. The model used here assumes that this opening always stays vertical—which appears to be what happens most often in practice. But following the notion of Frenet frames in differential geometry, it has often come to be supposed that the opening to a shell instead typically lies in a plane perpendicular to the helix traced out by the growth of the shell. This idea, however, leads to twisted shapes like those shown below that occur rarely, if ever, in actual shells. And in fact, despite elaborate efforts of computer graphics it has proved rather difficult with parametrizations based on Frenet frames to produce shells that have a reasonable range of realistic shapes.



▪ **Page 417 · Discrete folding.** See page 892.

▪ **Page 418 · Intrinsically defined curves.** With curvature given by a function $f[s]$ of the arc length s, explicit coordinates $\{x[s], y[s]\}$ of points are obtained from (compare page 1048)

```
NDSolve[{x'[s] == Cos[θ[s]], y'[s] == Sin[θ[s]], θ'[s] ==
    f[s], x[0] == y[0] == θ[0] == 0}, {x, y, θ}, {s, 0, s_max}]
```

For various choices of $f[s]$, formulas for $\{x[s], y[s]\}$ can be found using *DSolve*:

$f[s] = 1$: $\{\text{Sin}[\theta], \text{Cos}[\theta]\}$

$f[s] = s$: $\{\text{FresnelS}[\theta], \text{FresnelC}[\theta]\}$

$f[s] = 1/\sqrt{s}$: $\sqrt{\theta}\ \{\text{Sin}[\sqrt{\theta}\ ], \text{Cos}[\sqrt{\theta}\ ]\}$

$f[s] = 1/s$: $\theta\{\text{Cos}[\text{Log}[\theta]], \text{Sin}[\text{Log}[\theta]]\}$

$f[s] = 1/s^2$: $\theta\{\text{Sin}[1/\theta], \text{Cos}[1/\theta]\}$

$f[s] = s^n$: result involves $\text{Gamma}[1/n, \pm i\,\theta^n/n]$

$f[s] = \text{Sin}[s]$: result involves $\text{Integrate}[\text{Sin}[\text{Sin}[\theta]], \theta]$, expressible in terms of generalized Kampé de Fériet hypergeometric functions of two variables.

When $s_{max} \to \infty$, $f[s] = a\,s\,\text{Sin}[s]$ yields 2D shapes that are basically nested, with pieces overlapping for $\text{Abs}[a] < 1$.

The general idea of so-called natural equations for obtaining curves from local curvature appears to have been first considered by Leonhard Euler in 1736. Many examples with fairly simple behavior were studied in the 1800s. The case of $f[s] = a\,\text{Sin}[b\,s]$ was studied by Eduard Lehr in 1932. Cases related to $f[s] = s\,\text{Sin}[s]$ were studied by Alfred Gray around 1992 using *Mathematica*.

▪ **Multidimensional generalizations.** Curvatures for surfaces and higher-dimensional objects can be defined in terms of the principal axes of approximating ellipsoids at each point. There are combinations of these curvatures—in 2D Gaussian curvature and mean curvature—which are independent of the coordinate system used. (Compare page 1049.) Given such curvatures, a surface can in principle be obtained by solving certain partial differential equations. But even in the case of zero mean curvature, which corresponds to minimal surfaces of the kind followed by an idealized soap film, this is already a mathematical problem of significant difficulty.

If one looks at projections of surfaces, it is common to see lines of discontinuity at which a surface goes, say, from having three sheets to one. Catastrophe theory provides a classification of such discontinuities—the simplest being a cusp. And as emphasized by René Thom in the 1960s, it is possible that some structures seen in animals may be related to such discontinuities.

▪ **Page 419 · Embryo development.** Starting from a single egg cell, embryos first exhibit a series of geometrically quite precise cell divisions corresponding, I suspect, to a simple neighbor-independent substitution system. When the embryo consists of a definite number of cells—from tens to tens of thousands depending on species—the phenomenon of gastrulation occurs, and the hollow sphere of cells that has been produced folds in on itself so as to begin to form more tubular structures. In organisms with a total of just a few thousand cells, the final position and type of every cell seems to be determined directly by the genetic program of the organism; most likely what happens is that each cell division leads to some modification in genetic material, perhaps through rules like those in a multiway system. Beyond a few thousand cells, however, individual cells seem to be less relevant, and instead what appears to happen is that chemicals such as retinoic acid (a derivative of vitamin A)

produced by particular cells diffuse to affect all cells in a region a tenth of a millimeter or so across. Probably as a result of chemical concentration gradients, different so-called homeobox genes are then activated in different parts of the region. Each of these genes—out of a total of 38 in humans—yields proteins which then in turn switch on or off large banks of genes, allowing different forms of behavior for cells in different places.

■ **History of embryology.** General issues of embryology were already discussed in Greek times, notably by Hippocrates and Aristotle. But even in the 1700s it was still thought that perhaps every embryo started from a very small version of a complete organism. In the 1800s, however, detailed studies revealed the progressive development of complexity in the growth of an embryo. At the end of the 1800s experiments based on removing or modifying parts of early embryos began, and by the 1920s it had been discovered that there were definite pieces of embryos that were responsible for inducing various aspects of development to occur. That concentrations of diffusing chemicals might define where in an embryo different elements would form was first suggested in the early 1900s, but it was not until the 1970s and 1980s—after it was emphasized by Lewis Wolpert in 1969 under the name "positional information"—that there was clear experimental investigation of this idea. From the 1930s and before, it was known that different genes are involved in different aspects of embryo development. And with the advent of gene manipulation methods in the 1970s and 1980s, it became possible to investigate the genetic control of development in organisms such as fruit flies in tremendous detail. Among the important discoveries made were the homeobox genes (see note above).

■ **Page 420 · Bones.** Precursors of bones can be identified quite early in the growth of most vertebrate embryos. Typically the cells involved are cartilage, with bone subsequently forming around them. In hardened bones growth normally occurs by replication of cartilage cells in plates perhaps a millimeter thick, with bone forming by a somewhat complicated process involving continual dissolving and redeposition of already hardened material. The rate at which bone grows depends on the pressure exerted on it, and presumably this allows feedback that for example prevents coiling. Quite how the complicated collection of tens of bones that make up a typical skull manage to grow so as to stay connected—often with highly corrugated suture lines—remains fairly mysterious.

■ **General constraints on growth.** Given a system made from material with certain overall properties, one can ask what distributions of growth are consistent with those properties, and what kinds of shapes can be produced. With material that is completely rigid growth can occur only at boundaries. With material where every part can deform arbitrarily any kind of growth can occur. With material where parts can locally expand, but cannot change their shape, page 1007 showed that a 2D surface will remain flat if the growth rate is a harmonic function. The Riemann mapping theorem of complex analysis then implies that even in this case, any smooth initial shape can grow into any other such shape with a suitable growth rate function. In a 3D system with locally isotropic growth the condition to avoid tearing is that the Ricci scalar curvature must vanish, and this is achieved if the local growth rate satisfies a certain partial differential equation. (See also page 1049.)

■ **Parametrizations of growth.** The idea that different objects—say different human bodies or faces—can be related by changing a small number of geometrical parameters was used by artists such as Albrecht Dürer in the 1500s, and may have been known to architects and others in antiquity. (In modern times this idea is associated for example with the notion that just a few measurements are sufficient to specify the fitting of clothes.) D'Arcy Thompson in 1917 suggested that shapes in many different species could also be related in this way. In the case of shells and horns he gave a fairly precise analysis, as discussed above. But he also drew various pictures of fishes and skulls, and argued that they were related by deformations of coordinates. Largely from this grew the field of morphometrics, in which the relative positions of features such as eyes or tips of fins are compared in different species. And although statistical significance is reduced by considering only discrete features, some evidence has emerged that different species do indeed have shapes related by changes in fairly small numbers of geometrical parameters. Such changes could be accounted for by changes in growth rates, but it is noteworthy that my results above on branching and folding make it clear that in general changes in growth rates can have much more dramatic effects.

As emphasized by D'Arcy Thompson, even a single organism will change shape if its parts grow at different rates. And in the 1930s and 1940s it became popular to study differential growth, typically under the name of allometry. Exponential growth was usually assumed, and there was much discussion about the details and correctness of this. Practical applications were made to farm animals, and later to changes in facial bone structure during childhood. But despite some work in the past twenty years using models based on fluids, solid mechanics and networks of rigid elements, much about differential growth remains unclear. (A better approach may be one similar to general relativity.)

▪ **Schemes for growth.** After the initial embryonic stage, many general features of the growth of different types of organisms can be viewed as consequences of the nature of the elements that make the organisms rigid. In plants, as we have discussed, essentially all cells have rigid cellulose walls. In vertebrate animals, rigidity comes mainly from bones that are internal to the organism. In arthropods and some other invertebrates, an exoskeleton is typically the main source of rigidity. Growth in such organisms usually then proceeds by adding soft tissue on the inside, then periodically moulting the exoskeleton. In a first approximation, the mechanical pressure of internal tissue will typically make the shape of the exoskeleton form an approximate minimal surface.

▪ **Tumors.** In both plants and animals tumors seem to grow mainly by fairly random addition of cells to their surface—much as in the aggregation models shown on page 332.

▪ **Pollen.** The grains of pollen produced by different species of plants have a remarkable range of different forms. Produced in groups of four, each grain is effectively a single cell (with two nuclei) between a few and few hundred microns across. At an overall level most grains seem to have regular polyhedral shapes, though often with bulges or dents. Perhaps such forms arise through grains effectively being made with small numbers of roughly spherical elements being either as tightly or loosely packed as possible. The outer walls of pollen grains are often covered with a certain density of tiny columns that can form spikes, or can have plates on top that can form cross-linkages and can join together to appear as patches.

▪ **Radiolarians.** The silicate skeletons of single-cell plankton organisms such as radiolarians and diatoms have been used for well over a century as examples of complex microscopic forms in biology. (See page 385.) Most likely their overall shapes are determined before they harden through minimization of area by surface tension. Their pores and cross-linkages presumably reflect packings of many roughly spherical elements on the surface during formation (as seen in the mid-1990s in aluminophosphates).

▪ **Self-assembly.** Some growth—particularly at a microscopic level—seems to be based on objects with particular shapes or affinities sticking together only in specific ways—much as in the systems based on constraints discussed on page 210 (and especially the network constraint systems of page 483). (See also page 1193.)

▪ **Animal behavior.** Simple repetitive behavior is common, as in circadian and brain rhythms, as well as peristalsis and walking. (In a millipede there are, for example, typically just two modes of locomotion, both simple, involving opposite legs moving either together or oppositely.) Many structures built by animals have repetitive forms, as in beehives and spider webs; the more complex structures made for example by termites can perhaps be understood in terms of generalized aggregation systems (see page 978). (Typical models involve the notion of stigmergy: that elements are added at a particular point based only on features immediately around them; see also page 1184.) Nested patterns may occur in flocks of birds such as geese. Fairly regular nested space-filling curves are sometimes seen in the eating paths of caterpillars. Apparent randomness is common in physiological processes such as twitchings of muscles and microscopic eye motions, as well as in random walks executed during foraging. My suspicion is that just as there appear to be small collections of cells—so-called central pattern generators—that generate repetitive behavior, so also there will turn out to be small collections of cells that generate intrinsically random behavior.

Biological Pigmentation Patterns

▪ **Collecting shells.** The shells I show in this section are mostly from my fairly small personal collection, obtained at shops and markets around the world. (A few of the ones on page 416 are from the Field Museum.) The vast majority of shells on typical beaches do not have especially elaborate patterns. The Philippines are the largest current source of collectible shells: when molluscs intended as food are caught in nets interesting-looking shells are sometimes picked out before being discarded. Shell collecting as a hobby probably had its greatest popularity in the late 1700s and 1800s. In recent times one reason for studying animals that live in cone shells is that they produce potent neurotoxins that show promise as pain-control drugs.

▪ **Shell patterns.** The so-called mantle of soft tissue which covers the animal inside the shell is what secretes the shell and produces the pattern on it. Some species deposit material in a highly regular way every day; others seem to do it intermittently over periods of months or years. In many species the outer surface of the shell is covered by a kind of skin known as the periostracum, and in most cases this skin is opaque, thereby obscuring the patterns underneath until long after the animal has died. Note that if one makes a hole in a shell, the pattern is usually quite unaffected, suggesting that the pattern is primarily a consequence of features of the underlying mantle. In addition, patterns are often divided into three or four large bands, presumably in correspondence with features of the anatomy of the mantle. Sometimes physical ridges exist on shells in correspondence with their

pigmentation patterns. It is not clear whether multiple kinds of shell patterns can occur within one species, or whether they are always associated with genetically different species.

■ **Cowries.** In cowries the outside of the shell is covered by the mantle of the animal. The patterns on the shell typically involve spots, and are typical of those obtained from 2D cellular automata of the kind shown on page 428. The mantle is normally in two parts; the boundary between them shows up as a discontinuity in the shell pattern.

■ **History.** Elaborate patterns on shells have been noticed since antiquity, and have featured in a number of well-known works of art and literature. Since the late 1600s they have also been extensively used in classifying molluscs. But almost no efforts to understand the origins of such patterns seem to ever have been made. One study was done in 1969 by Conrad Waddington and Russell Cowe in which patterns on one particular kind of shell were reproduced by a specific computer simulation based on the idea of diverging waves of pigment. In 1982 I noticed that the patterns I had generated with 1D cellular automata looked remarkably similar to patterns on shells. I used this quite widely as an illustration of how cellular automata might be relevant to modelling natural systems. And I also made some efforts to do actual biological experiments, but I gave up when it seemed that the species of molluscs I wanted to study were difficult, if not impossible, to keep in captivity. Following my work, various other studies of shell patterns were done. Bard Ermentrout, John Campbell and George Oster constructed a model based on the idea that pigment-producing cells might act like nerve cells with a certain degree of memory. And Hans Meinhardt has constructed progressively more elaborate models based on reaction-diffusion equations.

■ **Page 426 · Animals shown.** Flatworm, cuttlefish, honeycomb moray, spotted moray, foureye butterfly fish; emperor angelfish, suckermouth catfish, ornate cowfish, clown triggerfish, poison-dart frog; ornate horned frog, marbled salamander, spiny softshell, gila monster, ball python; gray-banded kingsnake, guinea fowl, peacock, ring-tailed lemur, panda; cheetah, ocelot, leopard, tiger, spotted hyena; western spotted skunk, civet, zebra, brazilian tapir, giraffe.

■ **Animal coloration.** Coloration can arise either directly through the presence of visible colored cells such as those in freckles, or indirectly by virtue of cells such as hair follicles imparting pigments to the non-living elements such as fur, feathers and scales that grow from them. In many cases such elements are arranged in a highly regular way, often in a repetitive hexagonal pattern. Evolutionary optimization is often used to explain observed pigmentation patterns—with varying degrees of success. The notion that for example the stripes of a zebra are for camouflage may at first seem implausible, but there is some evidence that dramatic stripes do make it harder for a predator to recognize the overall shape of the zebra. Many of the pigments used by animals are by-products of metabolism, suggesting that at least at first pigmentation patterns were probably often incidental to the operation of the animal.

■ **Page 427 · Implementation.** Given a 2D array of values a and a list of weights w, each step in the evolution of the system corresponds to

```
WeightedStep[w_List, a_] := Map[If[# > 0, 1, 0] &,
    Sum[w[[1 + i]] Apply[Plus, Map[RotateLeft[a, #] &,
        Layer[i]]], {i, 0, Length[w] - 1}], {2}]
Layer[n_] := Layer[n] = Select[Flatten[Table[{i, j},
    {i, -n, n}, {j, -n, n}], 1], MemberQ[#, n | -n] &]
```

■ **Features of the model.** The model is a totalistic 2D cellular automaton, as discussed on page 927. It shows class 2 behavior in which information propagates only over limited distances, so that except when the total size of the system is comparable to the range of the rule, boundary conditions are not crucial.

Similar models have been considered before. In the early 1950s (see below) Alan Turing used a model which effectively differed mainly in having continuous color levels. In 1979 Nicholas Swindale constructed a model with discrete levels to investigate ocular dominance stripes in the brain (see below). And following my work on cellular automata in the early 1980s, David Young in 1984 considered a model even more similar to the one I use here.

There are simple cellular automata—such as 8-neighbor outer totalistic code 196623—which eventually yield maze-like patterns even when started from simple initial conditions. The rule on page 336 gives dappled patterns with progressively larger spots.

■ **Reaction-diffusion processes.** The cellular automaton in the main text can be viewed as a discrete idealization of a reaction-diffusion process. The notion that diffusion might be important in embryo development had been suggested in the early 1900s (see page 1004), but it was only in 1952 that Alan Turing showed how it could lead to the formation of definite patterns. Diffusion of a single chemical always tends to smooth out distributions of concentration. But Turing pointed out that with two chemicals in which each can be produced from the other it is possible for separated regions to develop. If $c = \{u[t, x], v[t, x]\}$ is a vector of chemical concentrations, then for suitable values of parameters even the standard linear diffusion equation $\partial_t c == d \,.\, \partial_{xx} c + m \,.\, c$

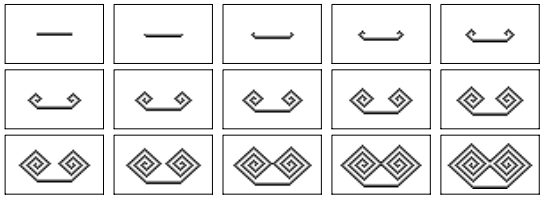
can exhibit an instability which causes disturbances with certain spatial wavelengths to grow (compare page 988). In his 1952 paper Turing used a finite difference approximation to a pair of diffusion equations to show that starting from a random distribution of concentration values dappled regions could develop in which one or the other chemical was dominant. With purely linear equations, any instability will always eventually lead to infinite concentrations, but Turing noted that this could be avoided by using realistic nonlinear chemical rate equations. In the couple of years before his death in 1954, Turing appears to have tried to simulate such nonlinear equations on an early digital computer, but my cursory efforts to understand his programs—written as they are in a 32-character machine code—were not successful.

Following Turing's work, the fact that simple reaction-diffusion equations can yield spatially inhomogeneous patterns has been rediscovered—with varying degrees of independence—many times. In the early 1970s Ilya Prigogine termed the patterns dissipative structures. And in the mid-1970s, Hermann Haken considered the phenomenon a cornerstone of what he called synergetics.

Many detailed mathematical analyses of linear reaction-diffusion equations have been done since the 1970s; numerical solutions to linear and occasionally nonlinear such equations have also often been found, and in recent years explicit pictures of patterns—rather than just curves of related functions—have commonly been generated. In the context of biological pigmentation patterns detailed studies have been done particularly by Hans Meinhardt and James Murray.

▪ **Scales of patterns.** The visual appearance of a pattern on an actual animal depends greatly on the scale of the pattern relative to the whole animal. Pandas and anteaters, for example, typically have just a few regions of different color, while other animals can have hundreds of regions. Studies based on linear reaction-diffusion equations sometimes assume that patterns correspond to stationary modes of the equations, which inevitably depend greatly on boundary conditions. But in more realistic models patterns emerge from long time behavior with generic initial conditions, making boundary conditions—and effects such as changes in them associated with growth of an embryo—much less important.

▪ **Excitable media.** In many physical situations effects become decreasingly important as they propagate further away. But in active or excitable media such as heart, muscle and nerve tissue an effect can maintain its magnitude as it propagates, leading to the formation of a variety of spatial structures. An early model of such media was constructed in 1946 by Norbert Wiener and Arturo Rosenblueth, based on a discrete array of continuous elements. Models with discrete elements were already considered in the 1960s, and in 1977 James Greenberg and Stuart Hastings introduced a simple 2D cellular automaton with three colors. The pictures below show what is probably the most complex feature of this cellular automaton and related systems: the formation of spiral waves. Such spiral waves were studied in 2D and 3D in the 1970s and 1980s, particularly by Arthur Winfree and others; they are fairly easy to observe in both inorganic chemical reactions (see below) and slime mold colonies.



▪ **Examples in chemistry.** Overall concentrations in chemical reactions can be described by nonlinear ordinary differential equations. Reactions with oscillatory behavior were predicted by Alfred Lotka in 1910 and observed experimentally by William Bray in 1917, but for some reason they were not further investigated at that time. An example was found experimentally by Boris Belousov in 1951 and extensive investigations of it were begun by Anatol Zhabotinsky around 1960. In the early 1970s spiral waves were seen in the spatial distribution of concentrations in this reaction, and by the end of the 1970s images of such waves were commonly used as icons of the somewhat ill-defined notion of self-organization.

▪ **Maze-like patterns.** Maze-like patterns occur in several quite different kinds of systems. Cases in which the underlying mechanism is probably similar to that discussed in the main text include brain coral, large-scale vegetation bands seen in tropical areas, patterns of sand dunes, patterns in pre-turbulent fluid convection, and ocular dominance stripes consisting of regions of brain tissue that get marked when different dye is introduced into nerves from left and right eyes. Cases in which the underlying mechanism is probably more associated with folding of fixed amounts of material include human fingerprint patterns and patterns in ferrofluids consisting of suspensions of magnetic particles.

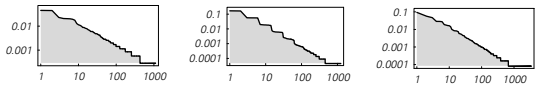
▪ **Origins of randomness.** The model in the main text assumes that randomness enters through initial conditions. If the two parts of a single animal—say opposite wings on a butterfly—

form together, then these initial conditions can be expected to be the same. But usually even the two sides of a single animal are never physically together, and they normally end up having quite uncorrelated random features. In cases such as fingerprints and zebra stripes there is some correlation between different sides, suggesting an intrinsic component to the randomness that occurs. (The fingerprints of identical twins are typically similar but not identical; iris patterns are quite different.) Note that at least sometimes random initial patterns are formed by cells that have the same type, but different lineages—as in cells expressing genes from the two different X chromosomes in a female animal such as a typical tortoiseshell cat. (In general, quite a few traits—particularly related to aging—can show significant variation in strains of organisms that are genetically identical.)

Financial Systems

■ **Laws of human behavior.** Over the past century there have been a fair number of quantitative laws proposed for features of human behavior. Some are presumably a direct reflection of human biological construction. Thus for example, Weber's law that the perceived strength of a stimulus tends to vary logarithmically with its actual strength seems likely to be related to the electrochemistry of nerve cells. Of laws for more complicated cognitive or social phenomena the vast majority are statistical in nature. And of those that withstand scrutiny, most in my experience turn out to be transformed versions of statements that some quantity or another can be approximated by perfect randomness. Gaussian distributions typically arise when measurements involve sums of random quantities; other common distributions are obtained from products or other simple combinations of random quantities, or from the results of simple processes based on random quantities. Exponential distributions (as seen, for example, in learning curves) and power-law distributions (as in Zipf's law below) are both, for example, very easy to obtain. (Note that particularly in economics there are also various laws derived from calculus and game theory that are viewed as being quite successful, and are not fundamentally statistical.)

■ **Zipf's law.** To a fairly good approximation the n^{th} most common word in a large sample of English text occurs with frequency $1/n$, as illustrated in the first picture below. This fact was first noticed around the end of the 1800s, and was attributed in 1949 by George Zipf to a general, though vague, Principle of Least Effort for human behavior. I suspect that in fact the law has a rather simple probabilistic origin. Consider generating a long piece of text by picking at random from k letters and a space. Now collect and rank all the "words" delimited by spaces that are formed. When $k = 1$, the n^{th} most common word will have frequency c^{-n}. But when $k \geq 2$, it turns out that the n^{th} most common word will have a frequency that approximates c/n. If all k letters have equal probabilities, there will be many words with equal frequency, so the distribution will contain steps, as in the second picture below. If the k letters have non-commensurate probabilities, then a smooth distribution is obtained, as in the third picture. If all letter probabilities are equal, then words will simply be ranked by length, with all k^m words of length m occurring with frequency p^m. The normalization of probabilities then implies $p = 1/(2k)$, and since the word at rank roughly k^m then has probability $1/(2k)^m$, Zipf's law follows.



■ **Motion of people and cars.** To a first approximation crowds of people seem to show aggregate fluid-like behavior similar to what is seen in gases. Fronts of people—as occur in riots or infantry battles—seem to show instabilities perhaps analogous to those in fluids. Road traffic that is constrained to travel along a line exhibits stop-start instabilities when its overall rate is reduced, say by an obstruction. This appears to be a consequence of the delay before one driver responds to changes in speed of cars in front of them. Fairly accurate cellular automaton models of this phenomenon were developed in the early 1990s.

■ **Growth of cities.** In the absence of geographical constraints, such as terrain or oceans, cities typically have patchy, irregular, shapes. At first an aggregation system (see page 331) might seem to be an obvious model for their growth: each new development gets added to the exterior of the city at a random position. But actual cities look much more irregular. Most likely the reason is that embedded within the cities is a network of transportation routes, and these tend to have a tree- or vein-like structure (though not necessarily with a single center)—with major freeways etc. as trunks. The result of following this structure is to produce a much more irregular boundary.

■ **Randomness in markets.** After the somewhat tricky process of correcting for overall trends, empirical price data from a wide range of markets seem to a first approximation to follow random walks and thus to exhibit Gaussian fluctuations, as noted by Louis Bachelier in 1900. However, particularly on timescales less than a day, it has in the past decade become clear that, as suggested by Benoit Mandelbrot in the early 1960s, large price fluctuations are significantly more common than a Gaussian distribution would imply.

Such an effect is easy to model with the approach used in the main text if different entities interact in clumps or herds—which can be forced if they are connected in a hierarchical network rather than just a line.

The observed standard deviation of a price—or essentially so-called volatility or beta—can be considered as a measure of the risk of fluctuations in that price. The Capital Asset Pricing Model proposed in the early 1960s suggested that average rates of price increases should be proportional to such variances. And the Black-Scholes model from 1973 implies that prices of suitably constructed options should depend in a sense only on such variances. Over the past decade various corrections to this model have been developed based on non-Gaussian distributions of prices.

▪ **Speculative markets.** Cases of markets that seem to operate almost completely independent of objective value have occurred many times in economic history, particularly in connection with innovations in technology or finance. Examples range from tulip bulbs in the mid-1630s to railroads in the mid-1800s to internet businesses in the late 1990s. (Note however that in any particular case it can be claimed that certain speculation was rational, even if it did not work out—but usually it is difficult to get convincing evidence for this, and often effects are obscured by generalized money supply or bankruptcy issues.)

▪ **Properties of markets.** Issues of how averaging is done and how irrelevant trends are removed turn out to make unequivocal tests of almost any quantitative hypothesis about prices essentially impossible. The rational expectations theory that prices reflect discounted future earnings has for example been subjected to many empirical tests, but has never been convincingly proved or disproved.

▪ **Efficient markets.** In its strong form the so-called Efficient Market Hypothesis states that prices immediately adjust to reflect all possible information, so that knowing a particular piece of information can never be used to make a profit. It is now widely recognized—even in academia—that this hypothesis is a fairly poor representation of reality.

▪ **Details of trading.** Cynics might suggest that much of the randomness in practical markets is associated with details of trading. For much of the money actually made from markets on an ongoing basis comes from commissions on trades. And if prices quickly settled down to their final values, fewer trades would tend to be made. (Different entities would nevertheless still often need liquidity at different times.)

▪ **Models of markets.** When serious economic theory began in the 1700s arguments tended to be based purely on common sense. But with the work of Léon Walras in the 1870s mathematical models began to become popular. In the early 1900s, common sense again for a while became dominant. But particularly with the development of game theory in the 1940s the notion became established, at least in theoretical economics, that prices represent equilibrium points whose properties can be derived mathematically from requirements of optimality. In practical trading, partly as an outgrowth of theories of business cycles, there had emerged all sorts of elaborate so-called technical analysis in which patterns of price movements were supposed—often on the basis of almost mystical theories—to be indicators of future behavior. In the late 1970s, particularly after the work of Fischer Black and Myron Scholes on options pricing, new models of markets based on methods from statistical physics began to be used, but in these models randomness was taken purely as an assumption. In another direction, it was noticed that dynamic versions of game theory could yield iterated maps and ordinary differential equations which would lead to chaotic behavior in prices, but connections with randomness in actual markets were not established. By the mid-1980s, however, it began to be clear that the whole game-theoretical idea of thinking of markets as collections of rational entities that optimize their positions on the basis of complete information was quite inadequate. Some attempts were made to extend traditional mathematical models, and various highly theoretical analyses were done based on treating entities in the market as universal computers. But by the end of the 1980s, the idea had emerged of doing explicit computer simulations with entities in the market represented by practical programs. (See also page 1105.) Often these programs used fairly sophisticated algorithms intended to mimic human traders, but in competitions between programs simpler algorithms have never seemed to be at much of a disadvantage. The model in the main text is in a sense an ultimate idealization along these lines. It follows a sequence of efforts that I have made since the mid-1980s—though have never considered very satisfactory—to find minimal but accurate models of financial processes.