EXCERPTED FROM

STEPHEN WOLFRAM

A NEW KIND OF SCIENCE

SECTION 7.4

Chaos Theory and Randomness from Initial Conditions

perfect randomness. And the reason for this is that there are almost inevitably correlations even in the supposedly random environment.

In an ocean for example, the inertia of the water essentially forces there to be waves on the surface of certain sizes. And during the time that a boat is caught up in a particular one of these waves, its motion will always be quite regular; it is only when one watches the effect of a sequence of waves that one sees behavior that appears in any way random.

In a sense, though, this point just emphasizes the incomplete nature of the mechanism for randomness that we have been discussing in this section. For to know in any real way why the motion of the boat is random, we must inevitably ask more about the randomness of the ocean surface. And indeed, it is only at a fairly superficial level of description that it is useful to say that the randomness in the motion of the boat comes from interaction with an environment about which one will say nothing more than that it is random.

Chaos Theory and Randomness from Initial Conditions

At the beginning of this chapter I outlined three basic mechanisms that can lead to apparent randomness. And in the previous section I discussed the first of these mechanisms—based on the idea that the evolution of a system is continually affected by randomness from its environment.

But to get randomness in a particular system it turns out that there is no need for continual interaction between the system and an external random environment. And in the second mechanism for randomness discussed at the beginning of this chapter, no explicit randomness is inserted during the evolution of a system. But there is still randomness in the initial conditions, and the point is that as the system evolves, it samples more and more of this randomness, and as a result produces behavior that is correspondingly random.

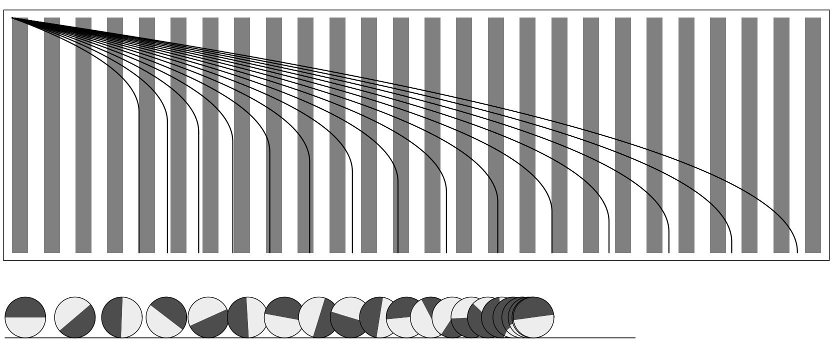
As a rather simple example one can think of a car driving along a bumpy road. Unlike waves on an ocean, all the bumps on the road are already present when the car starts driving, and as a result, one can consider these bumps to be part of the initial conditions for the system. But the point is that as time goes on, the car samples more and more of the bumps, and if there is randomness in these bumps it leads to corresponding randomness in the motion of the car.

A somewhat similar example is a ball rolled along a rough surface. A question such as where the ball comes to rest will depend on the pattern of bumps on the surface. But now another feature of the initial conditions is also important: the initial speed of the ball.

And somewhat surprisingly there is already in practice some apparent randomness in the behavior of such a system even when there are no significant bumps on the surface. Indeed, games of chance based on rolling dice, tossing coins and so on all rely on just such randomness.

As a simple example, consider a ball that has one hemisphere white and the other black. One can roll this ball like a die, and then look to see which color is on top when the ball comes to rest. And if one does this in practice, what one will typically find is that the outcome seems quite random. But where does this randomness come from?

The answer is that it comes from randomness in the initial speed with which the ball is rolled. The picture below shows the motion of a ball with a sequence of different initial speeds. And what one sees is that it takes only a small change in the initial speed to make the ball come to rest in a completely different orientation.



A plot of the position of a ball rolled with various initial speeds. Time goes down the page. The ball starts on the left, with an initial speed given by the initial slope of the curve. The ball slows down as a result of friction, and eventually stops. The ball is half white and half black, and the stripes in the picture indicate which color is on top when the ball is at a particular position. The divergence of the curves in the picture indicate the sensitivity of the motion to the exact initial speed of the ball. Small changes in this speed are seen to make the ball stop with a different color on top. It is such sensitivity to randomness in the initial conditions that makes processes such as rolling dice or tossing coins yield seemingly random output.

The point then is that a human rolling the ball will typically not be able to control this speed with sufficient accuracy to determine whether black or white will end up on top. And indeed on successive trials there will usually be sufficiently large random variations in the initial speed that the outcomes will seem completely random.

Coin tossing, wheels of fortune, roulette wheels, and similar generators of randomness all work in essentially the same way. And in each case the basic mechanism that leads to the randomness we see is a sensitive dependence on randomness that is present in the typical initial conditions that are provided.

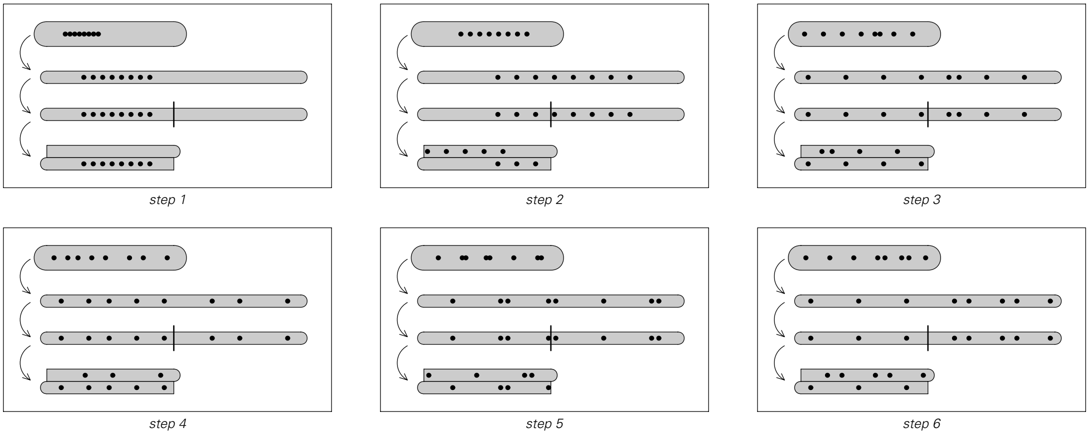
Without randomness in the initial conditions, however, there is no randomness in the output from these systems. And indeed it is quite feasible to build precise machines for tossing coins, rolling balls and so on that always produce a definite outcome with no randomness at all.

But the discovery which launched what has become known as chaos theory is that at least in principle there can be systems whose sensitivity to their initial conditions is so great that no machine with fixed tolerances can ever be expected to yield repeatable results.

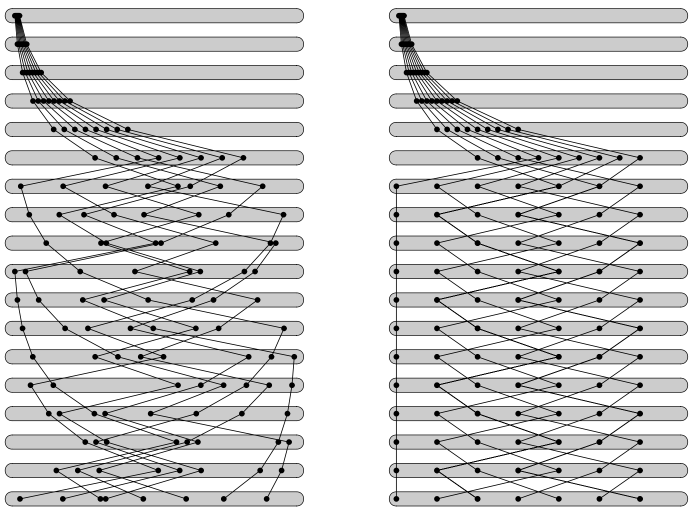
A classic example is an idealized version of the kneading process which is used for instance to make noodles or taffy. The basic idea is to take a lump of dough-like material, and repeatedly to stretch this material to twice its original length, cut it in two, then stack the pieces on top of each other. The picture at the top of the facing page shows a few steps in this process. And the important point to notice is that every time the material is stretched, the distance between neighboring points is doubled.

The result of this is that any change in the initial position of a point will be amplified by a factor of two at each step. And while a particular machine may be able to control the initial position of a point to a certain accuracy, such repeated amplification will eventually lead to sensitivity to still smaller changes.

But what does this actually mean for the motion of a point in the material? The bottom pictures on the facing page show what happens to two sets of points that start very close together. The most obvious effect is that these points diverge rapidly on successive steps. But after a while, they reach the edge of the material and cannot diverge any



A kneading process similar to ones used to make noodles or taffy, which exhibits very sensitive dependence on initial conditions. In the first part of each step, the material is stretched to twice its original length. Then it is cut in two, and the two halves are stacked on top of each other. The picture demonstrates that dots which are initially close together rapidly separate. (A more realistic kneading process would fold material rather than cutting it, but the same sensitive dependence on initial conditions would occur.)



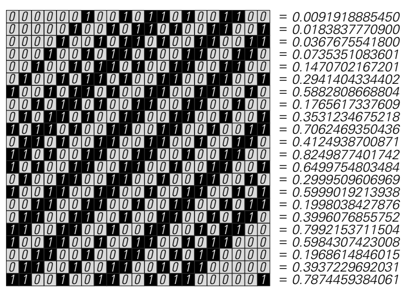
Two examples of what can happen when the kneading process above is applied to nearby collections of points. In both cases the points initially diverge exponentially, as implied by chaos theory. But after a while they reach the edge of the material, and although in the first case they then show quite random behavior, in the second case they instead just show simple repetitive behavior. What differs between the two cases is the detailed digit sequences of the positions of the points: in the first case these digit sequences are quite random, while in the second case they have a simple repetitive form.

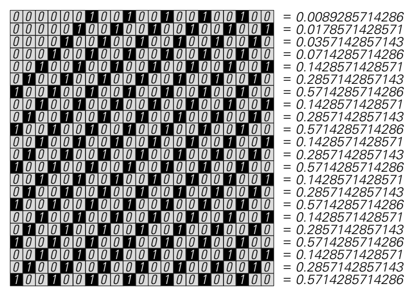
further. And then in the first case, the subsequent motion looks quite random. But in the second case it is fairly regular. So why is this?

A little analysis shows what is going on. The basic idea is to represent the position of each point at each step as a number, say x, which runs from 0 to 1. When the material is stretched, the number is doubled. And when the material is cut and stacked, the effect on the number is then to extract its fractional part.

But it turns out that this process is exactly the same as the one we discussed on page 153 in the chapter on systems based on numbers.

And what we found there was that it is crucial to think not in terms of the sizes of the numbers x, but rather in terms of their digit sequences represented in base 2. And in fact, in terms of such digit sequences, the kneading process consists simply in shifting all digits one place to the left at each step, as shown in the pictures below.





The digit sequences of positions of points on successive steps in the two examples of kneading processes at the bottom of the previous page. At each step these digit sequences are shifted one place to the left. So if the initial digit sequence is random, as in the first example, then the subsequent behavior will also be correspondingly random. But if the initial digit sequence is simple, as in the second example, then the behavior will be correspondingly simple. In general, a point at position x on a particular step will move to position *FractionalPart[2 x]* on the next step.

The way digit sequences work, digits further to the right in a number always make smaller contributions to its overall size. And as a result, one might think that digits which lie far to the right in the initial conditions would never be important. But what the pictures above show is that these digits will always be shifted to the left, so that eventually they will in fact be important. As time goes on, therefore, what is effectively happening is that the system is sampling digits further and further to the right in the initial conditions.

And in a sense this is not unlike what happens in the example of a car driving along a bumpy road discussed at the beginning of this section. Indeed in many ways the only real difference is that instead of being able to see a sequence of explicit bumps in the road, the initial conditions for the position of a point in the kneading process are encoded in a more abstract form as a sequence of digits.

But the crucial point is that the behavior we see will only ever be as random as the sequence of digits in the initial conditions. And in the first case on the facing page, it so happens that the sequence of digits for each of the initial points shown is indeed quite random, so the behavior we see is correspondingly random. But in the second case, the sequence of digits is regular, and so the behavior is correspondingly regular.

Sensitive dependence on initial conditions thus does not in and of itself imply that a system will behave in a random way. Indeed, all it does is to cause digits which make an arbitrarily small contribution to the size of numbers in the initial conditions eventually to have a significant effect. But in order for the behavior of the system to be random, it is necessary in addition that the sequence of digits be random. And indeed, the whole idea of the mechanism for randomness in this section is precisely that any randomness we see must come from randomness in the initial conditions for the system we are looking at.

It is then a separate question why there should be randomness in these initial conditions. And ultimately this question can only be answered by going outside of the system one is looking at, and studying whatever it was that set up its initial conditions.

Accounts of chaos theory in recent years have, however, often introduced confusion about this point. For what has happened is that from an implicit assumption made in the mathematics of chaos theory, the conclusion has been drawn that random digit sequences should be almost inevitable among the numbers that occur in practice.

The basis for this is the traditional mathematical idealization that the only relevant attribute of any number is its size. And as discussed on page 152, what this idealization suggests is that all numbers which are sufficiently close in size should somehow be equally common. And indeed if this were true, then it would imply that typical initial conditions would inevitably involve random digit sequences.

But there is no particular reason to believe that an idealization which happens to be convenient for mathematical analysis should apply in the natural world. And indeed to assume that it does is effectively just to ignore the fundamental question of where randomness in nature comes from.

But beyond even such matters of principle, there are serious practical problems with the idea of getting randomness from initial conditions, at least in the case of the kneading process discussed above.

The issue is that the description of the kneading process that we have used ignores certain obvious physical realities. Most important among these is that any material one works with will presumably be made of atoms. And as a result, the notion of being able to make arbitrarily small changes in the position of a point is unrealistic.

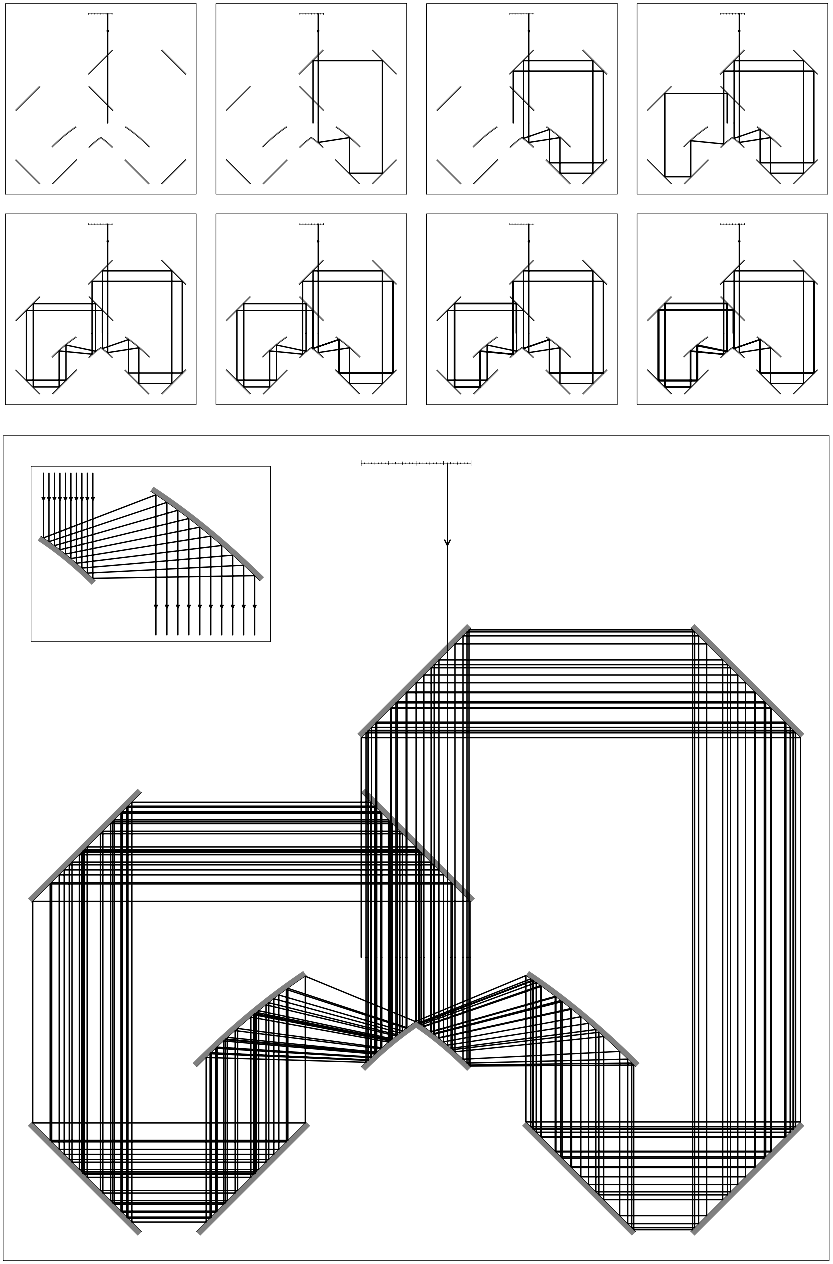
One might think that atoms would always be so small that their size would in practice be irrelevant. But the whole point is that the kneading process continually amplifies distances. And indeed after just thirty steps, the description of the kneading process given above would imply that two points initially only one atom apart would end up nearly a meter apart.

Yet long before this would ever happen in practice other effects not accounted for in our simple description of the kneading process would inevitably also become important. And often such effects will tend to introduce new randomness from the environment. So the idea that randomness comes purely from initial conditions can be realistic only for a fairly small number of steps; randomness which is seen after that must therefore typically be attributed to other mechanisms.

One might think that the kneading process we have been discussing is just a bad example, and that in other cases, randomness from initial conditions would be more significant.

The picture on the facing page shows a system in which a beam of light repeatedly bounces off a sequence of mirrors. The system is set up so that every time the light goes around, its position is modified in exactly the same way as the position of a point in the kneading process. And just as in the kneading process, there is very sensitive dependence on the details of the initial conditions, and the behavior that is seen reflects the digit sequence of these initial conditions.

But once again, in any practical implementation, the light would go around only a few tens of times before being affected by microscopic



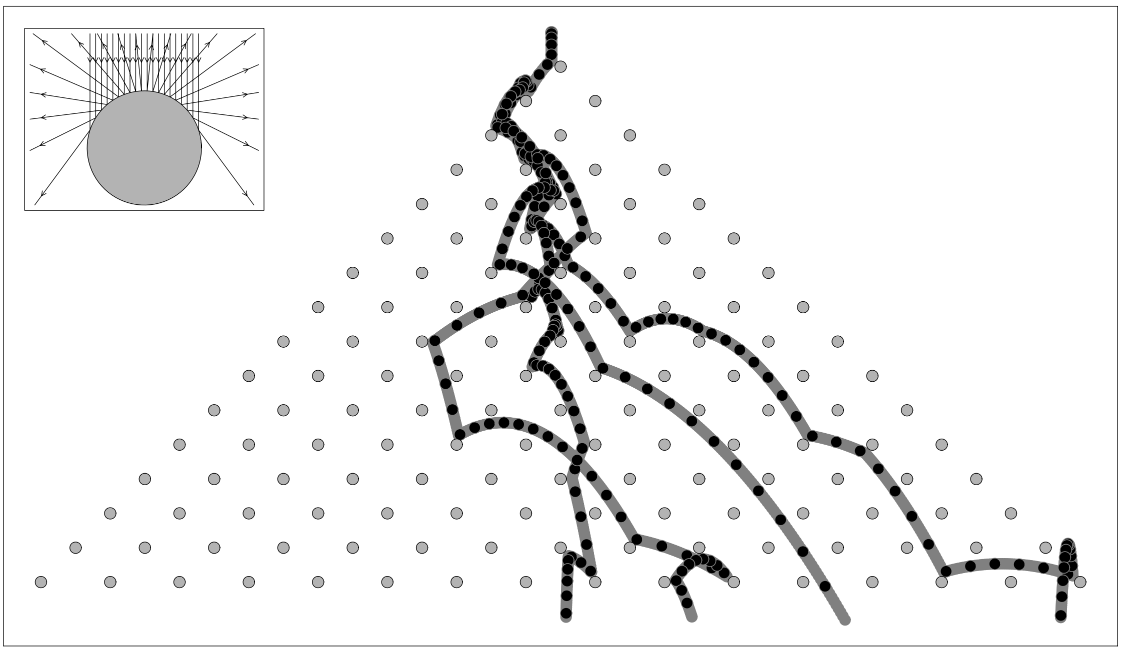
An arrangement of mirrors set up to exhibit randomness arising from sensitive dependence on initial conditions. The initial condition for the system is specified by the position of the incoming light ray in the gray region at the top of each picture. Whether the light ray goes to the left or to the right at each step is then determined by successive digits in the base 2 representation for the number that gives the initial condition. The heart of the system is the "amplifier" shown on the left which uses a pair of parabolic mirrors to double the displacement of each incoming ray. The initial condition used here is $\pi/4$, which has digit sequence 0.1100100100001111111.

perturbations in the mirrors and by other phenomena that are not accounted for in the simple description we have given.

At the heart of the system shown on the previous page is a slightly complicated arrangement of parabolic mirrors. But it turns out that almost any convex reflector will lead to the divergence of trajectories necessary to get sensitive dependence on initial conditions.

Indeed, the simple pegboard shown below exhibits the same phenomenon, with balls dropped at even infinitesimally different initial positions eventually following very different trajectories.

The details of these trajectories cannot be deduced quite as directly as before from the digit sequences of initial positions, but



Paths followed by four idealized balls dropped from initial positions differing by one part in a thousand into an array of identical circular pegs. The balls are taken to fall under gravity, and to bounce elastically whenever they hit a peg. As illustrated in the inset, small differences in direction are amplified—roughly doubling—at each bounce, with the result that after a few bounces the trajectories of the three balls are quite different. In a physical version of the system with balls of the same actual size as on this page perturbations from the environment will inevitably be amplified to have a significant effect on the trajectories after roughly the number of bounces shown. Versions of the system illustrated here—particularly with smaller peg spacings—are sometimes known as Galton or quincunx boards, and have been used since the late 1800s to demonstrate principles of probability theory. If balls are assumed to fall randomly on each side of each peg then with a large number of balls the final positions will approximate a binomial distribution.

exactly the same phenomenon of successively sampling less and less significant digits still occurs. And once again, at least for a while, any randomness in the motion of the ball can be attributed to randomness in this initial digit sequence.

But after at most ten or so collisions, many other effects, mostly associated with continual interaction with the environment, will always in practice become important, so that any subsequent randomness cannot solely be attributed to initial conditions.

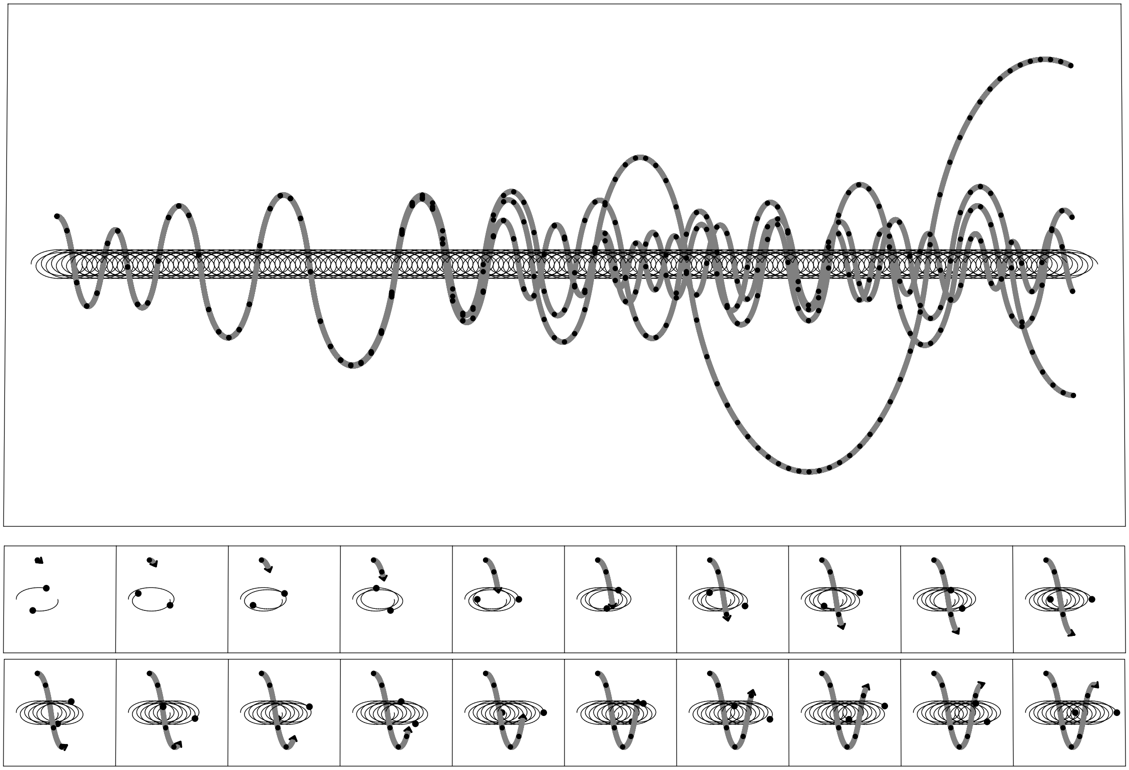
And indeed in any system, the amount of time over which the details of initial conditions can ever be considered the dominant source of randomness will inevitably be limited by the level of separation that exists between the large-scale features that one observes and small-scale features that one cannot readily control.

So in what kinds of systems do the largest such separations occur? The answer tends to be systems in astronomy. And as it turns out, the so-called three-body problem in astronomy was the very first place where sensitive dependence on initial conditions was extensively studied.

The three-body problem consists in determining the motion of three bodies—such as the Earth, Sun and Moon—that interact through gravitational attraction. With just two bodies, it has been known for nearly four hundred years that the orbits that occur are simple ellipses or hyperbolas. But with three bodies, the motion can be much more complicated, and—as was shown at the end of the 1800s—can be sensitively dependent on the initial conditions that are given.

The pictures on the next page show a particular case of the three-body problem, in which there are two large masses in a simple elliptical orbit, together with an infinitesimally small mass moving up and down through the plane of this orbit. And what the pictures demonstrate is that even if the initial position of this mass is changed by just one part in a hundred million, then within 50 revolutions of the large masses the trajectory of the small mass will end up being almost completely different.

So what happens in practice with planets and other bodies in our solar system? Observations suggest that at least on human timescales most of their motion is quite regular. And in fact this regularity was in



An example of the three-body problem, in which an idealized planet moves up and down through the plane of two equal-mass idealized stars in a perfect elliptical orbit. The trajectories obtained with four possible initial positions for the planet—differing by 10^{-8}—are shown. The pictures are made assuming the system to be in uniform motion from left to right. Successive black dots indicate where the planets are on each revolution of the stars. The main picture shows what happens over the course of 100 revolutions. The planet is assumed to be of negligible mass relative to the stars, and to start with zero vertical velocity at exactly an equal distance between the stars. The divergence of trajectories with slightly different initial vertical positions indicates sensitive dependence on initial conditions.

the past taken as one of the key pieces of evidence for the idea that simple laws of nature could exist.

But calculations imply that sensitive dependence on initial conditions should ultimately occur even in our solar system. Needless to say, we do not have the option of explicitly setting up different initial conditions. But if we could watch the solar system for a few million years, then there should be significant randomness that could be attributed to sensitive dependence on the digit sequences of initial conditions—and whose presence in the past may explain some observed present-day features of our solar system.