1989 vol 61

Chaology: the Emerging Science of Unpredictability

MICHAEL BERRY

Everybody knows that the world is unpredictable. We are continually being surprised. We cannot foretell whom we will meet, whom we will marry, and how we will meet our deaths. Only the fact of unpredictability is unsurprising.

Yet this familiar dominance of events by chance seems to contradict the picture of the world given by science. The contradiction reaches its extreme in the science where precise observation is most strongly connected with wide-ranging fundamental theory, namely physics and especially mechanics. In mechanics, the pendulum's swing is as "regular as clockwork", planets orbit the Sun with utter predictability, ancient eclipses can be reliably retrodicted and astronauts are guided to the Moon so as to land within a few kilometres of their target.

Until recently, most scientists would have explained life's unpredictability in terms of our ignorance: most events are influenced by many causes and we cannot know them all. The events that we initiate in laboratories, and call experiments, are exceptions, deliberately designed to be so simple that fundamental laws can be investigated without the inconvenience of multiple causes. That is the scientific method: isolate the relevant variables and study their effects one by one. (In astronomy we are lucky that simple situations often occur without our intervention.) Put more simply, the conventional view is that the world is unpredictable because it is complicated.

What I will now tell you is that this view is wrong. There are very simple systems moving according to Newtonian mechanics under the influence of completely known forces, and even simpler mathematical abstractions of these systems, whose behaviour is indistinguishable from a sequence of random events such as coin tosses. Nevertheless, these systems are causal in the sense that their future behaviour is completely determined by their present state.

Here is a simple example (Figure 1). A pendulum swings in two dimensions over a plane where some magnets are fixed. These repel another magnet inside the bob of the pendulum, which therefore moves under the influence of two forces: gravitational and magnetic. The outcome is an erratic weaving around between the magnets. We know that this randomness is intrinsic to Newton's equations and not

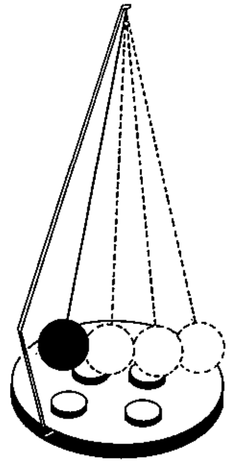


Figure 1
The magnetic conical pendulum.

the result of some uncontrollable external influence, because it persists in computer simulations (Figure 2) as the numerical accuracy is increased.

The unpredictable motion is chaos. Its study is chaology, which in recent decades has rescued mechanics from being a backwater about which we thought we knew everything (apart from details), and made it again an intensely-developing area of theoretical physics.

Chaology has emerged from a synergism between several different disciplines, each providing important contributions:

- *Pure mathematics*, through crucial theorems about the stability of solutions of differential equations such as those expressing natural laws;

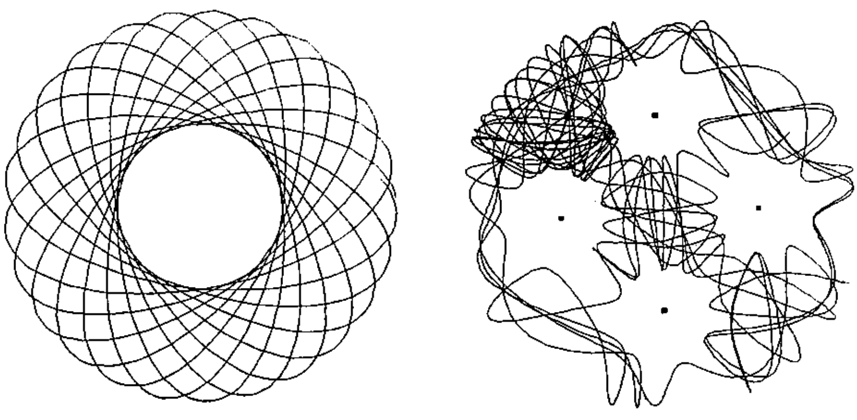


Figure 2 *Orbits of a conical pendulum:* **(a)** *without magnets;* **(b)** *with magnets (positions marked by square dots).*

- *Computers, and especially computer graphics*, giving a numerical "laboratory" enabling us to refresh mathematical parts that other techniques cannot reach;
- *Astronomy*, through erratic motion of stars in the galaxy, and in gaps in the asteroid belt between Jupiter and Mars;
- *Chemical physics*, through the motion of atoms in small vibrating molecules;
- *Plasma physics*, through the instabilities that prevent charged particles being kept bottled up by magnetic fields whilst they react to produce fusion power;
- *Accelerator physics*, through the need to ensure the stability of beams of particles, which have to stay together for trillions of circuits of their confining pipes;
- *Fluid dynamics*, through turbulence, which makes liquids flow more irregularly as the viscosity is reduced (think of a river in flood, rushing under a bridge) and which is arguably the deepest unsolved problem in physics that does not involve cosmology or the discovery of new fundamental laws;
- *Meteorology*, through the unpredictability of weather.

It is worth remarking that more than a century ago James Clerk Maxwell speculated that the reason why we cannot predict the weather is not that we are ignorant of the laws governing the dynamics of air masses but because these laws have inherent instabilities: two initially imperceptibly different weather systems evolving according to identical laws will soon come to behave very differently. He put it in this way. The fundamental causal principle of dynamics, that "identical causes produce identical effects" does not imply the more practical principle that "similar causes produce similar effects". Nowadays the idea that similar but not identical causes can produce very different effects has become central to chaology.

The question before us is therefore: how can something be unpredictable if it is the inevitable consequence of known laws operating on known antecedents? It is a mathematical question, because the laws are embodied in equations. Therefore the essence of chaos is to be found in mathematical ideas, and it is these, rather than the details of applications to particular fields, that I intend to tell you about. It would not be appropriate to give a formal mathematical discussion here, and in any case I am not a mathematician and would not be able to do that. So I have tried to bear in mind the proverb: "If you understand something, it is simple; if it is not simple, you don't understand it". But the ideas are quite subtle, and I am uncomfortably aware of Einstein's counter-proverb: "Physics (in this case mathematics too) should be made as simple as possible, but not simpler".

The Bouncer

Figure 3 shows a machine whose behaviour we will try to understand. I call it the bouncer; it is sold as a so-called executive toy under the name of "space ball". R is a rotator consisting of two light hollow balls containing magnets. Its axis A_1 swings to and fro at the top of a pendulum, pivoted at A_2, whose bob is the heavy driving ball D. The (predictable) swinging of D is itself driven by an electromagnet in the

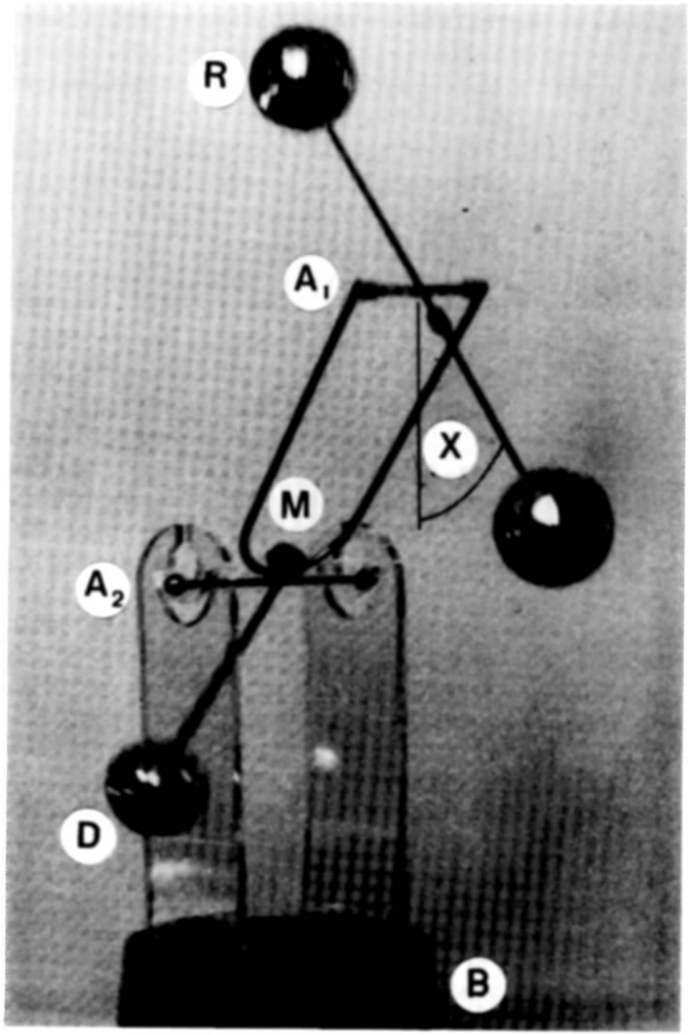


Figure 3
The bouncer.

base B that is switched on by a circuit that senses each approach of a magnet inside D. As well as rotating inertially and being swung by D, R occasionally bounces because its balls are repelled by a magnet M, situated just above A_2 on the pendulum. As a result of these influences the rotator turns in an apparently random manner, clockwise and anticlockwise, on time scales from several seconds (the driving period) to several weeks (the life of the battery in the base). We are unable to answer certain questions about the motion, such as: if the rotator is turning clockwise now, in which direction will it be turning immediately after the first bounce after five minutes from now?

To understand, we must first describe. The motion of the rotator is the history of the angle it makes with the vertical. Let us call this angle $x(t)$. Knowledge of x at any one time is not sufficient to determine the subsequent evolution. In order to apply Newton's laws we also need to know how fast the rotator is turning, that is we need the angular velocity, which we call $y(t)$. If we think of x and y as coordinates on a plane, then the motion is a curve in the plane, determined by one point on it - the initial conditions. We need only take x as running from 0° to 180°, because under a half turn the rotator looks the same. This natural arena for the motion of the bouncer is called its *phase space.*

Any reasonable model of the dynamics, fed into a computer and plotted in the phase space, generates a curve that is just a mess of lines. Now this need not be an indication of chaos. The mess is the result of showing the whole history of the motion and thereby presenting ourselves with too much information. We can understand more by seeing less. Imagine that instead of looking at the motion continuously we measure x and y only at regular intervals, for example once every period of the driving pendulum. If we plot just these points, we get what is called a *stroboscopic phase portrait* of the motion. Each orbit is thereby reduced to a pattern of dots, which is a coded representation of the motion. These patterns will reveal the chaos.

Maps

The dot pattern is the result of repeated application of what mathematicians call a map. A map is a rule for moving points about. We are familiar with geographic maps: these are made by a particularly simple rule: each point on the surface of part of the Earth (*e.g.* London) is moved to another point on a piece of paper, in such a way that shapes (*e.g.* of Hyde Park) are preserved but areas enormously diminished. Underlying the stroboscopic phase portrait of the bouncer is a map constructed from dynamics: each point x,y in the phase space is moved to a point further along its orbit (one pendulum period later). This map does not preserve shapes, but if we neglect friction in our model it does preserve areas.

A map of London is the result of applying the rule once to each of the points on the ground. We do not usually think of repeating the rule to get a map of the map, a map of the map of the map, *etc.*. If we did that with an Ordnance Survey map of London the first repetition would be only one-fiftieth of a millimetre across - although it is amusing to realise that in principle such repeated maps are contained in every map that lies in the territory it represents. In dynamics, though, we do repeatedly apply the map to individual points, because the pattern of dots gives immediate information about the orbit over long times.

To map the bouncer we need a dynamical model. Figure 4 shows the map generated by the simplest of a range of models. In this, which is not meant to be realistic, friction is neglected and the magnets are assumed to be infinitely strong,

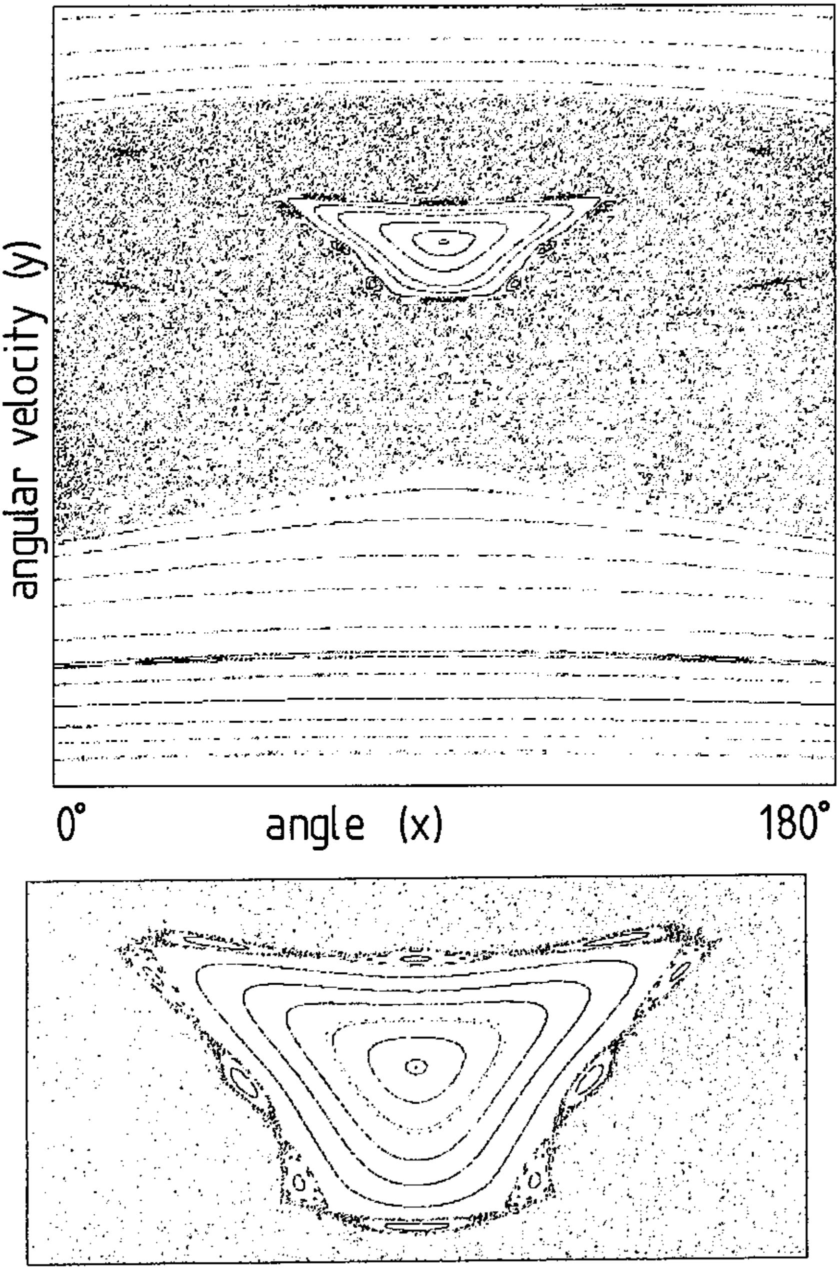


Figure 4(a) *Stroboscopic phase portrait of a model for the bouncer;* **(b)** *Magnification.*

so that bouncing occurs every time the rotator encounters them. In this model, the rotator moves like the clapper of a swinging bell. Several orbits are shown. I will discuss them separately.

The point at the centre of the eye represents a particularly simple orbit that repeats after each pendulum period, so that x,y just maps onto itself - it is a fixed point of the map. It can be quite tricky to reconstruct the motion between the mapped points; in this case there are two bounces from inside. Surrounding the fixed point are loops. Each of these represents an orbit which does not repeat but lies close to the repeating one. This is another way of saying that the repeating orbit is stable. Surrounding the eye are two chains of little islands. Each of these corresponds to stable orbits that repeat after five pendulum periods. None of these orbits are chaotic. Nor are those represented by the curves at the top and bottom of the map, where the angular velocity (y) of the rotator is so large that the driver hardly moves between bounces.

Now consider the great dust of points surrounding the eye. All these points lie on one orbit, and are generated erratically by the map. The orbit corresponds to a rotator motion that not only never repeats but is unstable in the sense that tiny differences in the starting conditions get rapidly magnified by bouncing and lead to very different orbits. This is the chaos whose origin we will soon try to understand.

First, though, I want to convince you through illustration that the complicated combination of random and stable orbits in Figure 4 is not a peculiarity of the bouncer or of our dynamical abstraction of its motion. On the contrary, it is typical of maps that preserve area. Some idea of the amazing complexity that even the simplest maps can generate is shown in Figure 5(a). Allow me to write the equation of this map:

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} Ax - B(y - x^2) \\ Bx + A(y - x^2) \end{pmatrix} \qquad (1)$$

where $A = 0.24005862$, $B = 0.97075839$

It is so simple that many thousands of repetitions can be computed, making it easy to display magnifications (Figure 5(b)). This reveals an endless hierarchy, with order and chaos intermingled down to the finest scales.

Into Chaos

Some maps are totally chaotic in the sense that all their orbits are unstable. One such map is shown in Figure 6. Its phase space is the unit square, on which points transform according to a rule so simple that it can be described in words: the new x is got by adding the old x and the old y, and the new y is got by adding the old x

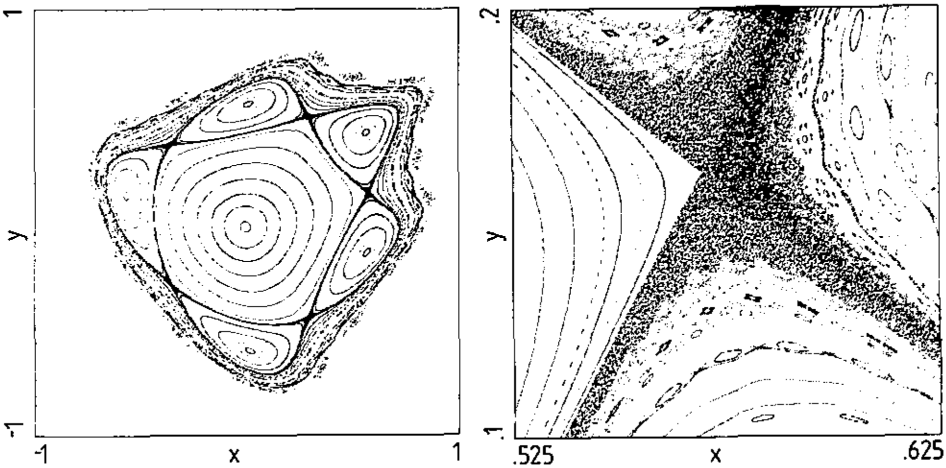


Figure 5 (a) *Orbits generated by repetition of the map defined by equation (1);* **(b)** *Magnification of (a) between the rightmost eyes (after M. Hénon).*

and twice the old y, subtracting one if necessary to ensure that the new x and y both lie between zero and one. In symbols,

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} x+y \\ x+2y \end{pmatrix} \pmod 1 \qquad (2)$$

The dust in Figure 6 is indistinguishable from what we would have got by scattering the points at random, rather than by the completely deterministic rule (2).

To lay bare the heart of the chaos we make one more abstraction. The need to consider at least two variables (x and y) comes from physics: to determine motion, it is necessary to specify a velocity as well as a coordinate. But it is possible to imagine a "lineland" whose history is described by just one quantity x, inhabiting the line segment between zero and one. "Dynamics" in lineland is then a rule transforming the old x into a new x, that is, a map on the unit interval. Even in this physically impoverished world some maps give predictable orbits and some give random orbits. A predictable map is the squaring transformation:

$$x \rightarrow x^2 \qquad (3)$$

under which any x gets rapidly smaller and so is inexorably attracted to zero. To get chaos we double x instead of squaring it, again subtracting one where necessary in order to stay in the unit interval:

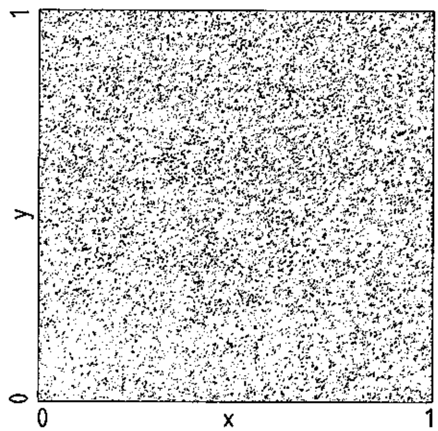


Figure 6
Chaotic orbit generated by 100,000 repetitions of the map defined by equation (2).

$$x \to 2x \pmod 1 \qquad (4)$$

A graph of the successive values of x (Figure 7) shows the chaos. One way it manifests itself is in the sequence describing whether successive x lie in the first half of the interval (0 to 1/2) or the second (1/2 to 1). The sequence is apparently random. Soon we shall see that it is really random.

For this is chaos we can understand. The successive values of x are determined by the starting value. We could write this as a decimal, that is as a point followed by a string of digits in the range zero through nine. It is, however, advantageous to write the starting x in binary notation, in which only zeros and ones follow the point, for example

$$x = 0.0110110001011.. \qquad (5)$$

If the first digit after the point is a zero, x is in the range 0 to 1/2; if it is a one, x is in the range 1/2 to 1. In binary, the doubling rule can be expressed very simply: to get the new x, move the point one place to the right. Thus

$$0.0110110001011.. \to 0.110110001011.. \to$$
$$\to 0.10110001011.. \to 0.0110001011..\ etc. \qquad (6)$$

Every x in the range 0 to 1 is represented in binary by a string of zeros and ones, and *vice versa.* Moving the point - that is operating the doubling "dynamics" - exposes these digits one by one. Therefore all possible jump sequences (between the two halves of the interval) can be achieved; to find the starting x that generates a given sequence, simply code the sequence by zeros and ones and put a point in front. The totality of possible sequences is exactly the same as the totality of

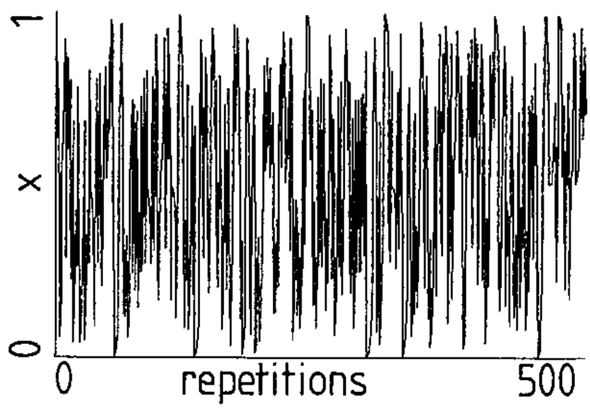


Figure 7 *Graph generated by 500 repetitions of the doubling map defined by equation (4).*

sequences of coin tosses, with zero representing heads and one representing tails. Which of these sequences are random? To answer this we need to know what randomness means. It is easy to write sequences that are not random, for example

$$00000000\ldots \text{ and } 10101010.. \qquad (7)$$

These are predictable because they can be generated by simple rules - embodied for example in computer programs or algorithms - specifying the nth digit for arbitrarily large n.

This suggests that we should define a sequence as random if it cannot be generated by a rule shorter than itself. The only way to reproduce such a sequence is to write it out and say: "Copy this". In other words, a random sequence is informationally incompressible. Each new digit comes as a surprise, so the definition accords with what we intuitively think randomness is.

Randomness is related to instability. Consider two initial x which are close together in the sense that their first n digits are the same. For the first few doublings they will remain close together; in particular they will explore the halves of the interval in the same sequence. After n doublings, however, their trajectories can be completely different. To predict the outcome beyond n doublings, we would need more initial digits. Another way of putting this is that by observing the system we continually get new information about the initial x. This contrasts with orbits that are not chaotic; these are insensitive to the initial condition, so that repeated observation does not generate new information about it.

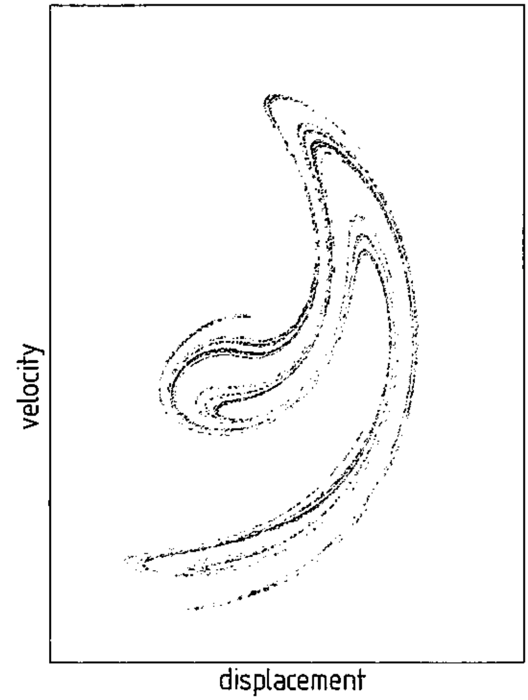


Figure 8
Strange attractor in the phase portrait of a pendulum driven periodically and damped by friction (after K. Ueda).

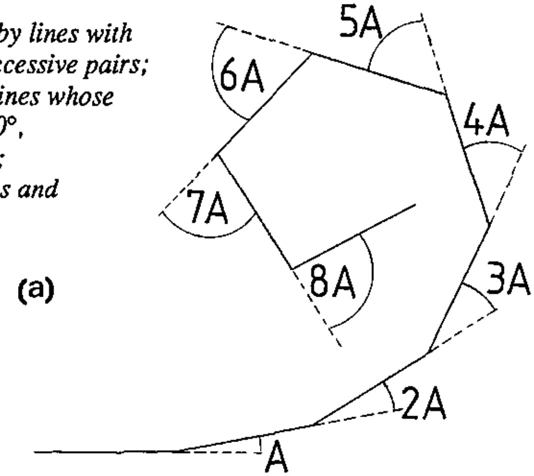
With a precise definition of randomness, purged of subjective elements, one can proceed to mathematics. A most important theorem has been proved. It states that *almost all sequences are random.* In terms of our doubling dynamics this means that unless we choose a very special x to start with (*e.g.* a rational number) the orbit will be totally unpredictable, even though it is fully determined by a dynamical law whose expression (equation (4)) contains no element of randomness. Here at last is the essence of chaos. Its ramifications are being explored in a new branch of mathematics called "algorithmic complexity theory".

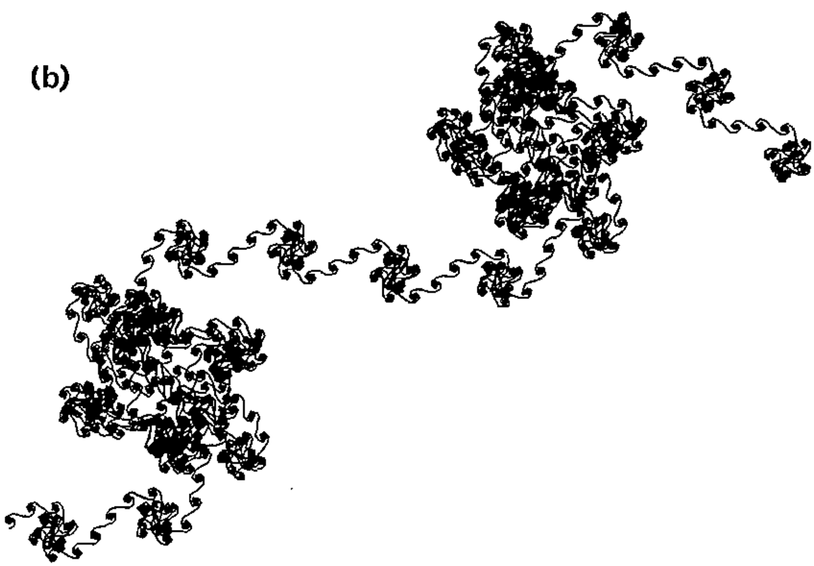
But has this understanding been won at the cost of abstracting all reality out of the original problem? Picasso's abstractions drew a similar criticism, to which Wallace Stevens replied in a poem that will serve equally well for us physicists:

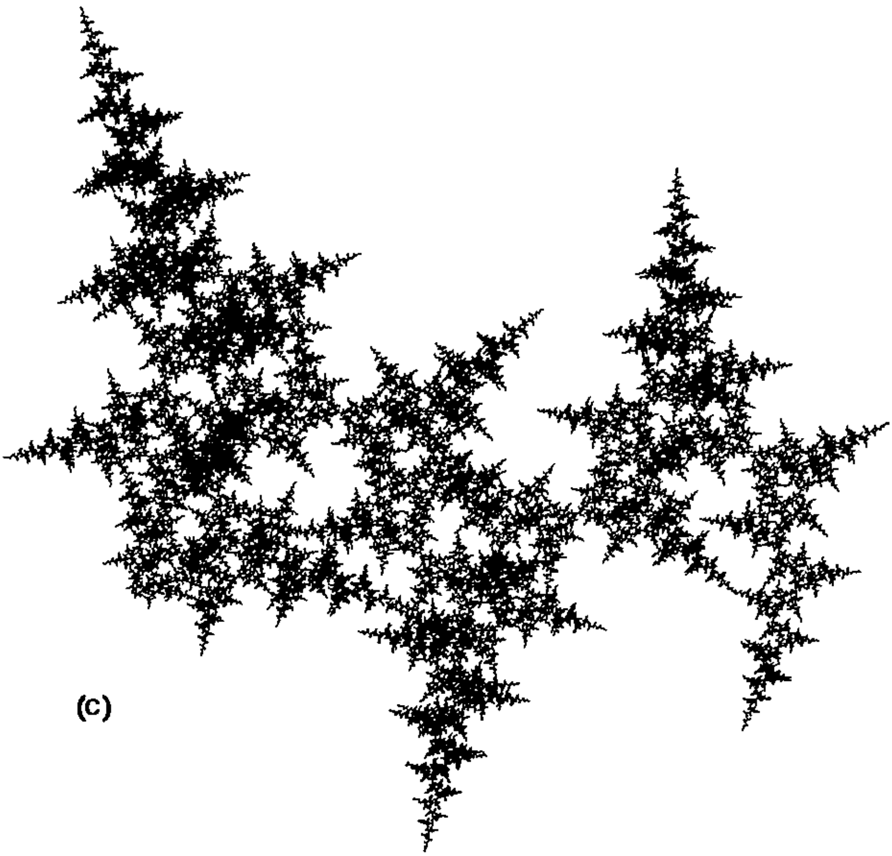
They said, "You have a blue guitar,
You do not play things as they are."
The man replied, "Things as they are
Are changed upon the blue guitar."

In fact, the abstraction is entirely appropriate, for it has been found that for any mechanical system the orbits can be coded by a sequence of digits constructed

Figure 9
(a) *Generation of a curlicue by lines with increasing angles between successive pairs;*
(b) *curlicue made by 12,000 lines whose angles are multiples of r x 360°, where r = (√82 - 9) (~ 0.055);*
(c) *as (b) but with 50,000 lines and r = (√5 - 1)/2 (~ 0.0618).*







(c)

according to certain rules. If the rules resemble the doubling transformation, the orbit is chaotic.

Strange Attractors

Back to the bouncer. There were two simplifications underlying the computed phase portrait in Figure 4. One was that the magnets were so strong as to make the rotator bounce at every encounter. This is easily put right by stipulating that a bounce occurs only for encounters with sufficiently small relative velocity, and the resulting phase portrait is qualitatively similar to Figure 4. The second simplification was the neglect of friction, and is much more serious.

In the abstract world of maps, friction is incorporated by making each transformation shrink areas. Consider the ordinary pendulum swinging in air.

Whatever its initial position and velocity x and y, it always comes to rest at the origin $x = y = 0$ because of damping from the air. Every point eventually maps to the origin, so that the state of rest is an *attractor* of the motion. This is quite different from what we have seen in phase maps so far (Figures 4, 5 and 6) where there are no attractors and orbits (chaotic or not) never rest.

But our rotator never comes to rest either, in spite of being continually damped by friction, because it is being *driven* by the magnet through the swinging of the heavy pendulum. Therefore its stroboscopic phase portrait must be generated by an area-shrinking map whose attractor is more complicated than a single point representing rest. Such an attractor is shown in Figure 8. It was calculated not for the bouncer (which gives a similar picture) but for a pendulum under the additional conflicting influences of steady damping and periodic driving (and differing from the usual pendulum in that the restoring force is proportional not to the displacement but to its cube).

The overall impression is of a ghostly swirl, reflecting the stretching and folding common to all maps of this type. Closer scrutiny resolves the main swirl into several subswirls, each of which consists of several subsubswirls. Computation at higher resolution reveals that this hierarchy continues, apparently *ad infinitum*. The attractor is a fractal, because it possesses structure on all scales, with a dimension that is not a whole number. Here the dimension is less than two (because any attractor must have zero area) and greater than one (the attractor contains infinitely many leaves and so has infinite length). This type of attractor is often called "strange".

A strange attractor is an exquisite combination of order and chaos. The attractor itself is a highly structured and systematically describable geometric object. But on the attractor the motion of the point representing the state of the dynamical system is random. Strange attractors were discovered during computer explorations of a model for the unpredictability of the weather.

There are a few more subtleties about the bouncer, such as the fact that its balls are inevitably unbalanced, which makes the rotator tend to orientations where the magnet will more easily keep it bouncing. Nevertheless, what I have told you captures the essence of its motion: chaos originates in the instability caused by the magnetic repulsion, and persists in spite of friction because of the continued driving.

Curlicues

Chaos need not show itself as instability of motion. It can govern aspects of the structure of apparently well-ordered processes. I will illustrate this with a family of geometric patterns arising in the theory of wave interference.

Draw short straight lines end to end, with the angle between successive pairs increasing by a fraction r of a circle, that is by $r \times 360°$ (Figure 9(a)). Evidently the

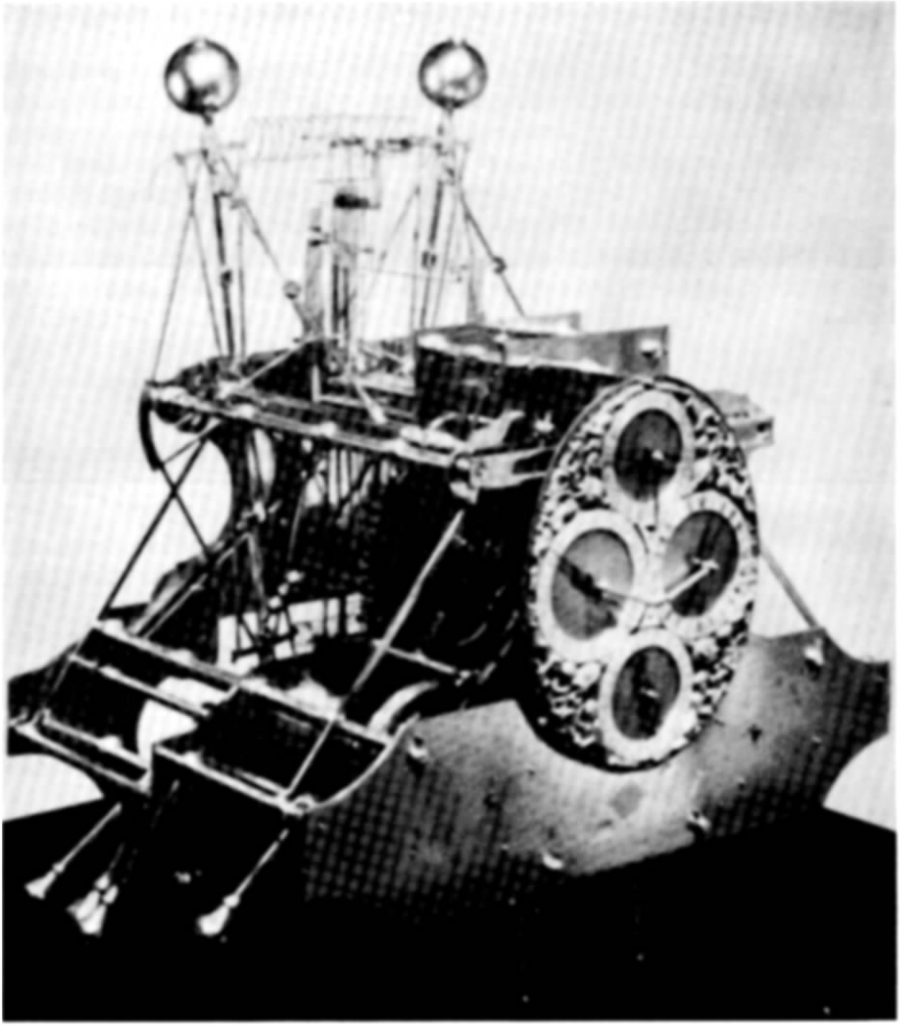


Figure 10 *Harrison's first sea clock (1735).*

lines wind up into a discrete spiral - a *curlicue*. Intricate patterns emerge on scales so large that there are thousands of lines individually too small to be resolved. The smallest curlicues assemble into hyperspirals which make hyperhyperspirals, and so on in a hierarchy. The curlicues are clearest when r is small (Figure 9(b)); larger values of r (Figure 9(c)) generate spiky patterns. The patterns give us a mathematical microscope revealing the *arithmetic* of the number r. If r is rational, the hierarchy of scales stops. If not, the hierarchy is endless and its details depend on what type of irrational number r is.

Chaos arises in the description of the hierarchy. The pattern on any given scale for a given value of r is obtained by magnification (by $1/\sqrt{r}$) of the pattern on the next smaller scale for a different value of r. Successive r, that is successive magnifications, are related by a map of the unit interval representing fractions. This map is chaotic, like the doubling transformation (equation (4)), although different in its details. Here is chaos entering subtly, not as irregularity in the curlicues as we see them but as randomness in the magnifications which enable us to comprehend the patterns. It is as though the chaos were in the mind, rather than the eye, of the beholder.

Law without Predictability

It is remarkably easy to make chaotic machines or mathematical mappings. Contrast this with the elaborate mechanisms (Figure 10) required to guarantee predictable motion. One may wonder why the techniques for analysing chaotic motion were not developed long ago and presented routinely in courses and textbooks on mechanics. Perhaps the reason is that engineers do not usually earn their living designing and producing unpredictable machines.

Chaology has deep implications. For several centuries, scientists (especially physicists) thought that the test of our understanding of a phenomenon is the ability to make predictions. Randomness was equated with ignorance. But now we discover an abundance of systems whose behaviour, although governed by precisely-known laws, cannot be predicted even in principle because they are so unstable. To know the law is not necessarily to know the behaviour (politicians please note). This discovery resolves what has sometimes seemed to be a contradiction, especially to nonscientists: the contrast between the richness and complexity of the world as we see it and the stark simplicity of the laws governing it. Laws encode behaviour in a much more compact way than we realised.

Bibliography

Michael Berry. 1986. In: Diner S., Fargue G. and Lochak G. (eds.), *The unpredictable bouncing rotator: a chaology tutorial toy*, in *Dynamical systems: a renewal of mechanism*, pp. 3-12. *World Scientific*.

Ivar Ekeland. 1988. *Mathematics and the Unexpected*. University of Chicago Press.

Joseph Ford. 1983. How random is a coin toss. *Physics Today*, April, 1-8.

James Gleick. 1987. *Chaos: Making a New Science*, Viking.

Ilya Prigogine and Isabelle Stengers. 1984. *Order out of chaos*, Heinemann.

Heinz Georg Schuster. 1988. *Deterministic chaos*. VCH Verlagsgesellschaft.