16

Mathematical Chaos and Strange Attractors

November, 1981

You can't know how happy I am that we met,
I'm strangely attracted to you.

—Cole Porter, "It's All Right with Me"

A few months ago, while walking through the corridors of the physics department of the University of Chicago with a friend, I spotted a poster announcing an international symposium titled "Strange Attractors". My eye could not help but be strangely attracted by this odd term, and I asked my friend what it was all about. He said it was a hot topic in theoretical physics these days. As he described it to me, it sounded quite wonderful and mysterious.

I gathered that the basic idea hinges on looking at what might be called "mathematical feedback loops": expressions whose output can be fed back into them as new input, the way a loudspeaker's sounds can cycle back into a microphone and come out again. From the simplest of such loops, it seemed, both stable patterns and chaotic patterns (if this is not a contradiction in terms!) could emerge. The difference was merely in the value of a single parameter. Very small changes in the value of this parameter could make all the difference in the world as to the orderliness of the behavior of the loopy system. This image of order melting smoothly into chaos, of pattern dissolving gradually into randomness, was exciting to me.

Moreover, it seemed that some unexpected "universal" features of the transition into chaos had recently been unearthed, features that depended solely on the presence of feedback and that were virtually insensitive to other details of the system. This generality was important, because any mathematical model featuring a gradual approach to chaotic behavior might provide a key insight into the onset of turbulence in all kinds of physical systems. Turbulence, in contrast to most phenomena successfully understood in physics, is a *nonlinear* phenomenon; two solutions to the equations of turbulence do not add up to a new solution. Nonlinear mathematical phenomena are much less well understood than linear ones, which is why a good mathematical description of turbulence has eluded physicists for a long time, and would be a fundamental breakthrough.

When I later began to read about these ideas, I found out that they had actually grown out of many disciplines simultaneously. Pure mathematicians had begun studying the iteration of nonlinear systems by using computers. Theoretical meteorologists and population geneticists, as well as theoretical physicists studying such diverse things as fluids, lasers, and planetary orbits, had independently come up with similar nonlinear mathematical models featuring chaos-pregnant feedback loops and had studied their properties, each group finding some quirks that the others had not found. Moreover, not only theorists but also experimentalists from these widely separated disciplines had simultaneously observed chaotic phenomena that share certain basic patterns. I soon saw that the simplicity of the underlying ideas gives them an elegance that, in my opinion, rivals that of some of the best of classical mathematics. Indeed, there is an eighteenth- or nineteenth-century flavor to some of this work that is refreshingly concrete in this era of staggering abstraction.

Probably the main reason these ideas are only now being discovered is that the style of exploration is entirely modern: it is a kind of experimental mathematics, in which the digital computer plays the role of Magellan's ship, the astronomer's telescope, and the physicist's accelerator. Just as ships, telescopes, and accelerators must be ever larger, more powerful, and more expensive in order to probe ever more hidden regions of nature, so one would need computers of ever greater size, speed, and accuracy in order to explore the remoter regions of mathematical space. By the same token, just as there was a golden era of exploration by ship and of discoveries made with telescopes and accelerators, characterized by a peak in the ratio of new secrets uncovered to money spent, so one would expect there to be a golden era in the experimental mathematics of these models of chaos. Perhaps this era has already occurred, or perhaps it is occurring right now. And perhaps after it, we will witness a flurry of theoretical work to back up these experimental discoveries.

In any case, it is a curious and delightful brand of mathematics that is being done. This way of doing mathematics builds powerful visual imagery and intuitions directly into one's understanding. The power of computers allows one to bypass the traditional "theorem-proof-theorem-proof" brand of mathematics, and to arrive quickly at *empirical* observations and discoveries that reinforce each other, and that form a rich and coherent network of results. In the long run, it may turn out to be easier to find proofs of these results (if proofs are still desired), thanks to the careful and thorough exploration of the conceptual territory in advance. It's an upstart's way of doing mathematics, and not all mathematicians approve.

One of the strongest proponents of this style of mathematizing has been Stanislaw M. Ulam, who, when computers were still young, turned them loose on problems of nonlinear iteration as well as on problems from many other branches of mathematics. It is from Ulam's early studies with Paul Stein that many of the ideas to be sketched here follow.

* * *

So much for romance. Let us work our way up to the concept of "strange attractors" by beginning with the more basic concept of an *attractor.* This whole field is founded on one concept: the iteration of a real-valued mathematical function—that is, the behavior of the sequence of values x, $f(x), f(f(x)), f(f(f(x)))$, . . . , where f is some interesting function. The initial value of x is called the *seed.* The idea is to feed f's output back into f as new input over and over again, to see if some kind of pattern emerges.

An interesting and not too difficult problem concerning the iteration of a function is this: Can you invent a function p with the property that for any real value of x, $p(x)$ is also real, and where $p(p(x))$ equals $-x$? The condition that $p(x)$ be real is what gives the problem a twist; otherwise the function $p(x) = ix$ (where i is the square root of -1) would work. In fact, you can even think of the challenge as that of finding a real-valued "square root of the minus sign". A related problem is to find a real-valued function q, whose property is that $q(q(x)) = 1/x$ for all x other than zero. Note that no matter how you construct p and q, each will have the property that, given any seed, repeated iteration creates a cycle of length four.

Now, more generally, what kinds of functions, when repeatedly iterated, are likely to exhibit interesting cyclic or near-cyclic behavior? A simple function such as $3x$ or x^3, when iterated, does not do anything like that. The nth iteration of $3x$, for example, is $3 \times 3 \times 3 \times \ldots \times 3x$, with n 3's—that is, $3^n x$—and the nth iteration of x^3 is just $(((x^3)^3)^3)^{\cdots 3}$ with n 3's again, which amounts to x^{3^n}. Nothing cycle-like here; the values just keep going up and up and up. To reverse this trend, one needs a function with some sort of switchback—a little zigzag or twist. A more technical way of putting it is that one needs a *nonmonotonic* function: a function whose graph is folded—that is, it starts moving one way—say upward—and then bends back the other way—say downward.

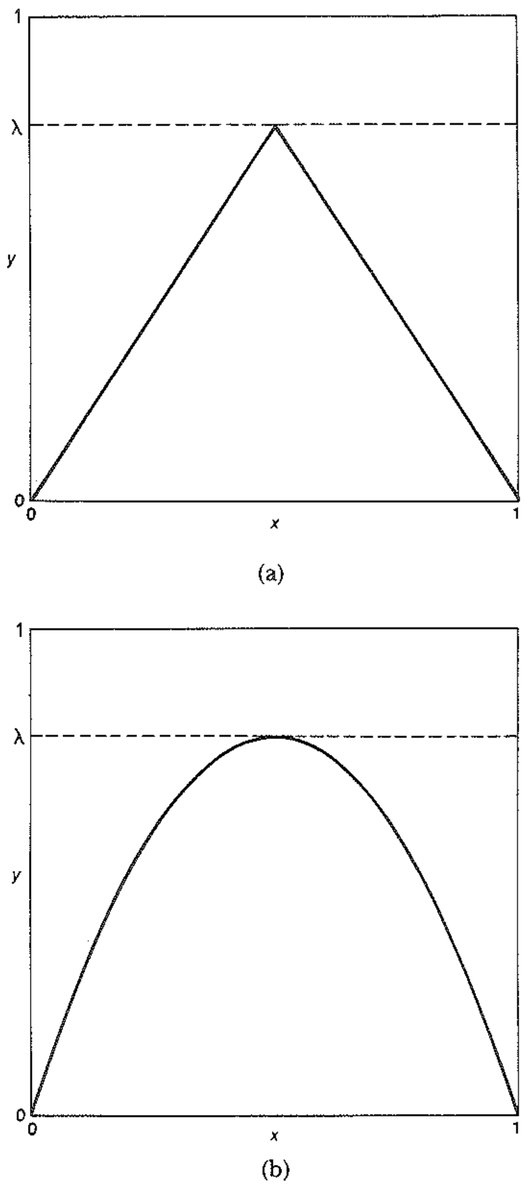


FIGURE 16–1. *Two nonmonotonic, or "folded", functions in the unit square. In (a), a sharp peak, and in (b), a parabola. The maximum height of both is defined by the parameter λ.*

In Figure 16-1*a*, we have a sawtooth with a sharp point at its top, and in Figure 16-1*b*, a smoothly bending parabolic arc. Each of them rises from the origin, eventually reaches a peak height called λ, and then comes back down for a landing on the far side of the interval. Of course there are uncountably many shapes that rise to height λ and then come back down, but these two are among the simplest. And of the two, the parabola is perhaps the simpler, or at least the more mathematically appealing. Its equation is $y=4\lambda x(1-x)$, with λ not exceeding 1.

We allow input (values of x) only between 0 and 1. As the graph shows, for any x in that interval, the output (y) always is between 0 and λ. Therefore the output value can always be fed back into the function as input, which ensures that repeated iteration will always be possible. When you repeatedly iterate a "folded" function like this, the successive y-values you produce will sometimes go up and sometimes down—always hovering, of course, between 0 and λ. The fold in the graph guarantees interesting effects when the function is iterated—as we shall see.

It turns out that the spectacular differences in the degree of regularity of patterns I mentioned above are due to variations in the setting of what we might call the "λ-knob". Depending on the value the knob is set at, the function yields an incredible variety of "orbits"—that is, sequences $x, f(x), f(f(x))$, and so on. In particular, for λ below a certain critical value $\lambda_c=0.892486417967\ldots$, the orbits are all regular and patterned (although there are various degrees of patternedness; generally the lower λ is, the more simply the orbit is patterned), but for λ at or beyond this critical value, hold onto your hat! An essentially chaotic sequence of values will be traced out by the values $x, f(x), f(f(x)), \ldots$, no matter what positive seed value of x you choose. In the case of the parabola, the critical role played by varying the λ-knob seems to have been first realized by P. J. Myrberg in the early 1960's, but his work was published in a little-known journal and did not attract much attention. Some ten years later, Nicholas C. Metropolis, Paul Stein, and Myron Stein rediscovered the importance of the knob not only for the parabola but also for many other functions. Indeed, they discovered that as far as certain topological properties were concerned, the function did not matter—only the value of λ did. This property has come to be called "structural universality".

*   *   *

In order to see how such a nonintuitive dependence on the setting of the λ-knob comes about, one must develop a visual sense for the process of iterating $f(x)$. This is readily done. Suppose we set λ to 0.7. The graph of $f(x)$ appears in Figure 16-2. In addition, the line $y=x$ appears as a 45-degree broken line. (This graph and most of the others in this article were produced on a small computer by Mitchell J. Feigenbaum of the Los Alamos National Laboratory.)

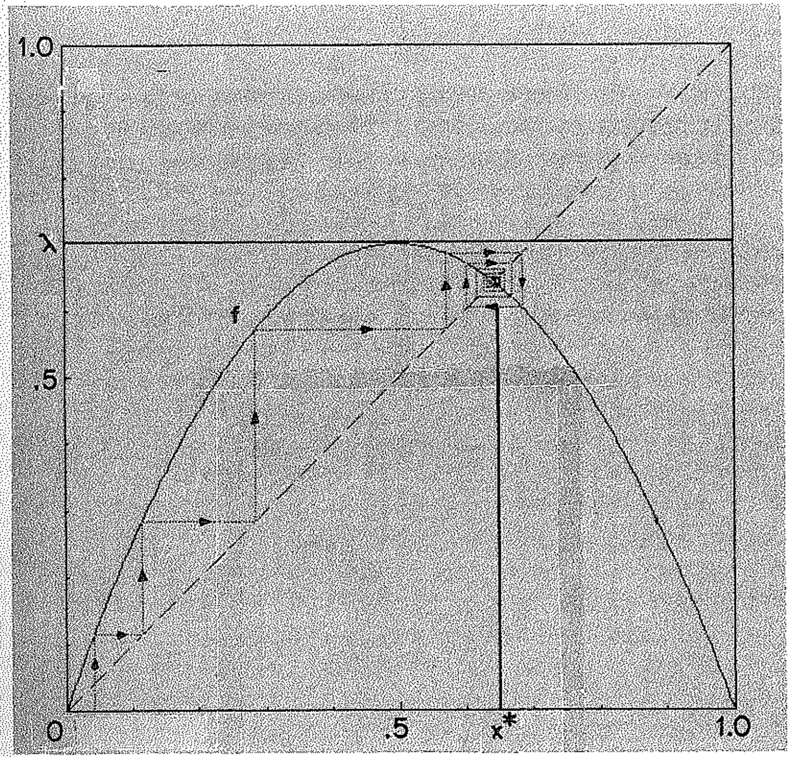


FIGURE 16–2. *The parabola defined by "λ-knob" setting of 0.7. An initial* x*-value of about 0.04 is used as a "seed" for iteration, and the pathway taken is shown. Eventually it settles down at a fixed point, denoted by* x^*.

Consider the two x-values where the 45-degree line and the curve intersect. They are at $x=0$ and $x=9/14=0.643$. Let us designate the nonzero value as x^*. By construction, then, $f(x^*)$ equals x^*, and repeated iteration of f at this x-value will get you into an infinite loop. The same happens if you start iterating at $x=0$: you get stuck in an endless loop. However, there is a significant difference between these two *fixed points* of f. It is best indicated by taking some other initial value of x, say one close to 0.04, as is shown in the same figure. Call this starting x-value x_0. There is an elegant graphical way to generate the orbit of any seed x_0. A vertical line at x-value x_0 will hit the curve at height $y_0=f(x_0)$. To iterate f, we must draw a new vertical line located at the new x-value equal to this y-value. This is where the 45-degree line $y=x$ comes in handy. Staying at height y_0, we simply move over horizontally until we hit that 45-degree line. Then, since along this line y equals x, both x and y equal y_0. Let us call this new x-value x_1. We now draw a second vertical line. This one will hit the curve at height $y_1=f(x_1)=f(y_0)=f(f(x_0))$. Now we just repeat.

In brief, iteration is realized graphically by a simple recipe:

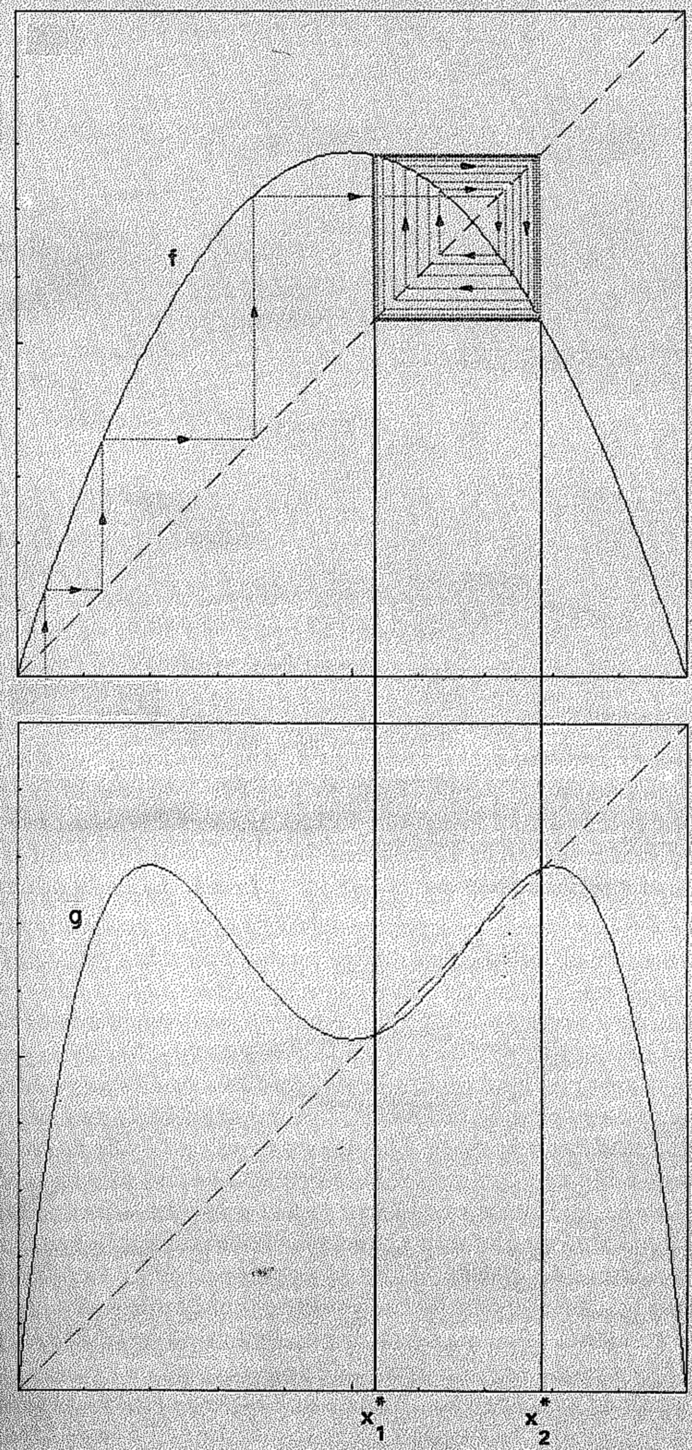
> (1) Move vertically until you hit the curve; then
> (2) Move horizontally until you hit the diagonal line.
> Repeat steps (1) and (2) over and over again.

The results of this procedure with seed $x_0=0.04$ are also shown in Figure 16-2. We are led in a merry chase 'round and 'round the point whose x-coordinate and y-coordinate are x^*. Gradually we close down on that point. Thus x^* is a special kind of fixed point, because it attracts iterated values of $f(x)$. It is the simplest example of an *attractor:* every possible seed (except 0) is drawn, through iteration of f, to this stable x-value. This x^* is therefore called an *attractive* or *stable* fixed point. By contrast, 0 is a *repellent* or *unstable* fixed point, since the orbit of any initial x-value, even one infinitesimally removed from 0, will proceed to move away from 0 and toward x^*. Note that sometimes the iterates of f will overshoot x^*, sometimes they will fall short—but they inexorably draw closer to x^*, zeroing in on it like swallows returning to Capistrano. One might also think of such familiar and charming metaphors of prey-seekers as heat-seeking missiles, mosquitos, bloodhounds, Nazi-hunters, sharks, and lastly, the children's rhyme, "Around the world, and around the world, goes a big bear; he bores a hole, and he bores a hole, right . . . in . . . *there*!"

What accounts for this radical qualitative difference between the two fixed points (0 and x^*) of f? A careful look at Figure 16-2 will show that it is the fact that at 0, the curve is sloped too steeply. In particular, the slope there is greater than 45 degrees. It is the local slope of the curve that controls how far you move horizontally each time you iterate f. Whenever the curve is steeper than 45 degrees (either rising or falling) it tends to pull you farther and farther away from your starting point as you repeatedly iterate by rules (1) and (2). Hence the criterion for the stability of a fixed point is: The slope at the fixed point should be less than 45 degrees. Now, this is the case for x^* when λ equals 0.7. In fact, the slope there is about 41 degrees, whereas at 0 it is much greater than 45 degrees.

What happens if we increase λ? The position of x^* (x^* being by definition the x-coordinate of the point where the curve f and the line $y=x$ intersect) will change, and the slope of f at x^* will increase as well. What happens when the slope hits 45 degrees or exceeds it? This occurs when λ is 3/4. We will call this special value of the λ-knob Λ_1. Let us look at the graph for a slightly greater λ-knob setting, namely $\lambda=0.785$. (See Figure 16-3.)

What if we begin with some random seed instead, again say $x=0.04$? The resulting orbit is shown in Figure 16-3*a*. As you can see, a very pretty thing happens. At first the values move up toward the vicinity of x^* (now an unstable fixed point of f), but then they spiral gradually outward and settle



down smoothly to a kind of "square dance" converging on two special values x_1^* and x_2^*. This elegant oscillation is called a *2-cycle,* and the pair of x-values that constitute it (x_1^* and x_2^*) is again an attractor—in particular, an attractor of period two. This term means that our 2-cycle is stable: it attracts x-values from far and wide as f is iterated. The orbit for any positive seed value (except x^* itself) will eventually fall into the same dance. That is, it will asymptotically approach the perfect 2-cycle composed of the points x_1^* and x_2^*, although it will never quite reach it exactly. From a physicist's point of view, however, the accuracy of the approach soon becomes so great that one can just as well say that the orbits have been "trapped" by the attractor.

An enlightening way to understand this is to look at a graph of a new function made from the old one. Consider the graph of $g(x)=f(f(x))$, shown in Figure 16-3*b*. This two-humped camel is called the *iterate* of f. First of all, observe that any fixed point of f is also a fixed point of g, so that 0 and x^* will be fixed points of g. But secondly observe that since $f(x_1^*)$ equals x_2^*, and conversely $f(x_2^*)$ equals x_1^*, g will have two new fixed points: $g(x_1^*)=x_1^*$ and $g(x_2^*)=x_2^*$. Graphically, x_1^* and x_2^* are easily found: they are intersection points of the 45-degree line with the two-humped graph of $g(x)$. There are four such points (0 and x^* being the other two). As we have seen, the criterion for the stability of any fixed point under iteration is that the slope at that point should be less than 45 degrees. Here we are concerned with fixed points of g, and hence with g's slope (as distinguished from f's slope). Indeed, in the same figure, you can clearly see that at 0 and at x^*, g is sloped more steeply than 45 degrees, whereas at both x_1^* and x_2^*, g's slope is less than 45 degrees. In fact, quite remarkably, not only are both slope values less than 45 degrees, but also, as it turns out through a simple bit of calculus, they are equal (or "slaved" to each other, as it is sometimes put).

* * *

We have now seen an attractor of period one get converted into an attractor of period two at a special value of λ (namely, $\lambda=3/4$). Precisely at that value, the single fixed point x^* splits into two oscillating values, x_1^* and x_2^*. Of course they coincide at "birth", but as λ increases, they separate and draw farther and farther apart. This increase of λ will also cause g's slope at these two stable fixed points (of g) to get steeper and steeper until finally, at some λ-value, g, like its progenitor f, will reach its own breaking point (*i.e.*, the identical slopes at both x_1^* and x_2^* will exceed 45 degrees), and each of these two attracting points will break up, spawning its own local 2-cycle. (Actually, the cycles are 2-cycles only as far as g is concerned; for f, the new points are elements of an attractor of period four. You must be careful to keep f and g straight in your mind!) These two splittings will happen at exactly the same "moment" (*i.e.*, at the same λ-knob setting), since the value of the slope of g at x_1^* is slaved to the value of the slope at x_2^*. This λ-knob setting will be called Λ_2, and it has the value of 0.86237 . . .

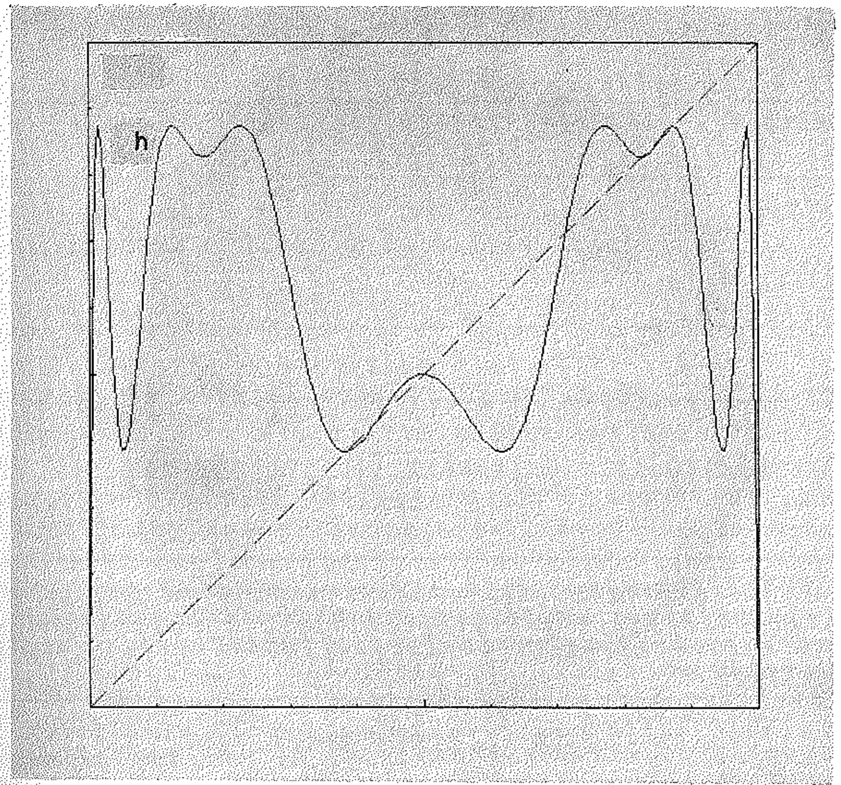


FIGURE 16–4. *A picture of* f*'s iterate's iterate* h *at a still higher value of* λ, *namely 0.87.*

Here, as with a joke, you may anticipate the punch line by the time you have heard the theme and one variation. Hence by now you have probably surmised that at some new value Λ_3, all four points in f's attractor will simultaneously fission, yielding a periodic attractor consisting of eight points; and thereafter this pattern will go on and on, doubling and redoubling as various special λ-knob settings are reached and passed. If this is your guess, you are quite right, and the underlying reason is the same each time: the (identical) slopes at all the stable fixed points of some graph reach the critical angle of 45 degrees. In the case of the first fission (at Λ_1) it was the slope of f itself at the single point x^*. The next fission was due to the slopes at g's two stable fixed points x_1^* and x_2^* simultaneously reaching 45 degrees. Analogously, Λ_3 is that value of λ at which the slope of $h(x)=g(g(x))=f(f(f(f(x))))$ hits 45 degrees simultaneouly at the four stable fixed points of h. And so it goes. Figure 16-4 shows the bumpy appearance of $h(x)$ at a λ-value of approximately 0.87.

In Figure 16-5, the locations on the x-axis of the stable fixed points of f are shown for Λ_1 through Λ_6 (by which time there are 32 of them, some clustered so closely that they cannot be distinguished). The points are pictured just at the moment of their becoming unstable, each one like a cell

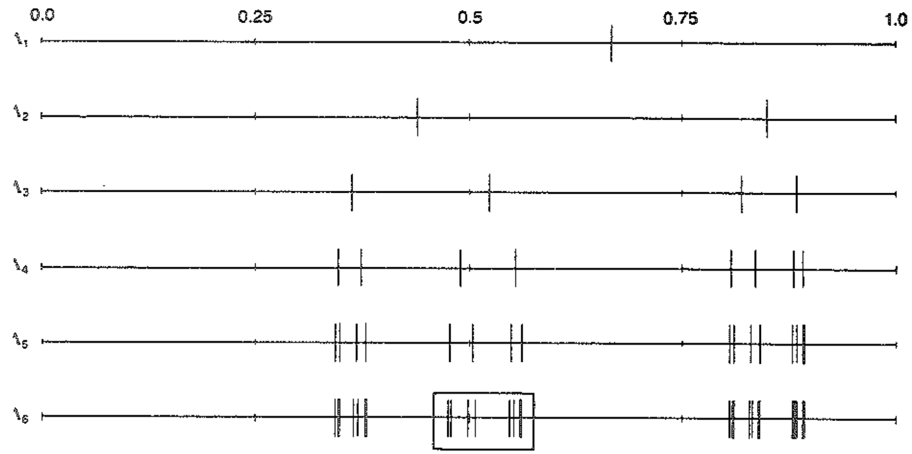
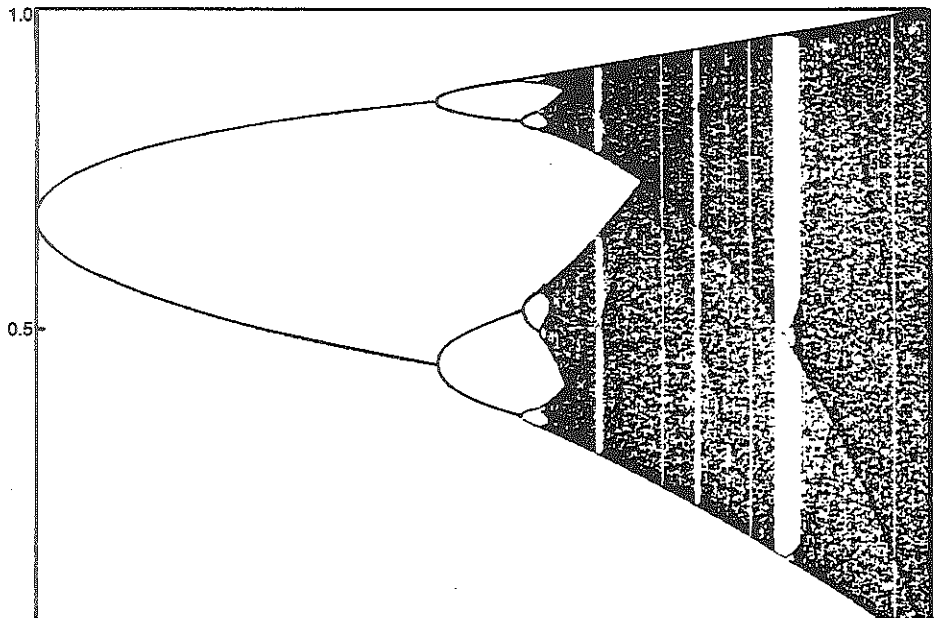


FIGURE 16–5. *Showing how stable attractors become unstable and undergo "fission" at a series of increasing λ-values, denoted* Λ_n *for n=1, 2, 3,... Note how the boxed subpattern on the lowest line resembles the entire pattern two lines above. This resemblance becomes more and more accurate the larger n gets.*

FIGURE 16–6. *A graph showing the evolution of attractors as λ increases from 0 to 1. Bifurcations begin at λ=0.75 and escalate towards chaos. The "chaotic region", beginning at λ=0.892...., shows unexpectedly beautiful fine structure. [From "Roads to Chaos" by Leo P. Kadanoff in* Physics Today, *December 1983 p.51; see also J. P. Crutchfield, D. Farmer, and B. A. Huberman,* Physics Reports, *Vol. 92, pp. 45–82, December, 1982.]*



on the verge of division. Notice the neat pattern in the distribution of the attracting points. Looking at these graphs of the spacings of the elements of the successive period-doubled attractors of *f*, you can see that each line can be made from the one above it through a recursive geometric scheme whereby each point is replaced by two "twin" points below it. Each local clustering pattern of points echoes the global clustering pattern, simply reduced in scale (and also, in alternating local clusters, left and right are reversed). For example, in the bottom line a local group of eight points has been outlined in color. Notice how the group of points is like a miniature version of the global pattern two lines above it.

The discovery of this recursive regularity, first made on a little calculator by Feigenbaum, is one of the major recent advances in the field. It states in particular that to make line $n+1$ from line n, you simply let each point on line n give birth to "twins". The new generation of points should be packed in about 2.5 times more densely than the old generation was. More exactly stated, the distance between new twins should be α times smaller than the distance between their parent and its twin, where α is a constant, approximately equal to 2.5029078750958928485 . . . This rule holds with greater and greater accuracy the larger n becomes.

What about the values of the Λ's? Are they headed asymptotically toward 1? Surprisingly enough, no. These Λ-values are quickly converging on a particular critical value λ_c, of size roughly 0.892486418 . . . And their convergence is remarkably smooth, in the sense that the distance between successive Λ's is shrinking geometrically. More precisely, the ratio $(\Lambda_n - \Lambda_{n-1})/(\Lambda_{n+1} - \Lambda_n)$ approaches a constant value called δ by Feigenbaum, its discoverer, but more often referred to simply as "Feigenbaum's number" by others. Its value is approximately 4.669201609102990 97 . . .

In short, as λ approaches λ_c, at special λ-values predicted by Feigenbaum's constant δ, *f*'s attractor doubles in population, and its increasingly many elements are geometrically arranged on the *x*-axis according to a simple recursive plan, the main determining parameter of which is Feigenbaum's other constant, α.

Then for λ beyond λ_c—called the *chaotic regime*—the results of iterating *f* can, for some seed values, yield orbits that converge to no finite attractor. These are *aperiodic* orbits. For most seed values, the orbit will remain periodic, but the periodicity will be very hard to detect. First of all, the period will be extremely high. Secondly, the orbit will be much more chaotic than before. A typical periodic orbit, instead of quickly converging to a geometrically simple attractor, will meander all over the interval [0,1], and its behavior will appear indistinguishable from total chaos. Such behavior is termed *ergodic*. Furthermore, neighboring seeds may, within a very small number of iterations, give rise to utterly different orbits. In short, a *statistical* view of the phenomena becomes considerably more reasonable beyond λ_c.

Figure 16-6 beautifully portrays the period-doubling route to chaos, as well as what happens after you've gotten there. The bifurcations are clear to the eye, and since the horizontal distance from each set of them to the next shrinks geometrically, the onset of chaos at λ_c is plainly visible. But the regularity of the structure to the *right* of λ_c—that is, in the chaotic regime—is quite unexpected. It is certain that there are many deep mathematical secrets locked up in this elegant graph.

* * *

Now, what do such novel concepts as the iteration of folded functions, period doubling, chaotic regime, and so on have to do with the study of turbulence in hydrodynamic flow, the erratic population fluctuations in predator-prey relations, and the instability of laser modes? The basic idea is embedded in the contrast between laminar flow and turbulent flow. In a peacefully flowing fluid, the flow is *laminar*—a soft and gentle word that means that all the molecules in the fluid are moving like cars on a multilane freeway. The key features are: (1) that each car follows the same path as its predecessor, and (2) that two nearby cars, whether they are in the same lane or in different ones, will, as time passes, slowly separate from each other—essentially in proportion to the difference in their velocities—which is to say, linearly. These features also apply to molecules of fluid in laminar flow; there, the lanes are called *streamlines* or *laminas*.

By contrast, when a fluid is churned up by some external force, this smooth behavior turns into turbulent behavior, as is seen in breakers at the beach and cream being stirred into coffee. Even the word "turbulent" sounds much harsher and more angular than the soft word "laminar". Here, the image of a multilane freeway no longer holds; the streamlines separate from each other and tangle in the most convoluted of ways, as shown in Figure 16-7. In such systems there are eddies and vortices and all sorts of unnamable whorls on many size-scales at once, and consequently, two points that were initially very close may soon wind up in totally different regions of the fluid. Such quickly diverging paths are the hallmark of turbulence. The distance between points can increase exponentially with time, instead of just linearly, and the coefficient of time in the exponent is called the *Lyapunov number*. When one speaks of chaos in turbulent flow, it is this rapid, nearly unpredictable separation of neighbors that is meant. Such behavior is strikingly reminiscent of the rapid separation, in the chaotic regime of λ, of two orbits whose seeds might originally have been very close together.

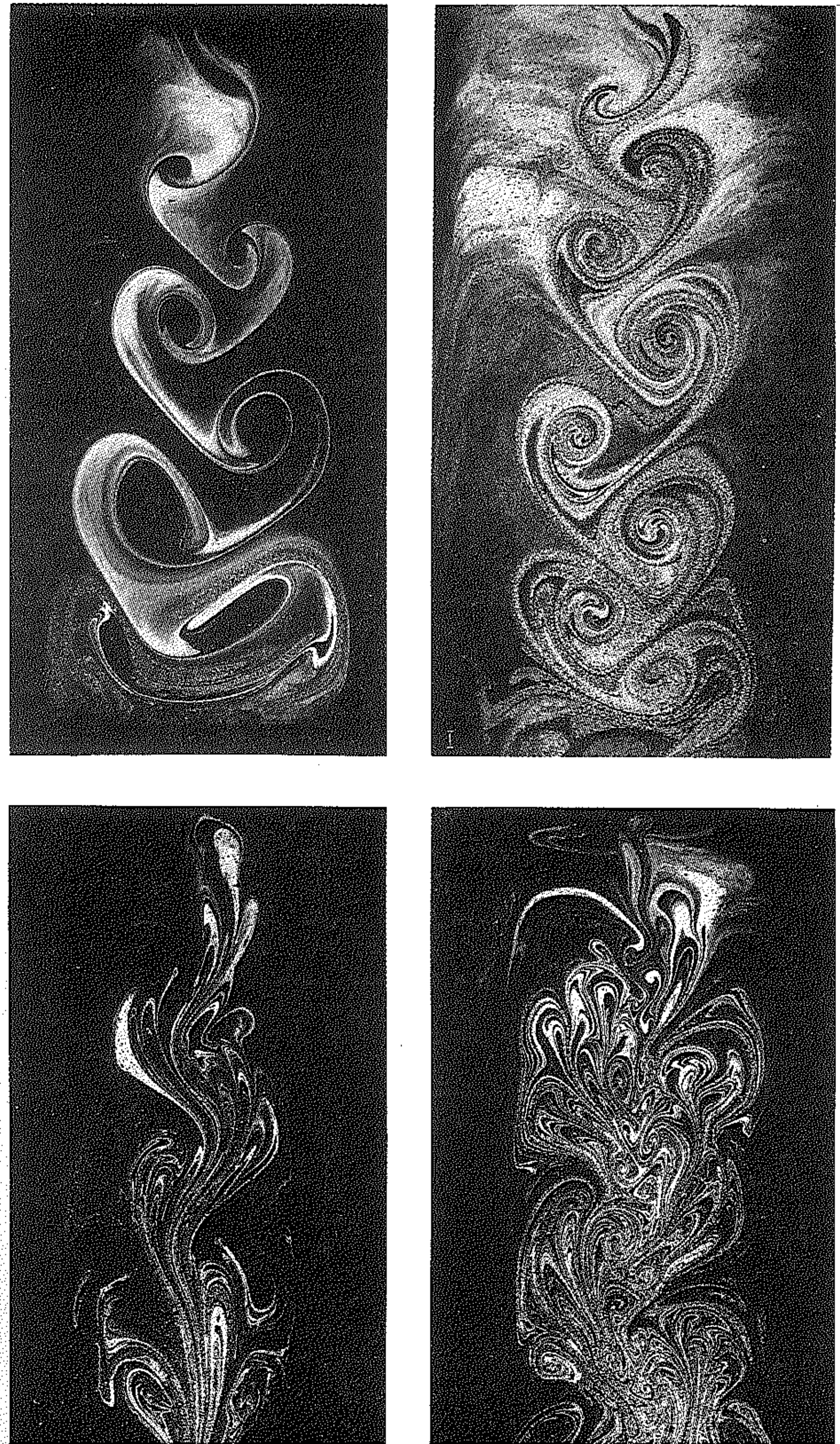


FIGURE 16–7. *Showing the approach to turbulence. In the upper two pictures, a rod was drawn through a viscous liquid once, setting up trains of vortices behind it. In the lower two, the rod was drawn more than once, and the forms are therefore more complicated and recursive-seeming. It is provocative to compare this figure with Figure 13-4.* [*From* Sensitive Chaos, *by Theodor Schwenk.*]

This suggests that the "scenario" (as it is called) by which pretty, periodic orbits gradually give way to the messy, chaotic orbits of our parabolic function might conceivably be mathematically identical to the scenario underlying the transition to turbulence in a fluid or other system. Exactly how this connection is established, though, requires some more detailed setting of context. In particular, we must briefly consider how the spatio-temporal flow of a fluid or some other entity, such as population density or money, is mathematically modeled.

In such real-world problems, the most successful equations yet found to model the phenomena are *differential equations.* A differential equation connects the continuous rate of variation of some quantity to that quantity's current size and the current sizes of other quantities. Moreover, the time variable is itself continuous, not jerking from one discrete instant to the next as some strange clocks and watches occasionally do, but indivisibly flowing, like a liquid. One way to visualize the patterns defined by differential equations is to imagine a multidimensional space—it could have thousands of dimensions, or merely a few—in which a point is continuously tracing out a curve. At any one moment, the single point contains all the information about the state of the physical system. Its projections along the various axes give the values of all the relevant quantities that pin down a unique state. Clearly the space—called *phase space*—would need to have an enormous number of dimensions for a mere point to store the entire shape of a wave breaking on a beach. On the other hand, in a simple predator-prey relation, only two dimensions suffice: one coordinate, say x, giving the predator population and the other, say y, giving the prey population. Two dimensions are more easily visualized, and so we will stick with that case for the time being. The ideas generalize, however, to higher-dimensional cases.

As time progresses, x and y determine each other in an intertwined manner. For example, a large population of predators will tend to reduce the population of prey, whereas a small population of prey will tend to reduce the population of predators. In such a system, x and y constitute a single point (x,y) that swirls around smoothly in a continuous orbit on the plane. (Here the sense of "orbit" is different from the preceding one—that of the discrete, or jumping, orbits we saw when our parabolic function was iterated.) One such possible orbit appears in Figure 16-8; it is generated by a differential equation called "Duffing's equation". It looks like the path of a buzzing fly in your bedroom—or rather, it looks like the *shadow* of the fly's path on a wall. As a matter of fact, this self-intersecting two-dimensional curve *is* the shadow of a non-self-intersecting three-dimensional curve. The motion of a point in phase space must *always* be non-self-intersecting. This arises from the fact that a point in phase space representing the state of a system encodes *all* the information about the system, including its future history, so that there cannot be two different pathways leading out of one and the same point.

In particular, in Duffing's equation there is a third variable, z, that I have

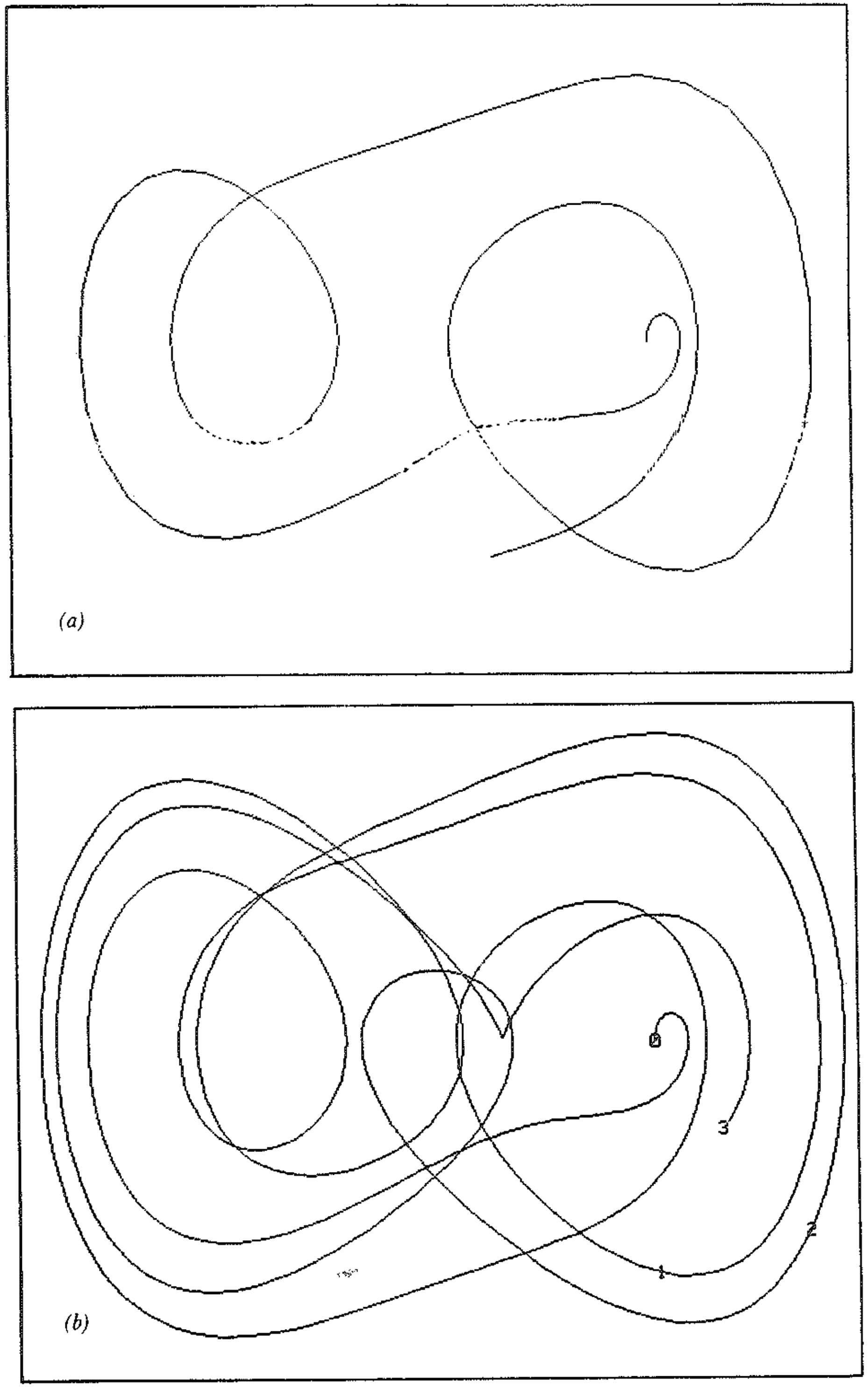


FIGURE 16–8. *If values of* x *and* y *mutually determine each other according to Duffing's equation as time passes, then the point* (x,y) *will trace out a curve* (a). *If a strobe light blinks*

not mentioned so far. If you think of x and y as representing predator and prey populations, then you can think of z as representing a periodically varying external influence, such as the sun's azimuth or the amount of snow on the ground. Now, if you will allow me to mix my buzzing-fly image with the predator-prey example, imagine a bedroom with a fly buzzing periodically back and forth between two walls. Let us say it takes the fly a year to cross the room and come back. (Perhaps it is a rather large bedroom, or maybe just a slow fly.) In any case, as the fly flies, its shadow on one of the two walls traces out the curve shown in Figure 16-8*a*. If the fly ever chances to come back to a point in the room which it has passed through before, it is doomed to loop forever, following the path it took the preceding time over and over again. This gives you a picture of the continuous orbit of a point in phase space representing the state of dynamic system controlled by differential equations.

* * *

Now suppose we wanted to establish some connection of these systems to *discrete* orbits. How might we do so? Well, the values of x, y, and z need not be watched at all moments; they can be sampled periodically, at some natural frequency. In the case of animal populations, a year is the obvious natural period. The sun's azimuth is exactly periodic, and the weather at least *tries* to repeat itself a year later. Thus a natural sequence of discrete points (x_1, y_1, z_1), (x_2, y_2, z_2), . . . can be singled out—one per year. It is as if a strobe light blinked regularly and froze the fly on special annual occasions—perhaps at midnight every Halloween. Or you can think of a firefly that flashes on for just a split second once every year. At all other times its peregrinations around the room are unseen. Figure 16-8*b* shows a sequence of discrete points along the fly-path's shadow, marked by numbers telling when they occurred. Gradually, as many "years" elapse, enough of these discrete points will accumulate that they will start to form a recognizable shape of their own. This pattern of points is a *discrete* "orbit", and so it is closely related to the discrete orbits defined by the iteration of our parabola $f(x)$. In that parabolic case, we had a simple one-dimensional recurrence relation (or an iteration):

$$x_{n+1}=f(x_n).$$

Here we have a *two*-dimensional recurrence:

$$x_{n+1}=f_1(x_n,y_n)$$

$$y_{n+1}=f_2(x_n,y_n)$$

This is a system of *coupled* recurrence relations, in which output values of

produce the $n+1$st generation. On and on it goes, generation after generation. In higher-dimensional cases, of course, there are more such equations. Nevertheless, the skeleton of all these systems remains the same: a multidimensional point $(x_n, y_n, z_n, \ldots)$ jumps from one discrete location in phase space to another, as a discrete variable, n, representing time jumping ahead in discrete units, is incremented.

Notice that we have finessed our way around the continuous time variable that is involved in differential equations. We have done it by focusing on the way the point is connected to its predecessor one "year" earlier (or whatever natural period is involved). But is there always a "natural period" at which to look at a system of mutually intertwined differential equations? Not always. In some situations, however, there is, and this happens to be the case in all situations where turbulent behavior occurs.

Why is this so? All systems that exhibit turbulent behavior are *dissipative*, which means that they dissipate, or degrade, energy from more usable forms such as electricity into the less usable form of heat. In the case of hydrodynamic flow, this dissipation is caused by friction, and in the other systems we have been considering, by abstract analogues of friction. A familiar consequence of friction is that objects in motion will grind to a halt unless energy is pumped in. Now if we "drive" a dissipative system with a *periodic* driving force (you can imagine, for example, stirring a cup of coffee with a spoon in a periodic, circular way), then, of course, the system will not grind to a halt; it will head for some kind of steady state. Such a steady state is a stable orbit—or in our terms, an attractor in phase space. And since we have driven the system with a periodic spoon, we have defined a natural frequency at which to flash our strobe light and freeze the system's state—namely, each time the spoon comes swinging around and passes some fixed mark on the cup, such as its handle. This will constitute our "year". In this way, continuous time can be replaced by a series of discrete instants, as long as we are dealing with a dissipative system driven by a periodic force. And so continuous orbits can be replaced by discrete orbits, which brings iteration back into the picture.

If the driving force itself has no natural period (it may be simply a constant force), there is still a way to define a natural period, as long as some variable in the system swings back and forth between extremes. Just flash your strobe whenever that variable hits its extreme value, and the fly will still be caught at discrete instants. This type of discrete representation of the fly's motion in a multidimensional space is called a *Poincaré map*.

This stirring argument is only hand-waving, of course, and needs much more rigor to be convincing to a mathematician. It nonetheless gives the flavor of how the study of a set of coupled differential equations can be replaced by the study of a set of coupled discrete recurrence relations. This is the vital step that brings us back to the recent discoveries about the parabola.

* * *

In 1975, Feigenbaum discovered that his numbers α and δ do not depend on the details of the shape of the curve defined by $f(x)$. Almost any smooth convex shape that peaks in the same spot will do as well. Inspired by the structural universality discovered by Metropolis, Stein, and Stein, Feigenbaum tried working with a sine curve instead of a parabola. He was flabbergasted by the reappearance of the same numerical values, to many decimal places, of the numbers α and δ, which had characterized the period-doubling and the onset of chaos for the parabola. For the sine curve just as for the parabola, there is a height-parameter λ and a set of special λ-values that converge to a critical point λ_c. Moreover, the onset of chaos at λ_c is governed by the same numbers α and δ. Feigenbaum began to suspect that there was something universal going on here. In other words, he suspected that what is more important than f itself is the mere fact that f is being iterated over and over. In fact, he suspected that f itself might play no role in the onset of chaos.

It is not quite that simple, in reality. Feigenbaum soon discovered that what *does* matter about f is just the nature of the peak at its very center. The long-term behavior of orbits depends only on an infinitesimal segment at the crest of the graph, and ultimately, it depends only on the behavior at the very point where the maximum occurs! The rest of the shape, even the region close to the peak, is irrelevant. A parabola has what is called a *quadratic* maximum, as do a sine wave, a circle, and an ellipse. In fact, the behavior of a randomly-produced smooth function at a typical maximum would be expected to be of the quadratic type, in the absence of any special coincidences. So the parabolic case, rather than being a quirky exception, begins to seem like the rule. This empirical discovery by Feigenbaum, involving two fundamental scaling factors α and δ that characterize the onset of chaos through period-doubling attractors, represents a new kind of universality, known as *metrical* universality, to distinguish it from the earlier-known *structural* universality. This empirically demonstrated metrical universality was later proved to be correct (in the more orthodox sense of proof) in the one-dimensional case by Oscar Lanford.

A truly exciting development occurred when Feigenbaum's constants unexpectedly turned up in some messy models of actual physical systems that exhibit turbulence, not just in pretty and idealized mathematical systems. Valter Franceschini of the University of Modena in Italy adapted the Navier-Stokes equation, which governs all hydrodynamic flow, for computer simulation. To do so, he turned it into a set of five coupled differential equations whose Poincaré maps he could then study numerically on his computer. He first found that the system exhibited attractors with repeated period-doubling as its governing parameters approached the values where turbulence was expected to set in. Unaware of Feigenbaum's work, he showed his results to Jean-Pierre Eckmann of the University of Geneva, who immediately urged him to go back and determine the rate of convergence of the λ-values at which period-doubling occurred. To their

amazement, Feigenbaum's α- and δ-values—accurate to about four decimal places—appeared seemingly out of nowhere! For the first time, an accurate mathematical model of true physical turbulence revealed that its structure was intimately related to the humble chaos lurking in the humble parabola $y=4\lambda x(1-x)$. Subsequently, Eckmann, Pierre Collet, and H. Koch showed that in the behavior of a multidimensional driven dissipative system, all dimensions but one tend to drop out after a sufficiently long period of time, and so one should *expect* the characteristic of one-dimensional behavior—namely Feigenbaum's metrical universality—to reappear.

Since then, experimentalists have been keeping their eyes peeled for period-doubling behavior in actual physical systems (not just in computer models). Such behavior has been observed in certain types of convective flow, but so far the measurements are too imprecise to lend very strong support to the idea that the parabola contains the clues revealing the nature of genuine physical turbulence. Still, it is tantalizing to think that somehow, all that really matters is that a dissipative set of coupled recurrence relations is being iterated—but that the detailed properties of those recurrences can be entirely ignored if one is concentrating on understanding the route to turbulence.

Feigenbaum puts it this way. One often sees a pattern of clouds in the sky—a celestial trellis composed of a myriad of small white puffs stretching from horizon to horizon—that clearly did not happen "by accident". Some systematic hydrodynamic law has got to be operating. Yet, says Feigenbaum, it must be a law operating at a higher level, or on a larger scale, than the Navier-Stokes equation, which is based on infinitesimal volumes of fluid and not on large "chunks". It seems that in order to understand such beautiful sky patterns, one must somehow bypass the *details* of the Navier-Stokes equation, and come up with some coarser-grained but more relevant way of analyzing hydrodynamic flow. The discovery that iteration gives rise to universality—that is, independence of the details of the function (or functions) being iterated—offers hope that such a view of hydrodynamics may be well on its way to emerging.

* * *

Well, we have covered attractors and turbulence; what about *strange* attractors? We have now built up the necessary concepts to understand this idea. When a periodically driven two-dimensional (or higher-dimensional) dissipative system is modeled by a set of coupled iterations, the successive points lit up by the flashes of the periodic strobe light trace out a shape that plays the role, for this system, that a simple orbit did for our parabola. But when one is operating in a space of more than one dimension, the possibilities are richer. Certainly it is possible to have a stable fixed point (an attractor of period one). This would just mean that at every flash of the strobe, the point representing the system's state is exactly where it was last time. It is also possible to have a *periodic* attractor: one where after some finite number of flashes, the point has returned to a preceding position. This would be analogous to the 2-cycles, 4-cycles, and so on that we saw occurring for the parabola.

But there is another option: that the point never returns to its original position in phase space, and that successive flashes reveal it to be jumping around quite erratically inside a restricted region of phase space. Over a period of time, this region may take shape before an observer's eyes as the strobe flashes periodically. In the majority of such cases so far studied, a most unexpected phenomenon has been observed to take place: the erratically jumping point gradually creates a delicate filigree that recalls the "faint fantastic tracery made by frost on glass". (I owe this poetic image to the American critic James Huneker, who used it to describe the magical effect of one of Chopin's piano études: Op. 25, No. 2—see Chapter 9.) The delicacy is of a rather specific kind, closely related to the "fractal" curves described by Benoît Mandelbrot in his book *The Fractal Geometry of Nature.* In particular, any section of such an attractor, when blown up, reveals itself to be just as exquisitely detailed as was the larger picture from which it was taken. In other words, there is an infinite regress of detail, a never-ending nesting of pattern within pattern. One of the earliest of such structures to be found, called the *attractor of Hénon,* is shown in Figure 16-9. It is generated by the sequence of points (x_n,y_n) defined by the following recurrence relations:

$$x_{n+1}=y_n-ax_n^2-1$$

$$y_{n+1}=bx_n$$

Here, a is equal to 7/5 and b to 3/10; the seed values are $x_0=0$ and $y_0=0$. The small square in Figure 16-9*a* is blown up in Figure 16-9*b* to reveal more detail, and then another square in Figure 16-9*b* is blown up in Figure 16-9*c* to reveal yet finer detail. Note that what we appear to have is a sort of three-lane highway each of whose lanes breaks up, when magnified, into more parallel lanes, the outermost of which is a new three-lane highway—and on and on it goes. Any perpendicular cross-section of this highway would be what is called a "Cantor set", formed by a simple and famous recursive process.

Begin with a closed interval, say [0,1]. ("Closed" means that the interval includes its endpoints.) Now eliminate some open central subinterval. (Since an open subinterval does not include its endpoints, those two points will remain in the Cantor set being constructed before your eyes.) Usually the deleted subinterval is chosen to be the middle third (1/3, 2/3), but this is not necessary. Two closed subintervals remain. Subject them to the same kind of process—namely, eliminate an open central subinterval inside each of them. Repeat the process *ad infinitum.* What you will be left with at the end of your infinite toil will be a delicate structure consisting of isolated

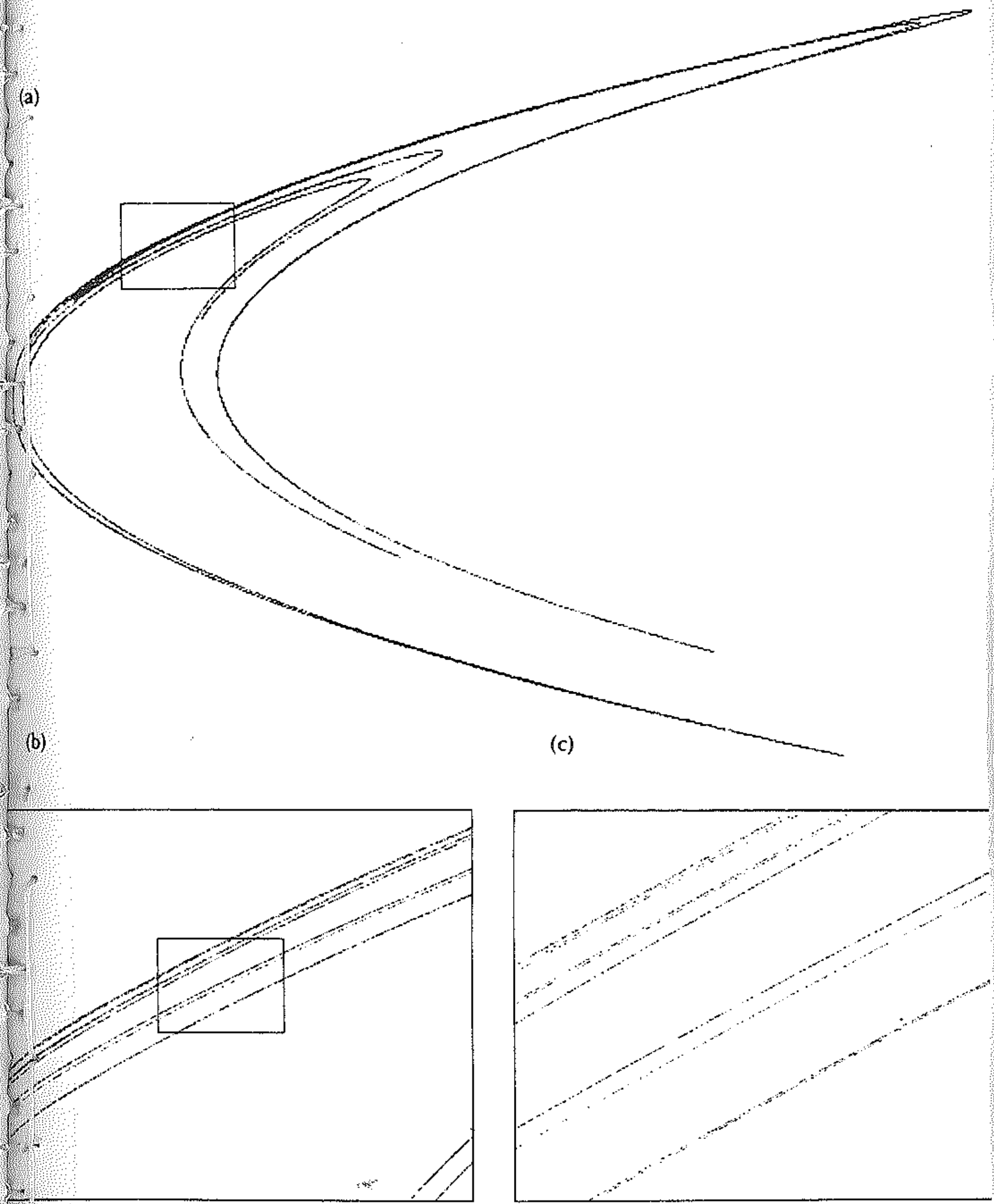


FIGURE 16–9. *The attractor of Hénon: a strange attractor. In (a), the full curve is shown. In (b), the boxed region of (a) is blown up to reveal hidden details. In (c), the boxed region of (b) is further blown up to reveal yet more deeply hidden details. And on and on it could go,* ad infinitum.

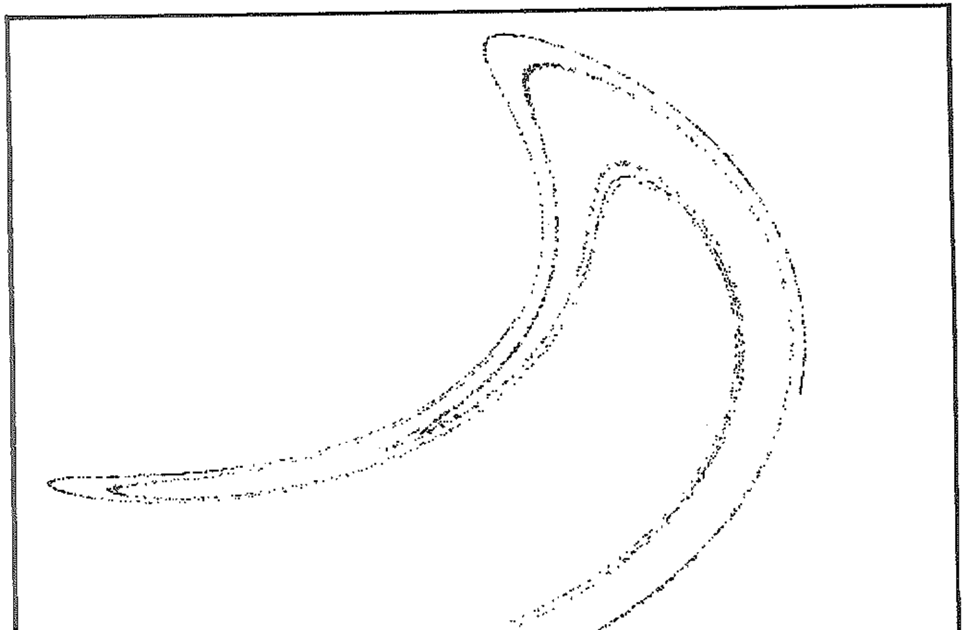
a wire. Their number, however, will be uncountably infinite, and their density will depend on the details of your recursive elimination process. Such is the nature of a Cantor set, and if an attractor's cross-sections have this weird kind of distribution, the attractor is said to be *strange,* and for good reason.

Another beautiful strange attractor is generated by the "stroboscopic" points 0, 1, 2, ... in Figure 16-8*b*. Since this pattern comes out of Duffing's equation, it is called "Duffing's attractor", and it is shown in a slightly expanded scale in Figure 16-10. Notice its remarkable similarity to the attractor of Hénon. Could this be universality showing its face again?

It is interesting that for the parabola, at the critical value λ_c, f's attractor suddenly consists of infinitely many points, since it is the culmination of an infinite sequence of bifurcations. You can visualize this set either as the limiting case of the horizontal point-sets in Figure 16-5, or as the vertical point-set belonging to $x = \lambda_c$ in Figure 16-6. The precise scatter-pattern of this uncountable point-set is determined by Feigenbaum's recursive rule involving his constant α. Given its recursive genesis, it seems probable that this particular attractor is a Cantor set. Hence the fertile parabola has provided us with an example of a *one-dimensional* strange attractor!

In the chaotic regime of the more general k-dimensional case, long-term prediction of the path that a point will take is quite impossible. Two nearly

FIGURE 16–10. *The strange attractor that emerges from a Poincaré map of Duffing's equation.*



touching points on a strange attractor will, after a few blinks of the strobe light, have wound up at totally different places. This is called *sensitive dependence on initial conditions* and is another defining criterion of a strange attractor.

* * *

At present, no one knows just why, how, or when strange attractors will crop up in the chaotic regimes of iterative schemes representing dissipative physical systems, but they do seem to play a central role in the mystery of turbulence. David Ruelle, one of the prime movers of this whole approach to turbulence, wrote: "These systems of curves, these clouds of points, sometimes evoke galaxies or fireworks, other times quite weird and disturbing blossomings. There is a whole world of forms still to be explored, and harmonies still to be discovered."

Robert M. May, a theoretical biologist, concluded his now quite famous review article covering the field in 1976 with a plea that I find most apt and would like to repeat:

> I would urge that people be introduced to the equation $y = 4\lambda x(1-x)$ early in their mathematical education. This equation can be studied phenomenologically by iterating it on a calculator, or even by hand. Its study does not involve as much conceptual sophistication as does elementary calculus. Such study would greatly enrich the student's intuition about nonlinear systems.
>
> Not only in research but also in the everyday world of politics and economics, we would all be better off if more people realized that simple nonlinear systems do not necessarily possess simple dynamical properties.

Post Scriptum.

Stanislaw Ulam, a uniquely inventive mathematician and a warm and delightful human being, died as I was working on this series of postscripts. I had the good fortune to get to know Stan Ulam and his French-born wife Françoise in the summer of 1980, when I visited Santa Fé and stayed with them for a few days. I had always admired and felt kinship with Ulam's strange style in mathematics, totally driven by a passion for the the quirky and the unpredictable, bored by the pure and regular. Ulam loved more than anything to find total chaos in the midst of pristine order. Of course, the thrill was in knowing that there was some kind of *law* to this chaos, so that in *reality*—that is, in God's eye—there was simply a deeper kind of order underneath it all. The bizarre yet tight connections between randomness and order are what all of Ulam's greatest discoveries are about. His style was iconoclastic, to be sure. He was perfectly able to do mathematics in the classical "theorem-proof-theorem-proof" way, but he delighted in the experimental approach, using computers to study crazy behaviors of oddball functions he dreamt up. In some sense, Ulam was a genuine mathematical artist, unlike so many mathematicians. A piece of math by Ulam often feels much more like a creation than like a discovery. It is more idiosyncratic, more easily recognizable as the product of a particular mind, than most mathematical discoveries are.

Aside from being fascinated by mathematics itself, Ulam was also fascinated by the human mind's workings, and he strove to express his vague but provocative intuitions in his writings. I always think of his "ten dogs" theory of memory. The idea is this: When you are searching for a memory that eludes you but that you know is there, what you in effect do is release ten "dogs" in your brain and let them go "sniffing" in parallel. Each dog will start to rummage around here and there, sometimes going in circles, sometimes smelling down wrong alleys, but since there are a bunch of them, you can afford to let them smell out many false pathways. They don't need to be very bright; they just need to have had a whiff of the original idea, and they will follow that spoor high and low. Eventually, it is likely that one dog or another will trot home carrying the desired memory in its mouth. Ulam's autobiography, *Adventures of a Mathematician,* is packed with such glimmerings about how minds work, as well as with droll anecdotes about many of this century's most brilliant mathematicians.

Ulam was very curious about language. He and his wife came to this country about 50 years ago, and both loved the English language. But whereas Stan never lost his strong Polish accent and constantly made errors in English, Françoise eliminated almost every trace of her French accent and became a virtually flawless speaker, whose mastery of idiomatic phrases exceeded that of most native speakers. This caused some amusing light-hearted bickerings between them that I witnessed. Françoise one day used some baseball idiom such as "he threw them a curve ball" or "in the ball park", and Stan immediately objected, saying "You can't use that expression! You didn't grow up playing baseball, so you don't really know what it means!" Françoise defended herself, saying that she had a good idea of its literal meaning but that in any case Stan's point was a red herring. I bought her argument lock, stock, and barrel. After all, how many native speakers of English know what domains such phrases as "red herring" or "lock, stock, and barrel" come from? Yet we certainly all use many such phrases and feel perfectly entitled to do so.

Like many of the brightest mathematicians and physicists working during and just after World War II, Stan Ulam got involved in military projects. His invention, with John von Neumann, of the Monte Carlo method was a key element in the development of the hydrogen bomb. The same forces that drove him to wonder about the cardinality of abstrusely defined sets and the dimensionality of peculiarly defined spaces also guided him to accurate ways of modeling the statistics of chain reactions. At the time he did the work, the nature of the dilemma it would lead humanity as a whole into was not so clear as it now is. To be sure, Einstein had warned us about our slow drift into unparalleled peril, but few people had Einstein's clarity of vision. One of the paradoxes about people is that they are so small compared to the events they can be involved in. Stan Ulam was an ant in a vast colony, and though his role was more significant than that of most ants, he still had no control over the nature of the colony itself. Human nature is one thing, but *humanity*'s nature is another thing.

A good and generous person like Stan Ulam can still be a part of a bad and frightful thing like the arms race. Clearly Ulam had many afterthoughts about his role in these developments, and it is to his credit that he tried to think it all through rationally. Others in similar positions have been far more trapped and narrow-minded, unable to see, or to admit seeing, the complex tragedy that has been unfolding as a consequence of their small actions joined with the small actions of many, many others.

For me it was a privilege to get to know and be friends with this warm and insightful man. I hope that in the long run, Stan Ulam's contributions to mathematics will prove to have outweighed his contributions to a potential Armageddon.

* * *

One of the basic themes of this column is what I call *locking-in.* For no particular reason, I failed to use that term in the column, but it is a good term. The imagery I wish to convey is that of a system that seeks and gradually settles into its own most stable states, and the mechanism whereby it seeks and attains such loci of stability is *feedback.* A system that locks into a state is in a stable equilibrium, which means that if you perturb it somehow, it will swiftly return to the state it was in—there are restoring forces that push it back. Perhaps the most primordial image is that of the particle in the potential well—for example, a marble sitting at the bottom of a round dish. If you ping it lightly with your finger, it will oscillate for a while, but eventually will come to rest again just where it was before: at the sole stable fixed point of the system. Here, as in the column, "fixed point" means that the system's "output" at time t (namely, the marble's position at time t) is identical to the "input" at time $t-1$ (namely, the marble's position at time $t-1$). In this case, the attractor is a single point in space, so it is ridiculously easy to visualize. Most of the attractors in the chapter, however, were *orbits* rather than single points, so they are slightly more abstract. However, if you think of an orbit as simply a point in a multidimensional space, then the concept of zeroing in on a fixed point and the concept of settling down in a stable orbit merge somewhat.

One of the most intuitive as well as charming examples of locking-in is the search for a solution to Raphael Robinson's puzzle in Chapter 2:

In this sentence, the number of occurrences of 0 is __, of 1 is __, of 2 is __, of 3 is __, of 4 is __, of 5 is __, of 6 is __, of 7 is __, of 8 is __, and of 9 is __.

One way to search for a solution to this puzzle is to fill in the blanks with an arbitrary sequence of ten numbers, such as $<0,1,2,3,4,5,6,7,8,9>$, and see what happens when you check out the truth of the resulting sentence. It turns out actually to have two occurrences of each digit. Thus the vector $<0,1,2,3,4,5,6,7,8,9>$ leads to the vector $<2,2,2,2,2,2,2,2,2,2>$ by the process we'll call "Robinsonizing". Where does *that* vector lead? Clearly to $<1,1,11,1,1,1,1,1,1,1>$, which leads to $<1,12,1,1,1,1,1,1,1,1>$, which leads to $<1,11,2,1,1,1,1,1,1,1>$, which leads to $<1,11,2,1,1,1,1,1,1,1>$ —and lo and behold, we've entered a closed loop!

This vector $<1,11,2,1,1,1,1,1,1,1>$ is like a whirlpool or a vacuum cleaner: it sucks things near to it into its vortex. It is a trap, a fixed point —an attractor. It is not unique; there is another such vortex, which I will leave it to you to find. Furthermore, there is at least one two-state loop, or period-two attractor, that I know of. I have reason to suspect that *everything* leads to one of those three attractors, but I could be wrong. You could search for a period-two attractor by writing down a vector of length twenty and generating its successor length-twenty vector as follows: Let the new vector's first half be derived from the old one's second half by Robinsonizing, and let the new one's second half be derived from the old one's first half by Robinsonizing. If you now iterate this double-barreled Robinsonizing operation starting with a random seed, you will eventually settle down on a fixed point.

Notice that we are now calling a period-two attractor a "fixed point". Notice also that this is a "point" in a twenty-dimensional space! The point is, we can view the system *either* as bouncing back and forth between two ten-dimensional points (a period-two attractor) *or* as sitting still on a fixed twenty-dimensional point. If by chance there were a loop of length four, we could similarly think of it as being a fixed point in a 40-dimensional space. As long as we're willing to "up" the dimensionality of the space, we can store more and more information in a single point. Thus fixed points and stable orbits are very close concepts.

* * *

This example serves to illustrate how feedback—plugging the system's output back into the system as input—ushers you to the fixed points. Why should this be so? Why could the system not thrash about randomly, somehow avoiding all fixed points? In short, why are fixed points so often attractive? Why could there not be a large number of fixed points that are totally isolated, like islands in a vast sea, unreachable via any obvious route? Could there not be fixed-point "anti-whirlpools" that repel any approacher

that is not dead on target? In the case of Robinson's puzzle, the answer is *no*; but there are such systems. Indeed, in the column I pointed out how there are repellent as well as attractive fixed points for functions of the form $4\lambda x(1-x)$. But in general, it seems to be a very good rule of thumb to search for fixed points by starting out somewhere at random and then hoping that you will get sucked into a stable orbit. Most likely you will, and you will thereby discover a locus of stability, a locked-in solution.

Even more remarkable, it seems generally reliable that you are more likely to be sucked into a short loop than a long one, if short ones exist. Thus, generally speaking, the *stablest* behavior of a system seems also to be its *simplest* behavior. This is true for systems of nearly any sort one can imagine. In the hydrogen atom, for instance, the ground state—the lowest-energy state—is spherically symmetric, and is the only one to have that simple property. Why should this be so, all across the board? Why are stable things the simplest things as well? Or, conversely, why are the simplest things the stablest of all? A toughie.

* * *

A puzzle more complex than Robinson's but similar in flavor is the search for self-documenting or self-inventorying sentences, which was carried out with such great gusto by Lee Sallows (see Chapter 3). His "logological rocket" was a machine for seeking attractive fixed points in a certain logological space. The book *Loopings* by Aldo Spinelli is a remarkable investigation of regions of a similar logological space, and his search is guided by the same old principle: that starting somewhere random and relying on feedback to get you somewhere "better" is the most likely way to discover a fixed point. This is a most strange way of looking for what might seem something elusive and precious, yet strange though it might be, it is very robust.

In Chapter 3's *Post Scriptum,* I stated that I felt Lee Sallows was overconfident in wagering that a computer search for a self-documenting sentence beginning "This computer-generated pangram contains . . ." would not succeed in ten years. The reason is simple. Lee did not consider the idea of "iterative convergence" to a solution—that is, the idea of Robinsonizing, applied to self-descriptive sentences. You begin with a sentence of the right form, but where all the numbers are randomly chosen. It's a blatant lie, but who cares? You just feed it to a program that counts all its letters and spits out a new sentence with the new letter-counts replacing the old guesses. Around and around you go . . . It is almost certain that you will pretty soon fall into an attractive orbit. Probably most orbits are fairly lengthy loops, and thus do not yield self-documenting sentences—but again, who cares? Just try it again with a different random seed, and keep on doing so, until you find a fixed point.

This method may sound too simple, but it works. I suggested it to Bob

French, one of the two translators into French of *Gödel, Escher, Bach,* and he was gung-ho about implementing such a program. Within a short time, he had one up and running. He sent me this note about his discoveries:

> I wrote a nice program to solve the Pangram Problem and got an answer, written, much to my annoyance, in "françaix". It is:
>
> *Cette phrase contient cinq a, cinq c, troix d, douze e, un f, un g, quatre h, treize i, huit n, six o, troix p, six q, huit r, six s, quatorze t, dix u, un v, sept x, & quatre z.*
>
> Unbelievably, in programming it, I had put the *wrong* goddam spelling of "trois" into the program. Oh well, when I corrected the mistake, I didn't get an answer immediately, but I'm confident that it'll come, in correctly spelled French, when I get back to work on the thing.

The point is, you don't need to perform a brute-force search through the entire space of all possible combinations of numbers filling the 26 blanks in order to find a perfect self-documenting sentence, not by a long shot! A Robinsonizing routine, together with a simple-minded loop detector, will do the trick quite easily, as long as you're willing to try a bunch of different seeds. The pulling-power of short loops will undoubtedly snag you sooner or later, and you'll have found your target sentence!

My friend Larry Tesler, equally spurred on by Sallows' challenge when it appeared in print in A. K. Dewdney's new *Scientific American* column called "Computer Recreations" in October 1984, coded up the Robinsonizing method in a program and soon his computer fell into a loop that seemed very close to a solution. By changing his program's search technique at that point, Tesler was then easily able to home in on a winner, which he gleefully sent off to both Dewdney and Sallows. Tesler's sentence runs as follows:

> This computer-generated pangram contains six a's, one b, three c's, three d's, thirty-seven e's, six f's, three g's, nine h's, twelve i's, one j, one k, two l's, three m's, twenty-two n's, thirteen o's, three p's, one q, fourteen r's, twenty-nine s's, twenty-four t's, five u's, six v's, seven w's, four x's, five y's, and one z.

* * *

Locking-in is perfectly illustrated by the hypothetical book *Reviews of This Book,* described in Chapter 3. There I characterized the method of its creation as resembling the construction of "self-consistent" solutions via the "Hartree-Fock" method. What does that mean? It boils down to the same thing once more. It turns out to be very hard—in fact, impossible—to give closed-form solutions to the equations describing any atom more complicated than a hydrogen atom, with its single electron. When you have three bodies, as in the helium atom with its two electrons and a nucleus, the mathematical complexity is overwhelming. The problem is in essence that

each electron would "like" to be in a simple hydrogen-like state around the nucleus, but the other one is blocking it from so doing. How can they "cooperate" with each other to find a stable mode of coexistence?

One way to study this mathematically, suggested first in 1928 by the English physicist Douglas Rayner Hartree, is to try to converge on a good description of the total system by starting out with a false solution—a mathematical description of a state known to be wrong, but easy to describe. (For instance, you could pretend that both electrons *are* in simple hydrogen-like states.) Then you see how each electron "perturbs" the other one out of the presumed state it was in. This leads you to a different—and probably no less fictitious—state. But at least you've made progress, in that you've taken into account the "first-order" effects each electron would have on the other one. Now you do the same thing over again—that is, you see how the perturbed states would perturb each other. This gives you "second-order" corrections—and so on and so on. Eventually—and this is the beauty of the method—the starting point of your calculations gets totally buried, and the state converges to what is called a "self-consistent" solution, very much like the solutions to Robinson's puzzle. What I mean by saying the starting point gets "buried" is that *no matter where you start,* you'll wind up at the same eventual solution—a fixed point, where further iteration has no effect. In this solution, the two electrons are in equilibrium with each other and do not perturb each other. And presto—one has "solved" the helium atom!

Of course, this type of solution is *numerical,* not analytic: there are no exact formulas that come out, only numbers. Nonetheless, that's good enough for most practical purposes. The Russian physicist Vladimir Fock later made a suggestion for improving the validity of this method of calculation, which involves taking into account the fact that electrons obey the Pauli exclusion principle, a complication that Hartree had ignored. That is the reason for the hyphenated name; however, Hartree is the inventor of the general principle of calculating self-consistent solutions for many-body systems.

* * *

This idea of locking-in recurs throughout science. In *Gödel, Escher, Bach,* I discussed the phenomenon called *renormalization*—the way that elementary particles such as electrons and positrons and photons all take each other into account in their very core. The notion is a mathematical one, but for a good metaphor, recall how your own identity depends on the identities of your close friends and relatives, and how theirs in turn depends on yours and on *their* close friends' and relatives' identities, and so on, and so on. This was the image I described for "I at the Center" in the *Post Scriptum* to Chapter 10. Another good graphic representation of this idea is shown in Figure 24-4, where identity emerges out of a renormalization process.

The tangledness of one's own self is a perfect metaphor for understanding what renormalization is all about. And the best way to imagine how *you* emerge from such a complex tangle is to begin by imagining yourself as a "zeroth-order person"—that is, someone totally unaware and inconsiderate of all others. (Of course, such a person would be barely a person, barely a self at all: a perfect baby.) Then imagine how "you" would be modified if you started to take other people into account, always considering others as perfect babies, or zeroth-order people. This gives a "first-order" version of you. You are beginning to have an identity, emerging from this modeling of others inside yourself. Now iterate: second-order people are those who take into account the identities of first-order people. And on it goes. The final result is *renormalized* people: people who take into account the identities of renormalized people. I know it sounds circular, and indeed it is, but *paradoxical* it is not—at least no more than are the fixed points of Raphael Robinson's puzzle! "Circular" is not synonymous with "paradoxical", although many people mistakenly assume it is. We shall re-encounter this notion of renormalized people in Chapter 30 and beyond, where it will in fact clear up some seeming paradoxes involving cooperation and egoism.

This close connection of locking-in to the deepest essence of personhood plays a central role also in Chapters 22 and 25, where "who" one is is portrayed as emerging from a "level-crossing feedback loop", in which a sophisticated perceiving system perceives limited aspects of its own nature, and by feeding them back into the system creates a type of locking-in. The locked-in loop itself is given a name, and that name, for every such system, is "I".

The idea of a system with an I, watching its own behavior, is closely related to the wellsprings of creativity (recall the cycle underlying creativity discussed in the *Post Scriptum* to Chapter 12, and that to Chapter 10 as well). We will delve into this in depth again in Chapter 23, trying to come to grips with another seeming paradox: that of mechanizing what seems by definition to be nonmechanical and nonmechanizable—the creative act. Once again we'll see vicious paradox dissolve into benign cycles.

In short, locking-in—that is, convergent and self-stabilizing behavior—will surely pervade the ultimate explanation of most mysteries of the mind. One example is the question of memory retrieval. How do things that are only vaguely similar to each other stir up rumblings of recollection, and eventually trigger the retrieval of amazingly deep abstract resemblances? One theory, best formulated and articulated by cognitive scientist Pentti Kanerva of Stanford University, sees the initial input as a *seed*—a vector in a very high-dimensional space, analogous to the seed vector that we fed into the Robinsonizing machine. The seed is fed into memory-retrieval mechanisms, which convert it into an output vector that is then fed back in again. This cyclic process continues until it either converges on a stable fixed point—the desired memory trace—or is seen to be wandering erratically without any likelihood of locking in, tracing out a chaotic

sequence of "points" in mind-space. The details of how this is accomplished in Kanerva's beautiful theory are beyond the scope of this book, but this "self-propagating search" provides another remarkable example of the many ways that locking-in can be exploited.

Closely related to memory retrieval is the problem of perception, or pattern recognition. As I mentioned in the *Post Scriptum* to Chapter 4, this central aspect of mind has been best modeled on computers in programs whose strategy is similar to that of Kanerva's model: there is a superficial sweep that narrows the field somewhat, followed by a deeper sweep that narrows it further, and so on (the "terraced scan" I described in the postscript to Chapter 5). This bottom-up processing is complemented by concurrent top-down processing driven not by the input, but by *expectations* of what is "out there" to be recognized. The swirling activity in which bottom-up and top-down processes seek a reconciliation with each other leads to a gradual kind of "crystallization", in which many small pieces of evidence align with, and mutually reinforce, each other. The ultimate justification for some of them resides, of course, in the raw perceptual input, while for others of them it resides in the richness of previous experiences stored in memory. The combination of all these mutually confirming hypotheses results in a globally optimal interpretation of the input: an act of recognition. Once again, locking-in carries the day.

One final example of locking-in is the subject of Chapter 27: the question of the inevitability (or evitability) of the genetic code. This central question about the molecular foundations of life turns out to revolve about two distinct senses of the word "arbitrary". I shall let that Chapter speak for itself, however.

* * *

In the Introduction, I described the space of my columns as gradually emerging as, month by month, I revealed one more dot in that space. What is this, if not a Poincaré map of my mental meanderings? During my column-writing era, my mind would light up like a monthly firefly and reveal where it was to the outside world! I just wonder: Would the shape I was thus tracing out turn out to be a strange attractor?

It seems appropriate that at this midpoint of the book, we have identified a unifying theme—or rather, *thema,* to be more faithful to the title. *Locking-in* seems to be a key to the metamagics of Snarls, of Society, of Slipping . . . of Strangeness, of Substrate, of Stability . . . of Survival.