Journal of Economic Perspectives—Volume 3, Number 1—Winter 1989—Pages 77–105

Chaos: Significance, Mechanism, and Economic Applications

William J. Baumol and Jess Benhabib

Imagine a bargaining model (say, involving diplomats negotiating tariff levels or a disarmament treaty) in which each party has been instructed by higher headquarters to respond to each new offer by her opposite number with a counteroffer that is to be calculated from a simple reaction function provided in advance. Both negotiating parties are prohibited from revealing their own reaction functions to the other. If the perfectly deterministic sequence of offers and counter-offers that *must* emerge from these simple rules were to begin to oscillate wildly and apparently at random, the negotiations could easily break down as each party, not understanding the source of the problem, came to suspect the other side of duplicity and sabotage. Yet all that may be involved, as we will see, is the phenomenon referred to as *chaos*, a case that is emphatically not pathological, but in which a dynamic mechanism that is very simple and deterministic yields a time path so complicated that it will pass most standard tests of randomness.

Chaos has become a subject of great interest to specialists and nonspecialists alike. Besides economics, it has entered the literature of geometry, physics, ecology and meteorology. It has been written about at length in the *New York Times Magazine* and *Scientific American*, as well as technical publications. This article seeks to describe what it is, how it works, and what it means for economics.

Roots in the Earlier Dynamic Models

The roots of economists' interest in complex dynamics are to be found in the vast nonmathematical literature on business cycles, with its large number of models each undertaking to provide a set of conditions sufficient to generate oscillatory behavior in

■ *William J. Baumol is Professor of Economics, Princeton University, Princeton, New Jersey, and New York University, New York, New York. Jess Benhabib is Professor of Economics, New York University, New York, New York.*

the economy. However, many of these models were vague and their logic difficult to verify.

All this changed in the 1930s with the work of Frisch (e.g., 1933), Lundberg (1937), and Samuelson (1939), who used difference equations, differential equations and mixed models to generate deterministic time paths. It was readily demonstrated for such models that any parameter values chosen within broad ranges that are sometimes easily calculated must yield an oscillatory time path.

For example, the simplest sort of difference equation is the first order (one-period lag) linear equation $y_{t+1} = ay_t$, where y_t is the value at time t of its single variable. Given initial value y_0, this obviously generates the time path

$$y_1 = ay_0, \qquad y_2 = ay_1 = a^2 y_0, \ldots, \qquad y_t = a^t y_0.$$

Clearly, *any* negative value of parameter a yields an oscillatory time path, as $a^t y_0$ successively goes from positive to negative and back.

An example that is somewhat less oversimplified is Samuelson's justly famous multiplier-accelerator model (1939) which is made up of the three standard relationships

$$Y_t = C_t + I_t, \qquad C_t = cY_{t-1} + k, \qquad I_t = b(Y_{t-1} - Y_{t-2}),$$

where Y is national income (output), C is consumption, c is the marginal propensity to consume, and I is investment. The C equation is an obvious linear consumption function with a one-period lag. The investment function is a linear lagged accelerator with investment assumed proportionate to the preceding period's rate of growth of output. Substitution of the two latter equations into the first at once yields

$$Y_t = (c + b)Y_{t-1} - bY_{t-2} + k,$$

which is Samuelson's second-order linear difference equation. It is easy to show that for broad ranges of values of c and b this equation, too, generates oscillations, and it is not hard to explain intuitively the economics of the oscillatory mechanism.

These models were received enthusiastically and generated many writings by leading economists. Still, before long disappointment seemed to set in and publication slowed. There were two basic reasons. First, it became clear that the behavior of the time path generated by such a linear dynamic system can be extremely sensitive to changes in the values of the parameters, as well as the structure of the model. That made it hard to formulate models (and econometric estimates of their parameter values) that constituted robust and reliable representations of reality. Still, as we will see, this is not really a shortcoming of the models, but a weakness of some of their interpretations. Such sensitivity, though of a rather different variety, holds with even greater force for chaos models, and is one of the main reasons for their interest. Second, it was recognized early that, qualitatively speaking, such linear models were capable of generating only four types of time path: (1) oscillatory and stable (that is, converging with oscillations of decreasing amplitude toward some fixed equilibrium value); (2) oscillatory and explosive (cycles of ever-increasing amplitude); (3) nonoscillatory and stable; and (4) nonoscillatory and explosive. It was soon recognized that linear equations even more complex (that is, of higher order) than Samuelson's would not generate any time paths basically different from these four. This range of possible time path configurations simply was not sufficiently rich for the economists' purposes, since in reality time paths are often more complicated and many oscillations do not seem either to explode or dampen toward disappearance.

A solution to the problem, brought to our attention by Hicks and Goodwin, was the nonlinear models, perhaps of the general form $Y_t = f(Y_{t-1}, \ldots, Y_{t-h})$. Responding to real economic issues, not just to formal mathematical problems, these authors, for example, showed that such a nonlinear model can yield a stable limit cycle toward which all possible time paths of the variable Y_t converge. That is, rather than converging to a fixed equilibrium value, Y^*, with the oscillations dampening out toward zero amplitude, the nonlinear models could instead yield a stable equilibrium cycle, with Y_t forever wandering from peak to trough along the equilibrium cyclical path. This is pretty much where matters were left, with the work stopping short of introducing explicitly a degree of nonlinearity sufficiently great to generate chaotic behavior.

In essence, chaos theory shows that a simple relationship that is deterministic but *nonlinear*, such as a first order nonlinear difference equation, can yield an extremely complex time path. Intertemporal behavior can acquire an appearance of disturbance by random shocks and can undergo violent, abrupt qualitative changes, either with the passage of time or with small changes in the values of the parameters. Chaotic time paths can have the following attributes, among others: a) a trajectory (time path) can sometimes display sharp qualitative changes in behavior like those we associate with large *random* disturbances (for example, very sudden changes from small-amplitude to large-amplitude cycles, and vice-versa), so at least some of the standard tests of randomness cannot distinguish such chaotic patterns of change from "truly random" behavior; b) a time path is sometimes extremely sensitive to microscopic changes in the values of the parameters—a change in, say, the fifth decimal place of one parameter can completely transform the qualitative character of the path; c) they may never return to *any* point they had previously traversed, but display in a bounded region an oscillatory pattern which is consequently very "disorderly."

Where chaos occurs economic forecasting becomes extremely difficult (for further discussion see Baumol and Quandt, 1985). The two basic forecasting devices—extrapolation (of various degrees of sophistication) and estimation of a structural forecasting model—both become questionable. Extrapolation is hardly appropriate for a time path that might exhibit two-period oscillations of steadily increasing amplitude for 50 periods, with the fluctuations but all disappearing for the next 20 periods, and still another pattern abruptly emerging thereafter. Forecasting carried out with the aid of estimates of the parameters of an underlying model also runs into difficulties if an error in calculation of the third decimal place of a parameter can change the qualitative character of the forecast beyond recognition.

The work on chaotic dynamics suggests that disenchantment with earlier dynamic models is perhaps attributable to failure to recognize their most promising role —that of revealing sources of uncertainty, and enriching the list of recognized *possible* developments. Three brief examples will suffice to illustrate the point and to lead us toward economic applications of chaos analysis.

The first is the demonstration by even the earliest formal models of how easy it is for any deterministic time path to produce oscillation, a fact well recognized by engineers who work with control theory (the theory of automatic adjustment mechanisms such as steering devices or thermostats). The analysis demonstrated that the construction of a model sufficient to imply the presence of fluctuations requires neither convoluted reasoning nor premises that are implausible or pathological.

In addition, despite their sensitivity, the dynamic models proved to be effective instruments for *dis*proof of the universal validity of propositions that had previously been accepted too readily, and for corresponding warnings to policy designers. For instance, such a model was used to disprove by counterexample the allegation that profitable speculation is always and necessarily stabilizing (Baumol, 1957). That is, even if speculators buy when price is low and sell when price is high they can conceivably increase the amplitude of any fluctuations in the price of the good in which they are speculating, if its price happens to be rising at the time they buy and declining at the time they sell. Similarly, it was shown that slight lags in response can undermine apparently rational countercyclical policy (Baumol, 1961). A government which pursues an "obvious" policy such as spending more whenever the economy's output (or its growth rate) is below some target level and reducing its expenditure when output (or growth rate) is above target can increase the amplitude of fluctuations even if the lag in its countercyclical measures is assumed unrealistically small (see also Phillips, 1954, 1957).

Chaos theory has at least equal power in providing *caveats* for both the economic analyst and the policy designer. For example, it warns us that apparently random behavior may not be random at all. It demonstrates dramatically the dangers of extrapolation and the difficulties that can beset economic forecasting generally. It provides the basis for the construction of simple models of the behavior of rational agents, showing how even these can yield extremely complex developments. It has served as the basis for models of learning behavior and has been shown to arise naturally in a number of standard equilibrium models. It offers additional insights about the economic sources of oscillations in a number of economic models. Some of this will be illustrated later, once the required tools have been described. Meanwhile, some references to the literature should suggest the range encompassed so far by economic writings on the subject.[1]

[1]In economics, the possibility of cyclical and chaotic dynamic behavior was perhaps first suggested by May and Beddington (1975), and has been shown to arise in simple ad hoc macroeconomic models (Stutzer, 1980; Day and Shafer, 1983, in duopoly models (Rand, 1978), in models of growth cycles (Day, 1983; R. A. Dana and P. Malgrade, 1984), in cobweb models of demand and supply (R. V. Jensen and R. Urban, 1982), in models of the firm subject to borrowing constraints (Day, 1982), in dynamic models of choice with endogenous tastes (Benhabib and Day, 1981), in models of productivity growth (Baumol and Wolf, 1983), in dynamic models of advertising expenditures (Baumol and Quandt, 1985), in models analyzing military arms races and disarmament and negotiations (Baumol, 1986). Cyclic and chaotic dynamics have been

How Complex Cyclical Patterns Arise

Since much of the discussion that follows relates to cyclical and oscillatory behavior, it is important to define precisely what we mean by those terms. A time path, y_t, will be taken to be characterized by a cycle whose duration is p periods if it always replicates itself *precisely* every p periods from any initial point in its trajectory, and does not always repeat itself precisely in any smaller number of periods. In short, any pattern which repeats itself exactly every p periods is said to be a *p-period cycle.*

In contrast, an *oscillatory* time path is defined more vaguely as one which is not monotonic, involves "frequent" rises and declines in the values of its variables, but in which the time path may rarely or never replicate an earlier portion of its trajectory.

The simplest and most common chaos model involves a nonlinear one-variable difference equation of first order, that is, one of the form $y_{t+1} = f(y_t)$, whose graph (the *phase diagram*) showing $f(y_t)$ as a function of y_t (Figure 1) is *hill-shaped* and "*tunable*"; in other words, the height, steepness and location of the hill can be adjusted as desired by a suitable modification in the values of the parameters of $f(y_t)$. This phase diagram, which will presently be described explicitly, is the geometric instrument used to analyze the time path generated by a difference equation model, and it is employed extensively in chaos analysis.

The function most commonly used to illustrate the chaos phenomenon is the quadratic equation with a single parameter, w

$$(1) \qquad y_{t+1} = f(y_t) = wy_t(1 - y_t), \quad \text{where} \quad dy_{t+1}/dy_t = w(1 - 2y_t).$$

The *phase curve* (the hill-shaped curve) in Figure 1 is defined as the graph of the difference equation $y_{t+1} = f(y_t)$; in this case, it shows this function when $w = 3.5$. The figure also shows the generation of a time path graphically, using the phase curve in the manner familiar in elementary economic dynamics, to find y_1 from y_0 (point A), next, using the 45° ray to transfer y_1 from the vertical to the horizontal axis (point B) then repeating the procedure to find y_2 from y_1 (point C) and so on.[2]

Figure 2 shows that the height of the phase curve hill, and its slope at the (equilibrium) intersection point with the 45° ray depends on the value w of the parameter(s) of the difference equation. There are four general cases, each illustrated

---

shown to arise in a number of competitive models of intertemporal general equilibrium. These results are of particular interest since they demonstrate that prices and outputs can oscillate even under standard competitive assumptions such as market clearing, perfect information and perfect foresight. For overlapping generations models of exchange, Benhabib and Day (1982) have provided sufficient conditions for cyclic and chaotic dynamics under perfect foresight when the young are net borrowers (the classical case; see Gale, 1973). Grandmont studied the case where young are net savers (the Samuelsonian case; see Gale 1973) and correctly learn to forecast periodic equilibria. In equilibrium models with infinitely lived agents Benhabib and Nishimura (1979; 1985, 1989), provided sufficient conditions for cyclic equilibrium and recently, albeit in a more abstract setting, Boldrin and Montracchio (1985) and Deneckre and Pelikan (1984) have shown that chaotic trajectories can occur in such models. See also Woodford (forthcoming). These are only part of a growing list.

[2]We see immediately from (1) that whatever the value of w, the graph (the *phase curve*) for the equation always must reach its maximum at $y_t = 0.5$, where $dy_{t+1}/dy_t = w(1 - 2y_t) = 0$. At that point its height is $w\ 0.5\ (1 - 0.5) = w/4$ which increases in proportion to w.

Fig. 1. Phase diagram, periods 0–9 $y(t+1) = 3.5y(t)[1 - y(t)]$, $y(0) = 0.034$

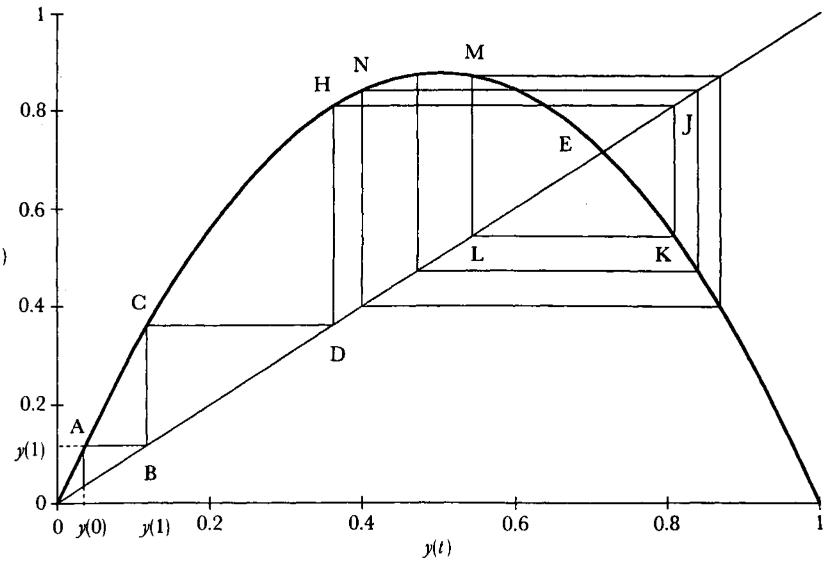

Figure 1 can be used to remind the reader in a bit more detail how one uses a phase diagram to calculate the behavior of an ongoing time path. We must start our calculation at some arbitrarily chosen date, call it $t = 0$, at which the value of y_t is indicated by the point labelled $y(0)$ on the horizontal axis of the graph. Then the next period's value of y, that is, $y(1)$, is given by the height of point A on the phase graph of $y_{t+1} = t(y_t)$ directly above $y(0)$. Next, we want to repeat the process, this time starting from $y(1)$, in order to find $y(2)$, the next value of y_t. In order to do this, we must, obviously, first transfer $y(1)$, i.e., the height of point A, from the vertical axis to the horizontal axis. For this purpose we first move horizontally from point A to point B on the 45° line. The point directly below B is the position of $y(1)$ on the horizontal axis, because the two coordinates of any point of the 45° line must be equal. Having found $y(1)$ on the horizontal axis we now move directly upward to point C on the phase graph, whose height is $y(2)$. Continuing in this way we trace out the time path of y_t.

We notice that in the leftward region of the diagram, where the phase graph is upward sloping, the time path ABCDH... does not change direction (i.e., in this case it goes steadily upward). Thus, as, e.g., where $y_t = a^t y_0$ with $a > 0$, the time path has no oscillations. However, toward the righthand end of the diagram, where the phase graph has a negative slope (as where $a < 0$ in $y_t = a^t y_0$) the time path starts to oscillate. It goes up and down in a cobweb pattern (such as HJKLM) around the equilibrium point E. (E is the equilibrium point since that is where the phase curve cuts the 45° line, so that there $y_{t+1} = y_t$, as equilibrium requires.

in Figure 2: As is easily shown directly from equation 1, (1) if $w < 1$ the phase curve will lie entirely below the 45° ray in the positive quadrant;[3] but if $w > 1$ there will be a positive valued ($y_t > 0$) intersection (equilibrium) point, E, between the phase curve and the 45° ray. In particular, (2) if $1 < w < 2$, the phase curve's slope at the intersection point will be positive; (3) if $2 < w < 3$ that slope will be negative but less than unity in absolute value, and (4) if $w > 3$ the slope will be less than -1.

For two reasons, this last case, $w > 3$, is the one that is of interest here. First, since the slope of the phase curve at the equilibrium point is negative, then the

[3]For at a crossing of the phase line with the 45° ray $y_{t+1} = y_t = y_e$ so that $y_e = wy_e(1 - y_e)$ or $wy_e^2 = (w - 1)y_e$. Hence, $y_e = (w - 1)/w$ will not have a positive value for $w < 1$. Note that, by (1), at E the slope of the phase graph is $w(1 - 2y_e) = w - 2(w - 1) = 2 - w$.

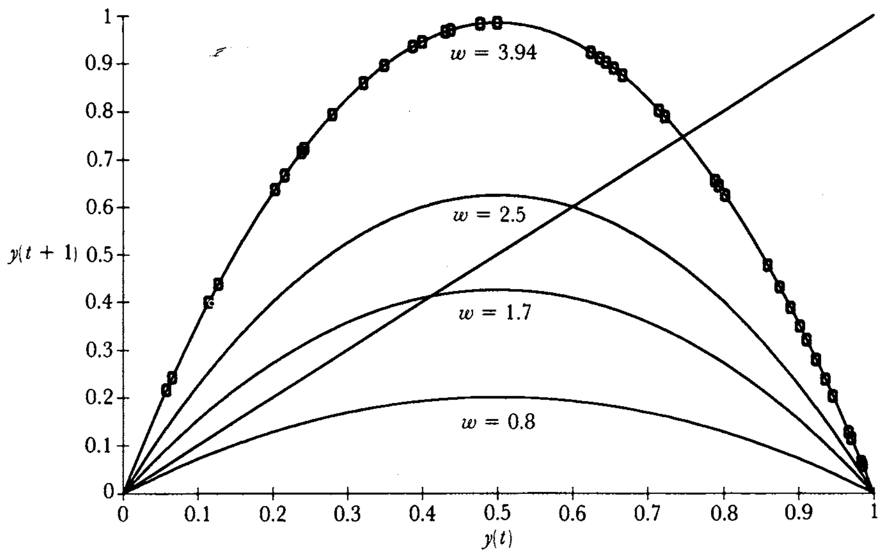

Fig. 2. $y(t+1) = wy(t)[1 - y(t)]$ various values of w and a simulated path.

elementary theory of difference equations tells us that the time path must be oscillatory. The cobweb-like oscillations will involve a rise and a fall in every two successive periods, with the high point of one period, y_t, followed by the low point, y_{t+1}, of the next, just as in $y_t = a^t y_0$ when $a < 0$. Second, since the slope is greater than unity is absolute value, the oscillations will be explosive (of ever-growing amplitude), moving ever further away from the equilibrium value, y_e, in the neighborhood of that value of y. An example of such a time path is the case where $y_{t+1} = -2y_t$ and $y_1 = 1$, so that $y_2 = -2$; $y_3 = +4$; $y_4 = -8$; and so on. In Figure 1 such a path is shown by the cobweb-like trajectory surrounding point E.

If the graph were not hill-shaped that would be the end of the story, with the cobweb oscillations moving ever further away from the equilibrium point. However, with a hill-shaped phase curve, eventually, as the cobweb expands it will encounter the positively sloping side of the hill and "bounce off it" at a y_{t+1} value closer to the equilibrium level than some earlier y_{t+1}. (Thus, in Figure 1 the height of point N is closer to that of equilibrium point E than is earlier point M's.) Such a return toward equilibrium *must* occur eventually, since as the cobweb expands further to the left during its explosive stage, its height in the next move that follows must be reduced because the slope of the pertinent portion of the graph is positive. When this happens, the cycles will begin converging toward E once more, but that can only be temporary, because the slope of the phase curve at E is greater than unity in absolute value, so that E is an unstable equilibrium that generates an explosive time path, as we have seen.

The analogy with a billiard ball bouncing off the sides of the table in a complicated pattern is suggestive here. It is easy to imagine why, in such circumstances, the time path can turn out to be complex, as chaos requires. What is rather more surprising is that the pattern of chaotic behavior will then follow some very simple and orderly rules.

To understand this orderly pattern it is necessary to describe chaotic behavior more carefully. While the discussion is based largely on our illustrative chaos equation $y_{t+1} = wy_t(1 - y_t)$, exactly the same sort of behavior holds for a very wide set of relationships $y_{t+1} = f(y_t)$ whose graph is hill-shaped and "tunable" by adjustment of the parameter values.

Let us first preview the results that will be shown, initially without explanation. We will see that for values of w slightly less than 3 there will be cobweb time paths whose oscillations converge to a stable equilibrium point E (Figures 3a and 3b). When w increases slightly above 3, these time paths will be replaced by one which is explosive but which converges to a stable limit cycle, two periods long (the rectangular trajectory LL′HH′ in phase diagram 3c and the corresponding time path in Figure 3d with its clear two-period limit cycle). Then, for higher values of w, the two-period cycles will in turn (at a known value of w) give rise to a cycle of four-period length. At a higher value of w this cycle will in turn give birth to an 8-period cycle, and then to a 16-period cycle, etc. Here, a term such as "four-period cycle" means a path such as "high point; low point; still higher point; still lower point," which is then repeated over and over again.

We begin the story just where the tuning (or controlling) parameter attains a value ($w = 3$ in our case) at which the equilibrium point, E, becomes unstable because the slope of the phase curve exceeds unity in absolute value. *Exactly* at that value of w the stable two period limit cycle, y_1, y_2 (with $y_3 = y_1$) will be shown to make its appearance. Then, as w increases further (to $w \approx 3.4495$ in our case) the two-period cycle, in its turn, will become unstable. At exactly that value of w we will see that a stable four-period cycle, $y_1^*, y_2^*, y_3^*, y_4^*$ (with $y_5^* = y_1^*$) makes its appearance. Two of these four points, say, y_1^*, y_3^*, of the four period-cycle are generated from one of the points, say y_1, of the two-period cycle, via a process called "bifurcation" (which will be explained presently), while the other two new points y_2^*, y_4^* "bifurcate" from the other point of the two-period cycle.

As the value of the parameter w is increased further, the four-period cycle itself becomes unstable in its turn, and from *each* of the four values $y_1^*, y_2^*, y_3^*, y_4^*$ that constitute the four-period cycle, two additional points emerge via a new bifurcation. These new eight points now constitute a stable eight-period cycle. When the four-period cycles are introduced, the two-period cycle remains present; similarly, when the eight-period cycles enter, those of four and two periods remain. As w is increased, the period-doubling bifurcation tale will repeat itself, successively adding new stable cycles. Ultimately, at suitable values of the parameter, the time path must involve an infinite number of cycle lengths.

At first, these will only have even periods of increasing length, all of them powers of two. Eventually, cycles whose length involves an odd number of periods will appear. The first such odd-period cycles to enter the time path will be very long, but they will be joined by odd-period cycles of shorter and shorter duration, encompassing every positive odd integer (for a complete characterization, see Sarkovskii, 1964).

Finally, at some value of the controlling parameter w even three-period cycles will occur. There will then be an uncountable number of initial values yielding bounded time paths which *never* repeat any past behavior, no matter how long a set of time periods one permits the calculation to encompass. When this set of conditions holds, one may wish to say that chaos is present.

It should be made clear, however, that while an infinite number of cycle lengths are present then, they need in general not all be equally influential upon the time

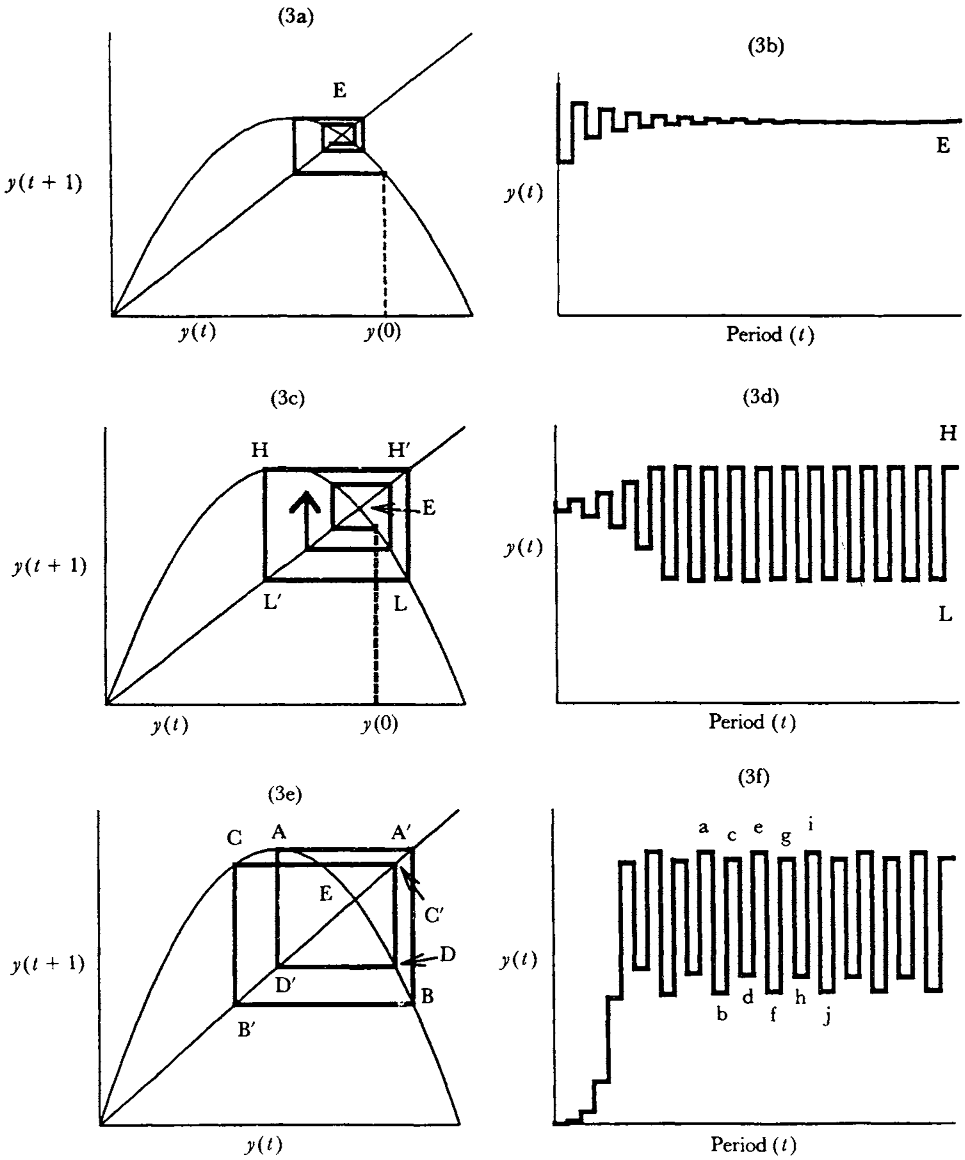

Fig. 3. Three types of time path and equilibria.

path. For example, when the four-period cycle enters the picture, the two-period limit cycle continues to be present in the structure of the model, but since it is now unstable it will generally no longer influence the long-run behavior of the time path. Similarly, when the eight-period cycle emerges, the four-period limit cycle will remain in the background, largely deprived by its instability of influence on the time path.

Figures 1 and 3e show for the case $w = 3.5$ how the four-period limit cycle is gradually approached. In addition to showing $f(y_t)$, Figure 1 also includes the time path of the first 10 periods, starting with $y_0 = 0.034$. The resulting messy cobweb displays oscillatory behavior about point E, but the pattern is not obvious. However, Figure 3e represents the time path toward which y_t converges as t approaches infinity. We see here that the time path has settled down into a recurrent pattern (so that w has not yet entered into the region of chaotic behavior). We seem to have two nested cycles with the time path alternating between them. The cyclical path AA′BB′ does not return to starting point A but instead goes to neighboring point C; then it follows the cycle CC′DD′ and then, apparently as something of a miracle, returns to starting point A of the other cycle. However, this is no great coincidence, but a normal part of the process, for reasons the analysis of the next section will indicate.

To see just how this produces a four-period cycle we turn to Figure 3f, which is the time path generated by the phase diagrams in Figures 1 and 3e. At first glance we see only a persistent (but imperfectly replicated) oscillation exactly two periods in length, which clearly dominates behavior. But superimposed on it are two cyclical disturbances, each four periods long, which together constitute the four-period cycle. To see them one must first look exclusively at the upper horizontal segments a, c, e, g and i of the time path, and then by looking in turn only at the lower segments, b, d, f, h and j.

The upper segments describe the first of the oscillations. Starting from a, *and skipping one period*, y_t falls to c. Then, after a gap of another period, y_t rises again to e. Continuing in this way we see that one has an oscillatory disturbance whose high points are a, e, i, ... and whose low points are c, g, ..., with four periods elapsing between, say, one high point and the next. This oscillation corresponds to the difference in height between horizontal segments AA′ and CC′ in the time path of the phase diagram 3e. The reader will now readily recognize the other four period oscillatory disturbance in Figure 3f by looking at the lower horizontal segments b, d, f, h, and j. The net result (in the limit) is a single four-period cycle, that is, a cycle that repeats itself precisely every four periods, as is confirmed by careful examination of the right hand end of the time path in Figure 3f.

The Orderly Structure of Chaotic Behavior[4]

Let us begin to consider, now, why chaos should be approached by such an orderly progression: as w is increased, first there is one stable two-period cycle; then at

[4]The bulk of the following discussion is based on the beautiful analysis in May (1976).

the value of w at which the two-period cycle becomes unstable, a stable four-period cycle emerges; that four-period cycle, as we have seen, is produced by two sets of oscillation one around the top points and one around the lower points of the two-period cycle. Then, at the value of w where the four-period cycle becomes unstable, eight-period cycles appear, and so on, *ad infinitum*. We will seek to provide an intuitively appealing view of the matter.

We start by studying how the stable two-period cycles enter, since the mechanism of the subsequent introduction of 4, 8, 16 and higher period cycles is perfectly analogous. Because we want to study the transition from a stable equilibrium point E to a stable two-period cycle we must extend our analytic tools correspondingly. An equilibrium *point* is for obvious reasons defined by the requirement $y_{t+1} = y_t = y_e$. This clearly compares the values of y in *two successive periods*.

But if an equilibrium is constituted by a two-period cycle it is, obviously, defined by $y_t = y_{t+2}$, $y_{t+1} = y_{t+3}$. This clearly requires comparison of the values of y in every *other* period. Thus, to investigate the genesis of two-period cycles we need a relationship between y_{t+2} and y_t, not one between y_{t+1} and y_t. Such a two-period relationship is obtained via a second iteration of the relationship $y_{t+1} = f(y_t)$. That is, we first find y_{t+1} from $f(y_t)$ and then we find y_{t+2}, in turn, from $f(y_{t+1})$ to give us $y_{t+2} = f[f(y_t)]$. To carry out this calculation, if, for example, $y_{t+1} = wy_t(1 - y_t)$ so that also $y_{t+2} = wy_{t+1}(1 - y_{t+1})$, we substitute the expression for y_{t+1} from the first equation into the second to obtain a direct relation, $y_{t+2} = f[f(y_t)]$, between y_{t+2} and y_t. (Henceforth, we will use $f^{(2)}$ to represent $f[f(y_t)] = y_{t+2}$, $f^{(3)}$ to represent $f\{f[f(y_t)]\} = y_{t+3}$, and so on.)

To tell the story of the introduction of the two-period cycles we must first consider some properties of the graph of the general two-period relationship $y_{t+2} = f^{(2)}(y_t)$. This may be considered the equation of a two-period phase curve in the graph which has y_{t+2} rather than y_{t+1} on its vertical axis. In Figure 4a such a phase curve [labelled $y(t+2)$], with its typical double hump, is superimposed on the hill-shaped two-period phase curve [labelled $y(t+1)$].

Let us examine the relations between the two phase curves, $y(t+2)$ (or $f^{(2)}$), and $y(t+1)$ (or f). First we note that the two-phase curves in Figure 4a cross the horizontal axis at the same points, that is, at the points $y_t = 0$ and $y_t = 1$. This is generalized in *Proposition 1*:[5] If the graph of f goes through the origin, then all roots of f (that is, points at which $y_{t+1} = f(y_t) = 0$) must also be roots of $f^{(2)}$.

Next, we note that the two-phase curves cross the 45° ray at a common equilibrium point, E. This is generalized in *Proposition 2*: Any equilibrium point of f must also be an equilibrium point of $f^{(2)}$.

We also have *Proposition 3*:[6] The slope of $f^{(2)}$ at an equilibrium point of f must equal the square of the slope of f. That is, at any $y_e = y_t = y_{t+1}$ we must have $df^{(2)}/dy_t = (df/dy_t)^2$.

[5] *Proof* of proposition 1: Let y^* be a root of f. Then, since $f(y^*) = 0$, $f[f(y^*)] = f(0) = 0$

[6] *Proof* of proposition 2: If y_e is an equilibrium point of f then $f(y_e) = y_e$, so that $f^{(2)}(y_e) = f[f(y_e)] = f(y_e) = y_e$. *Proof* of proposition 3: $df^{(2)}/dy_t = (dy_{t+2}/dy_{t+1})(dy_{t+1}/dy_t)$, but at $y_t = y_{t+1} = y_e$ we must have $dy_{t+1}/dy_t = dy_{t+2}/dy_{t+1} = df/dy_t$.

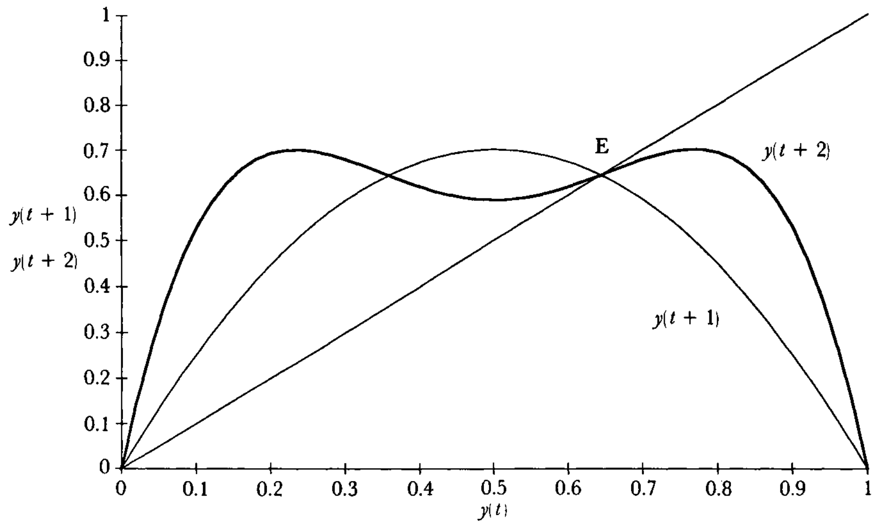

Fig. 4a. $f[y(t)]$, $f\{f[y(t)]\}$, $y(t+1) = f[y(t)] = 2.8y(t)[1 - y(t)]$

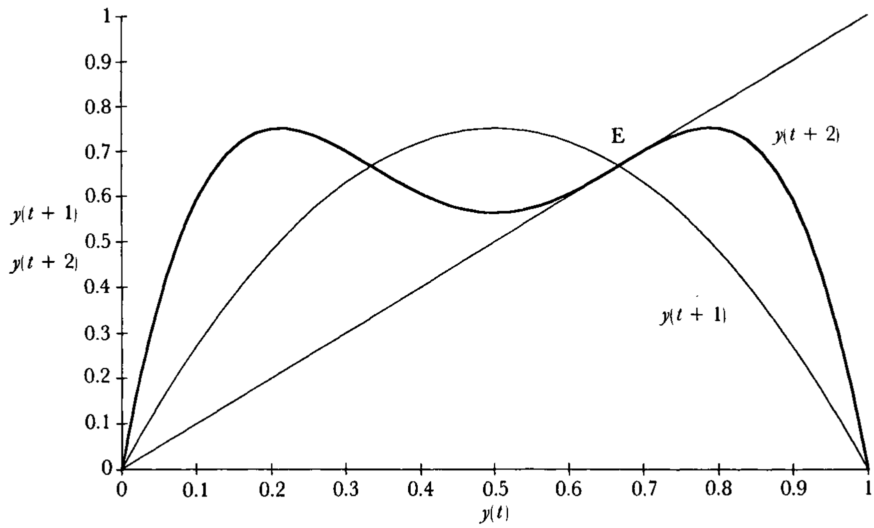

Fig. 4b. $f[y(t)]$, $f\{f[y(t)]\}$, $y(t+1) = f[y(t)] = 3y(t)[1 - y(t)]$

Comment: Proposition 3 must hold, in particular, at the origin, where $y_{t+1} = y_t = 0$. *Note that corresponding propositions also hold for any* $f^{(n)}$ *where n is any positive integer.*

The final key observation linking the graphs of $f^{(2)}$ and f is that where (as in the case of equation 1) the basic relationship, $f(y_t)$ is quadratic (it includes a term with y_t^2 in it), and consequently has one peak, the relationship $f^{(2)}(y_t) = f[f(y_t)]$ will be of fourth degree; it has a term involving y_t^4, as the reader can verify by direct substitution. (For example, if $y_{t+1} = y_t^2$, so that $y_{t+2} = y_{t+1}^2$, then $y_{t+2} = y_t^4$.) So $f^{(2)}$

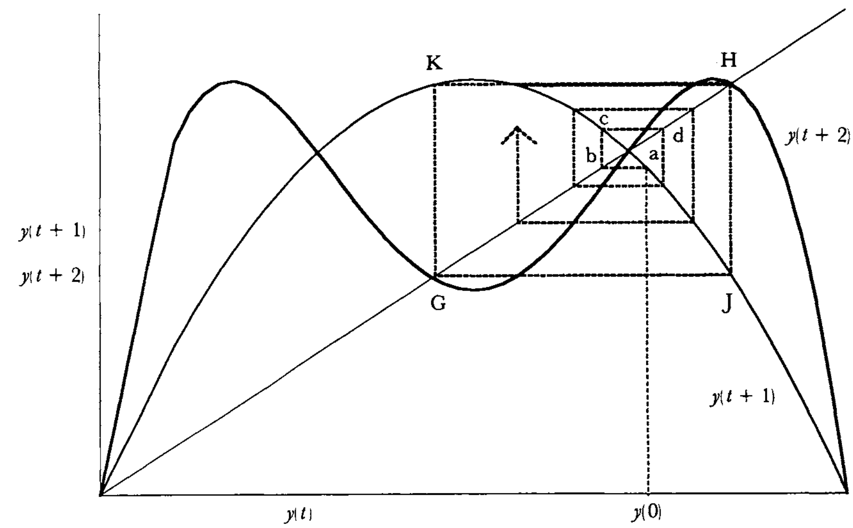

Fig. 4c. $f[y(t)]$, $f\{f[y(t)]\}$, $y(t+1) = f[y(t)] = 3.43y(t)[1 - y(t)]$

can then be expected to have either two maxima and one minimum, or the reverse. That is why the $f^{(2)}$ graph, as is shown in the figure, typically has a double hump.[7]

Figures 4a–4c illustrate the behavior of the preceding relationships as the value of w increases. Each graph is derived from our basic illustrative equation for a different value of w, as will now be explained, and we will see precisely why two new equilibrium points must appear just at the value of the tuning parameter ($w = 3$ in our example) where the initial equilibrium point E becomes unstable.

By Proposition 3, in all three graphs $f^{(2)}(y_t)$ must cut the horizontal axis at the same points as $f(y_t)$. It is easy to show that each of these curves will have a slope greater than 1 near the origin and so lie above the 45° ray. Finally, at equilibrium point E all three phase curves, $f(y_t)$, will have the slope $2 - w$ (see footnote 3) so that by proposition 3 $f^{(2)}$ will have the slope $(2 - w)^2$.

It follows at once that where (as in Figure 4a) $2 < w < 3$, then the slope of $f^{(2)}$ will be positive and less than unity. That is, the $f^{(2)}$ curve will cut the 45° ray at E *from above* as one moves from left to right.

Figure 4b shows the case $w = 3$ so that at E, $df/dy_t = -1$ exactly; and the slope of $f^{(2)}$ is $(df/dy_t)^2 = (2 - w)^2 = +1$. Then $f^{(2)}(y_t)$ will be tangent to the 45° ray at E, as the figure shows.

Finally, the full story emerges in Figure 4c where $w > 3$ so that $df^{(2)}/dy_t = (2 - w)^2 > 1$. Therefore, $f^{(2)}(y_t)$ must cut the 45° ray *from below* at E as we move from left to right. This means that as $f^{(2)}(y_t)$ leaves the origin, initially lying above

[7]This also suggests in another way how the bifurcation process works. Equilibrium points for $y_{t+2} = f^{(2)}(y_t)$ by definition require $y_{t+2} = y_t$, call their common value y_{e2}. Then the equation $y_{e2} = f^{(2)}(y_{e2})$ which gives these equilibrium points is a quartic. Two of its roots will be imaginary until w attains the value at which the new equilibria appear.

the 45° ray, at some point G to the left of E, it must cross that ray for it to be possible for that curve to cut the ray at point E from below. Similarly, to the right of E, $f^{(2)}(y_t)$ must first lie above the ray, and so must cross it again at some point, H, in order to reattain the horizontal axis again at $y_t = 1$. We see that happen just at that value of w at which the slope of the $f(y_t)$ graph begins to exceed unity in absolute value, so that when its equilibrium becomes unstable, there appear two new intersection points G and H, which have no counterparts in Figure 4a. These are the two new equilibrium points[8] that constitute the bifurcation which generates a two-period cycle from the equilibrium E. This cycle will be stable because the absolute value of the slope of $f^{(2)}$ at G and H is negative and less than 1. The limit cycle is approached by an oscillatory cobweb which is shown in Figure 4c by the time path abcd... which, analogously to that in Figure 3c, approaches the rectangular limit cycle HJGK. It should be noted that the size and position of the limit cycle are given by the two corner points G and H, which are clearly the stable two-period equilibrium points of intersection between the curve $y_{t+2} = f^{(2)}(y_t)$ and the 45° ray. This limit cycle entails a perpetual and successive rise and fall of y_t from the height of point G to that of point H, and so on, forever.

How Four-Period Limit Cycles Arise

We can now quickly see by analogy how the two-period equilibrium points G and H that generate the two-period limit cycle in Figure 4c become unstable and four new equilibrium points then appear. The story is a precise replication, with a second bifurcation step, of that in the previous section. As the value of w increases, the absolute slopes of $f^{(2)}(y_t)$ at G and H will increase monotonically, and must ultimately exceed unity, so their surrounding cobwebs must again become unstable. If we form the function $y_{t+4} = f^{(4)}(y_t) \equiv f(f(f(f(y_t))))$, for precisely the same reasons as before, the curve representing $f^{(4)}$ will become tangent to the 45° ray at G just when the slope of $f^{(2)} = -1$, and the same will be true at H. For w slightly larger than this, G and H will each be surrounded, via bifurcation, by two new equilibrium points, producing a four-period limit cycle, just like that shown in Figure 3e. These four new equilibrium points constitute a four-period limit cycle which is initially stable but will grow unstable as w increases still further. The process obviously can repeat itself *ad infinitum*, thus giving rise to an infinite set of superimposed oscillations, each one with a period that is a power of two.

How Three-Period Cycles Arise

So far we have dealt only with cycles whose duration is an even number of periods. However, it is now easy to show how odd-period cycles arise, using essentially

[8]It can be proved that the slope of the phase curve at equilibrium points G and H must be the same, and the analogous result holds at any equilibrium points that emerge at any subsequent bifurcation.

the same approach as before. To find the three-period cycles, for example, we plot the phase diagram for $y_{t+3} = f^{(3)}(y_t) = f(f(f(y_t)))$.

If $f(y_t)$ has a single hill, $f^{(3)}(y_t)$ will normally exhibit four hills when the tuning parameters have made the $f(y_t)$ hill sufficiently steep. The graph is shown in Figures 5a and 5b. In 5a, with w relatively small, the phase curve only crosses the 45° ray once, at a location, E, which is also the nonzero equilibrium point of $f(y_t)$. However, as w increases in value the hilltops of $f^{(3)}$ will rise and the valleys will deepen, and eventually the 45° ray will be crossed seven times. The corresponding six new crossings will correspond to points on two distinct cycles of three periods, one of which is stable and the other unstable. Unlike the bifurcations of cycles of even order which derive from cycles of lower period as their equilibrium points undergo a loss of stability, odd period cycles do not bifurcate from lower order cycles, but simultaneously emerge or disappear in pairs, with a stable and unstable cycle constituting each pair.

Chaos and Strange Attractors

Despite its aura of erotic kinkiness, "strange attractor" is a technical term which offers yet another insight into the workings of the chaos phenomenon. An *attractor* is what most of us might describe as the equilibrium or limit time path of a stable dynamic system, whether or not that system is chaotic. For example, the difference equation $y_{t+1} = 0.5y_t$ clearly converges toward the equilibrium value $y_e = 0$ so that any time path of the equation, whatever the initial point, will converge in the limit to the origin in the phase diagram. The origin is then said to be the attractor for this relationship; and in this case the attractor is clearly a single point.

In other cases, the attractor is more complex. For example, all time paths of the system may be cobwebs which converge toward a simple rectangle in the phase diagram. This means that the time path will settle down in the limit to a two-period oscillation—a repeated traversing of that rectangle, going endlessly back and forth from its upper to its lower edge and then back up again. Here the attractor is the rectangle, that is, it is a two-period limit cycle, toward which all time paths of the system converge (Figure 3c).

Attractors can grow more complex still, as illustrated in Figures 1 and 3e. In the former we see a complicated cobweb path which converges to the attractor shown in Figure 3e, an attractor which can perhaps be described as a pair of intertwined rectangles. The result is an equilibrium time path involving somewhat messy oscillations approximating those in Figure 3f.

Now intuition suggests, correctly, that in the stable case, as the attractor of the system is made increasingly complex by changes in the pertinent parameter values, some time paths will increasingly take on chaotic attributes. Before defining a strange attractor we must note that a bounded time path is called *aperiodic* if it will never return to any point it had previously visited. Second, an infinite set is called

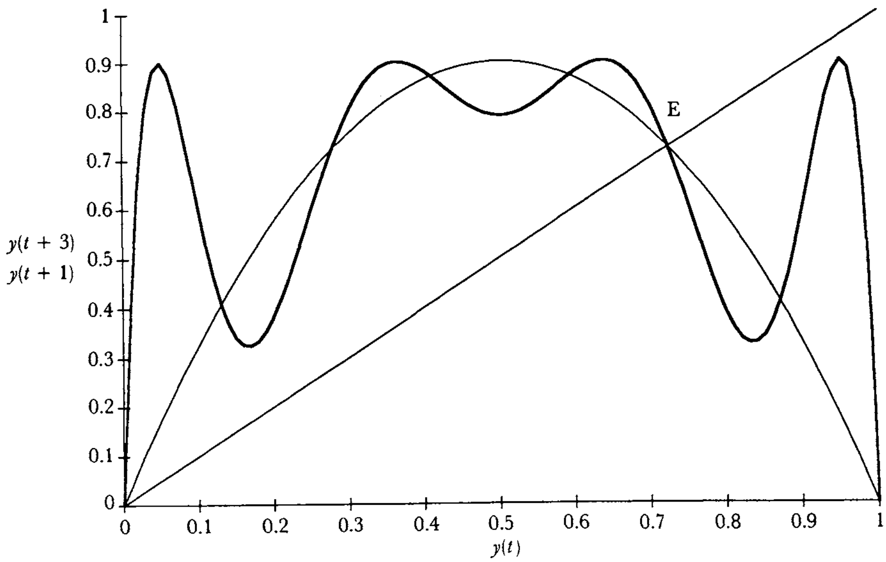

Fig. 5a. $y(t+3) = f(f\{f[y(t)]\}), y(t+1) = 3.6y(t)[1 - y(t)]$

"countable" if its elements correspond in number to that of the set of all integers. Otherwise, the set is called *uncountable*.

Turning next to a *strange attractor* we may think of it, roughly, as a set of points toward which complicated paths starting off in its neighborhood are attracted. More accurately, it is an uncountable set of points such that all time paths that start off within it will remain in that set, that neighboring time paths will be attracted to it and such that time paths that start in the set can be aperiodic or whose period is as long as any arbitrarily preselected number.

It is possible to provide pictures of strange attractors, but they are sufficiently convoluted that it is fairly difficult to do so without recourse to three-dimensional colored diagrams (for nice examples, see Crutchfield, Farmer, Packard and Shaw, 1980, pp. 50–51).

Sensitivity of the Time Path

An important matter, especially for forecasting purposes, is whether initial points that are close together give rise to time paths that diverge. Such divergence has been termed "sensitive dependence to initial conditions" and has attracted a great deal of attention in disciplines outside of economics which have also utilized chaos theory. For example, meteorologists have dubbed this sensitivity the "butterfly's wing phenomenon." They refer to the possibility that a butterfly fortuitously flapping its wings in Hong Kong can cause tornados in Oklahoma if weather is controlled by chaotic relationships.

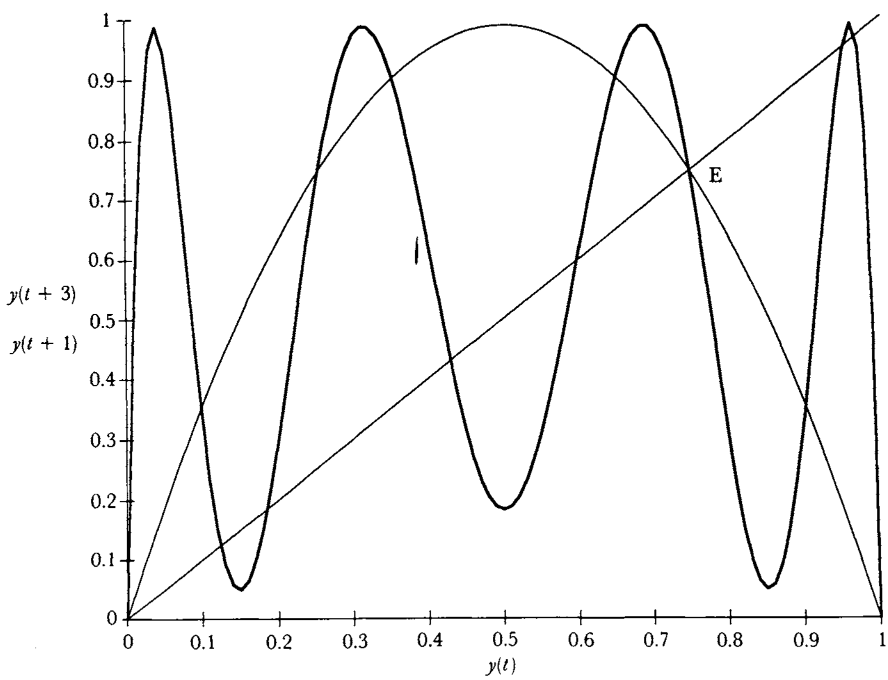

Fig. 5b. $y(t+3) = f(f\{f[y(t)]\}), y(t+1) = 3.95y(t)[1 - y(t)]$

A few graphs will illustrate this degree of sensitivity. In Figures 6a and 6b there is no difference in initial conditions or anything else, except that in 6a $w = 3.935$, while in 6b $w = 3.94$. The figures show that changes in the third decimal place in the parameter value can transform the entire picture unrecognizably. A similar qualitative jump occurs if we move to $w = 3.945$. It is easy to demonstrate that far smaller changes in the value of the parameter can cause similar upheavals. Not only that. If we hold the parameter value constant and change the initial condition by microscopic amounts in a chaotic regime equally startling qualitative changes in the time path may follow.

The sensitivity of the time path of a variable governed by a chaotic time path can be brought out in another way. In a calculation by Richard Quandt, the time path of our illustrative equations $y_{t+1} = wy_t(1 - y_t)$ was determined twice, each time for 640 periods. The first calculation was carried out by a process that rounded after 7 decimal places while, in the second, rounding occurred after 14 decimal places. With w sufficiently low so that the equation had not yet actually entered the chaotic region, the two calculated time paths remained virtually identical even after 600 iterations. In contrast, with a w value sufficiently large to produce chaos, after only 30 iterations the two series lost virtually any resemblance to one another.

Sensitive dependence to initial conditions will not be observed if there exist stable periodic time paths that attract trajectories from almost all initial points. For the quadratic case there will be many values of w between 3 and 4 for which stable orbits

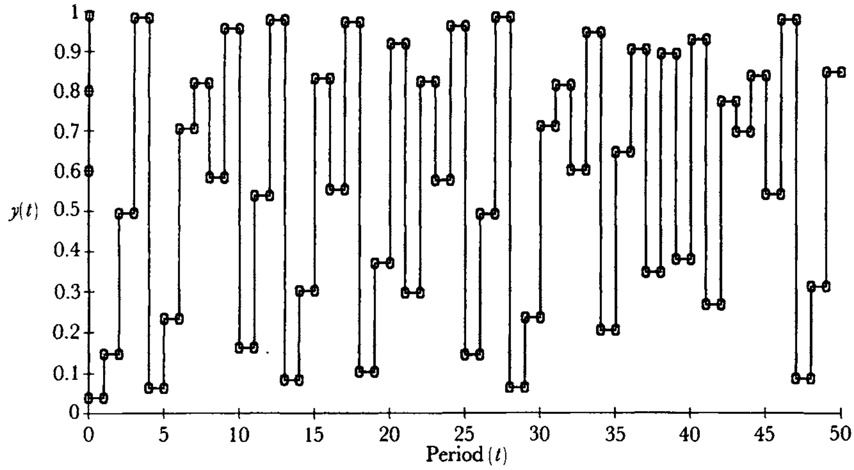

Fig. 6a. Time path, periods 0–50, $y(t + 1) = 3.935y(t)[1 - y(t)]$, $y(0) = 0.99$

exist but there will also be a "large" set of w's for which trajectories will be sensitive to initial conditions. The studies of Shaw (1981) also suggest that for the quadratic case sensitive dependence will be prevalent for values of w between 3.5 and 4 (in the sense that sensitive dependence will occur for a set of w's of positive Lebesque measure, a standard measure of the number of points in a set, defined in terms of the total area or volume completely occupied by those points; see Jacobson, 1981).

These figures indicate the difficulties that are apt to beset forecasting in the presence of chaos. Even a forecasting procedure of unprecedented accuracy is likely in such a case to yield results that differ vastly from the actual course of future developments.

Sudden Qualitative Breaks in the Time Path

Figure 6b also dramatizes another of the characterizing attributes of chaotic trajectories—their propensity to introduce sharp and unheralded qualitative breaks in time path. From the initial point, A, of the time path until point B, some 25 periods later, there is a fairly homogeneous regime of (somewhat lopsided) cycles which seem to exhibit no clear trend in amplitude. Then, suddenly, the time path becomes almost horizontal, and for 10 periods (from B to C) cyclical behavior all but disappears. At that point, just as unexpectably, several fairly sharp oscillations arise, apparently out of nowhere, abruptly becoming very moderate again to the right of point D. It is difficult to imagine how any forecasting technique that relies upon extrapolation, direct or indirect, could have correctly predicted events during the period encompassed between points B and C from even the most accurate and fullest set of data about the 25-period interval that preceded it.

This graph suggests that chaotic behavior does not generally mimic pure randomness in the performance of its basic variable. Rather, the time path can resemble

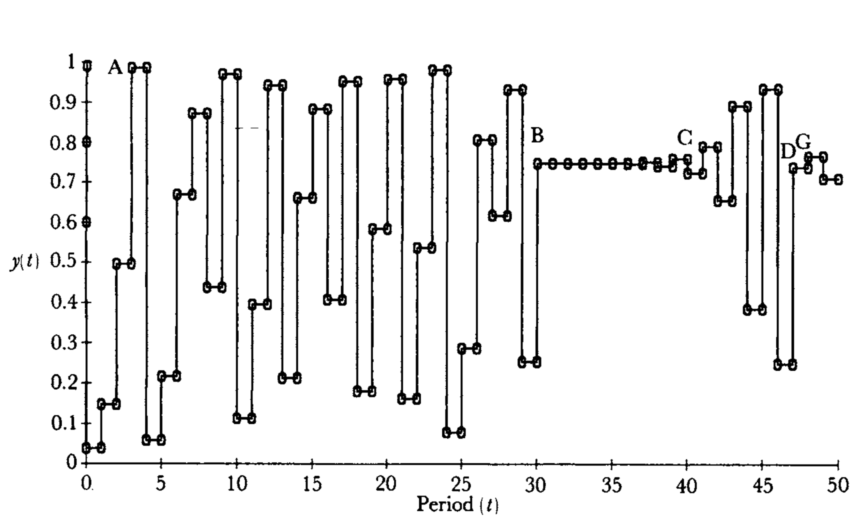

Fig. 6b. Time path, periods 0–50, $y(t+1) = 3.94y(t)[1 - y(t)]$, $y(0) = 0.99$

one that might be expected of a deterministic model, but which is at the same time subject to very large random disturbances occurring at randomly determined intervals.[9]

Some Basic Mathematical Results

Modern methods of qualitative analysis of dynamic systems go back to Poincare (1880, 1892). Since the classic work of Smale (1967), it has become clear that very complicated trajectories (time paths) can easily arise in certain dynamical systems and that such complicated trajectories can persist when small perturbations of the underlying system occur. (For a clear exposition see Guckenheimer and Holmes, 1983.) The papers of Li and Yorke (1975) and others and the work of Sarkovskii (1964) which has recently been rediscovered (see Stefan, 1977) have greatly facilitated exploration of the pertinence of such complicated dynamics, arising in simple first order dynamic systems, to a variety of fields, such as physics, biology or economics. (For a recent excellent exposition, also see Grandmont, 1986.)

Let us now describe a widely used and very useful result on chaotic dynamics, translating it into terms that economists can follow more easily. The theorem describes the superimposition of cycles of periods of different length and the resulting behavior of the time path when cycles of every integer periodicity are included. We will also

[9] Professor Quandt has carried out a simulation exercise in which the behavior of a chaotic time path generated by our basic illustrative difference equation was contrasted with one that followed an uncomplicated deterministic regime that was subject to substantial random disturbance of moderately low probability. Spectral analysis then yielded very similar results for the chaotic series and the series subject to random disturbances, properties very different from those that held for the time paths of series generated by our equation with w values not far from the chaotic region. The implication is that standard statistical procedures may fail to determine correctly in any particular case whether a set of observations has been subject to random disturbances or whether it has been generated by a model that is perfectly deterministic but chaotic. For details of the simulations see Baumol and Quandt (1985).

note a few serious pitfalls besetting interpretation of this theorem, and continue to consider alternate definitions of chaos. (Readers less interested in mathematical foundations may wish to skip over this section and the next.)

The Li-Yorke Theorem. Let f be a difference equation that is continuous, and for which there exist two numbers, $\underline{a}$ and $\underline{b}$, such that if $a \leq y_t \leq b$, then $a \leq y_{t+1} \leq b$. Now, if one can find a y_t such that when y_t rises for two successive periods it will fall back to below its initial value in the next period, that is,

$$(7) \quad y_{t+1} = f(y_t) > y_t \quad \text{and} \quad y_{t+2} = f^{(2)}(y_t) > y_t \quad \text{but } y_{t+3} = f^{(3)}(y_t) \leq y_t,$$

then two major consequences follow:

(a) For *any* integer $k > 1$ there is at least one initial point y_0 between $\underline{a}$ and $\underline{b}$ such that the subsequent time path, y_t, is characterized by cycles of period k;
(b) There exists an uncountable set, S, of initial points in the interval between $\underline{a}$ and $\underline{b}$ such that if initial points x_0 and y_0 both lie in S, then (i) at some time t in the future the difference $(x_t - y_t)$ will come arbitrarily close to zero, that is, the two paths will (temporarily) move as close to one another as may be desired; (ii) however, after some interval of close proximity the two time paths must always diverge again; (iii) moreover, no such time path will ever converge asymptotically to any periodic time path and a time path originating in S will not converge asymptotically to any time path that originates outside S.

While a difference equation that, depending on the initial condition selected, will either give rise to time paths with cycles of any period length, as in (a) above, or can generate aperiodic time paths described in (b) above, may be said to "generate chaotic dynamics," this definition of chaos may be misleading. As has been noted already, the aperiodic trajectories, whose existence is shown in the Li-Yorke theorem, will normally be generated by an *uncountable* infinity of different initial conditions. Speaking very roughly, this may make it appear that this region in the realm of initial condition is "very large." Yet, the Lebesque measure of these initial points may be zero, that is, for some chaotic models their behavior may be *non*chaotic "almost everywhere." Here it should be noted that the Lebesque measure is defined so that the measure of any interval is equal to its length, while a set of "isolated" points, even if there is a nondenumerable infinity of them, has Lebesque measure zero (that is, these points can be covered with a countable set of intervals whose total length is arbitrarily small).

Under certain conditions it has been shown that the time paths will have at most one stable periodic orbit which will attract all initial points except for a set sufficiently small that its Lebesque measure is zero. To illustrate the point, again consider the family of equations $y_{t=1} = wy_t(1 - y_t)$, with w lying between 2 and 4. As has been seen, as the value w rises the time path will be subject to a sequence of period-doubling bifurcations, with that period tending to infinity as w approaches a limit point w' lying between 3 and 4. Now for many values of w between w' and 4 there will be a single attracting and periodic time path, with a period that is not a power of two, and toward which almost all time paths starting off from initial points between zero and one will converge. For many values of w between w' and 4 our equation will be chaotic in the sense of Li and Yorke, in that the set of initial points that give rise to "aperiodic" time paths, is infinite and, indeed, uncountable. Yet this set of points can be so small as to be of (Lebesque) measure zero, with almost all initial points being followed by time paths that converge to the stable cycle that was just mentioned. For such a value of w the chaotic behavior can be "unobservable." That is, almost all choices of initial point will lead to behavior that is simply periodic.

On the other hand, there will also exist many other values of w between w' and 4 for which there is a set of initial points S sufficiently large to have measure greater than zero, each such initial point yielding an aperiodic time path which can therefore be described as "chaotic." We can interpret this as a case in which chaos is "observable" because the number of initial points generating chaotic time paths is so large. Moreover, here any such chaotic time path which starts off from a point in set S must remain in that set.

In such a case we can define a limiting frequency with which the aperiodic (chaotic) time paths visit particular subsets of S. This permits us to interpret observable chaos as a case in which such a frequency distribution can only be defined on a set sufficiently large to have positive measure. It is, however, not known yet whether "observable chaos" is implied by sensitive dependence on initial conditions. It is also possible to show in some cases that the frequency distribution will be the same for almost all initial points in S (see Woodford, forthcoming).[10]

The key issue that remains to be settled is whether the class of difference equations that give rise to "observable chaos," or to strange attractors, itself constitutes a large set. In particular, for our illustrative family of difference equations (1), the question is whether the set of values of w that gives rise to observable chaos is large. Jacobson (1981) has shown that this set has positive (Lebesque) measure. But this does not rule out the possibility that the set S is small under still another definition: S may contain no (complete) intervals (see Jacobson, 1981; Collet and Eckmann, 1980).

Matters are different if one turns to difference equations whose graph is piecewise linear and has the shape of an inverted V. If the slope of this graph is everywhere greater than unity in absolute value (except at the apex, where slope is, of course, not defined), then it is easily shown that the equation will give rise to aperiodic time paths from almost all initial points and all of its periodic time paths will be unstable. Moreover, the chaos in this case must be observable in the sense just defined. If we perturb (change the parameter values of) such a difference equation slightly, but the slope continues to be greater than unity everywhere that the equation is defined,

[10]Nonetheless, since the value say of y_{t+1} is completely determined by y_t in a deterministic system, using this frequency distribution is of no use for forecasting. (The conditional distribution of y_{t+1} given y_t is degenerate.) It is also not possible to forecast y some periods ahead by a linear stochastic difference equation that somehow makes use of this frequency distribution. Therefore, for modelling purposes, the temptation actually to treat the dynamics of the nonlinear system as if it were generated by a simple stochastic system should be resisted.

except the apex, these properties will clearly continue to hold. An economic equilibrium model which demonstrably gives rise to observable chaos that is also robust under perturbation is provided by Woodford (forthcoming).

Chaos in Higher Order and Multivariate Systems

A number of economic models use n simultaneous difference equations of first order to relate a vector of n dated variables $(x_{1t}, x_{2t}, \ldots, x_{nt})$ to their values $(x_{1t+1}, x_{2t+1}, \ldots, x_{nt+1})$ in the subsequent period. Other models employ an nth order difference equation in a single variable $y_t = f(y_{t-1}, \ldots, y_{t-n})$, like the justly famed Samuelson model of the accelerator-multiplier cycle. Any theorem about chaotic behavior in n simultaneous first order systems (that is, about systems whose variables are n dimensional) must also apply to a single nth order equation. This is so since, as is well-known, such an nth order equation can easily be rewritten as the simultaneous first order system in n variables

$$x_{1t} \equiv y_{t-1}; \quad x_{2t} \equiv y_{t-2}, \ldots, \quad x_{nt} \equiv y_{t-n}, \qquad y_t = f(x_{1t}, \ldots, x_{nt}).$$

Almost all of the chaotic economic models referred to in this paper employ one-dimensional (that is, single variable first-order) difference equations. One exception is the paper by Benhabib and Day (1981) which studies the dynamics of endogenous choice and provides conditions on preferences under which chaotic choice sequences of n-commodity vectors arise under stationary conditions. They use the results of P. Diamond (1976) which generalize the Li and Yorke (1975) propositions to the n dimensional (n variable or n period lag) case. A further generalization is also reported by Llibre (1981) and Marotto (1978). However, these results all are subject to the same limitations that beset the Li-Yorke result. Though they imply that there is an infinite (and uncountable) set of initial values that give rise to a time path sufficiently complex to exhibit no cycles (it is "aperiodic"), they do not show that the set must be sufficiently great to have a positive Lebesque measure (roughly speaking, it does not have enough points to "fill in" a line segment or a region). These results are not easy to employ and their use up to now has been limited. They only suggest the conjecture that in higher-order systems sufficient conditions for chaos to arise are "easier" to satisfy than in the case of first-order systems; that is, chaos is "more likely" to occur in higher-order systems.

How Hill-Shaped Phase Diagrams Arise in Economics

The key to construction of a model in which chaotic behavior may arise, as we have seen, is the generation of a *hill-shaped* phase graph, at least if the model is built upon a difference equation of first order. We must, then, indicate how such hill-shaped dynamic relationships can arise in economics. Let us provide brief discussions of several models which have this property. We begin with an example that is an oversimplification, to say the least, but in which there is a very clear connection with the shape of phase curve in which we are interested. Its weakness is the degree of (not wholly unrealistic) irrational adherence to rules of thumb that it assumes for the firm, a problem that does not beset the bargaining model with which the paper began, or the models that follow this one.

Consider the relationship between a firm's profits and its advertising budget decision. Suppose that without any expenditure on advertising the firm cannot sell anything. As advertising outlay rises, total net profit first increases, then gradually levels off and finally begins to decline, yielding the traditional hill-shaped profit curve. If P_t represents total profit in period t and y_t is total advertising outlay, P_t can, for illustration, be taken to follow the expressions $P_t = ay_t(1 - y_t)$. If, in addition, the firm devotes a fixed proportion, b, of its current profit to advertising outlays in the following period so that $y_{t+1} = bP_t$, the first equation is immediately transformed into our basic chaotic equation (1b), with $w = ab$.

The reason the slope of the phase graph turns from positive to negative in this case is clear and widely recognized. Even if an increase in advertising outlay always raises total revenue, after a point its marginal net *profit* yield becomes negative and, hence, the phase diagram exhibits a hill-shaped curve.

A moment's thought also indicates why the time path of y_t can be expected to be oscillatory. Suppose the initial level of advertising, y_0, is an intermediate one that yields a high profit figure P_0. That will lead to a large (excessive) advertising outlay y_1 in the next period, thereby bringing down the value of profit P_1. That, in its turn, will reduce advertising again and raise profit and so on *ad infinitum*.

The thing to be noted about this process is that it gives us good reason to expect the time paths of profit and advertising expenditure to be oscillatory. *But it does not give us any reason to expect that these time paths need either be convergent or perfectly replicatory.* Exactly the same logical structure is consistent with "sloppiness" in the cycles, so that past behavior is reproduced only imperfectly in the future. That, then, is how chaotic behavior patterns can arise.

Another example has been provided in the theory of productivity growth (Baumol and Wolff, 1983). It involves the relationship between the rate of productivity growth, $(\Pi_{t+1} - \Pi_t)/\Pi_t$, (which we can write as Π_t^*) and the level r_t of R&D expenditures by private industry. Obviously, a rise in r_t can be expected to increase Π_t^*. However, because research can be interpreted as a service activity with a more or less fixed labor component, its cost will be raised by productivity growth in the remainder of the economy and the resulting stimulus to real wages. This, in turn, will cut back the quantity of R&D demanded. The result, as a formal model easily confirms, will be analogous to the corn-hog (cobweb) cycle with high productivity growth rates leading to high R&D prices which restrict the next period's productivity growth, and so reduce R&D prices, and so on. If R&D costs ultimately increase disproportionately with increases in productivity growth it is clear that the relation $\Pi_{t+1}^* = f(\Pi_t^*)$ can generate the sort of hill-shaped phase graph that is consistent with a chaotic regime.

Another model that can generate cyclic or chaotic dynamics is a standard growth model of Solow type in which the propensity to save out of wages is lower than that for profits (for a more complex version of this model see Akerlof and Stiglitz, 1969). Suppose that at low levels of capital stock K one obtains increasing marginal returns to increased capital and the elasticity of substitution of labor for capital is initially low; but diminishing returns eventually set in and the elasticity of substitution moves the other way. Then total profits can rise at first, relative to total wages, but later profits may fall both relative to wages, and even absolutely. This can immediately generate a hill-shaped relationship between K_{t+1} and K_t as rising K_t at first elicits rising savings and then eventually depresses them as profits fall.

Similar results can be obtained for a model in which the propensity to save out of profits and wages is the same but where this propensity declines as the society grows progressively richer. (For a formulation in terms of an overlapping generations model in which the discount factor increases with wealth see Benhabib and Day, 1980a.)

Questions have been raised about the possibility of constructing simple chaotic macromodels that are consistent with the presence of long-lived agents who optimize intertemporally and have perfect foresight and in which market clearing occurs. Much of the macroeconomic literature on cycles and chaos uses life-cycle models. It has been suggested that previous models of this type may generate chaos only because they involved agents whose lifespans were less than the duration of many of the cycles, and who were thereby prevented from eliminating the cycles through acts of arbitrage. A number of studies (Benhabib and Nishimura, 1979, 1985, 1989; Boldrin and Montrucchio, 1986; Deneckere and Pelikan, 1986; Woodford, forthcoming) have demonstrated that such models exhibiting cycles and chaos can easily be constructed. (For an earlier discussion, see Stiglitz, 1973.) Discounting, for example, permits relative prices to cycle as long as their percentage change in each period does not exceed the rate of discount. Alternatively, storage costs or imperfect financial markets may prevent complete arbitrage (as in Woodford, forthcoming) and allow cycles or chaos to persist even under rational expectations.

Empirical Evidence on the Presence of Chaos

The evidence on whether chaos does or does not occur in economic phenomena so far is only suggestive.

Brock (1986) has used some new techniques (see also Brock and Dechert, 1986), to test whether a particular time series is most likely to have been generated by a stochastic system or instead by a regime that is (predominantly) chaotic, defined as a deterministic system giving rise to complicated dynamics (perhaps with minor random influences). Brock and Sayers (1985) have used these techniques to study a number of macroeconomic series. While the evidence is weak and somewhat inconclusive, there seem to be grounds for the tentative conclusion that the use in econometric analysis of simple linear systems with stochastic disturbances may in some particular cases be inadequate and misleading and that nonlinear systems may be more appropriate.

On the other hand, even though it has so far been used mostly in macroeconomics, macro variables may not be the most promising place to look for chaos. Rather, from its logic, one may suspect that chaotic dynamics is more likely to affect disaggregated variables (such as the production of pig iron) rather than an aggregate series such as GNP, particularly when the micro variables are inherently subject to resource constraints that interconnect future values of the variables with their current levels (as in the case of resource depletion). All in all, the evidence for the existence of chaotic behavior in real economic time series is far from compelling so far, though what there is does suggest the value of further research in that direction.

How does one test for chaos empirically? One rather simplistic approach proceeds by seeking the underlying dynamic system generating a possibly "chaotic" time series. To dramatize its simplicity, we use the data for the highly "disorderly" time path in Figure 6b, to reconstruct the hill-shaped graph of our generating equation $y_{t+1} = 3.94\ y_t(1 - y_t)$. Assume we have guessed that the underlying relationship is a single difference equation of first order. This enables us to proceed directly from pairs of adjacent observations of y in Figure 6b and merely plot each y_{t+1} against the corresponding y_t. The result is shown back in Figure 2, which indicates that despite the complicated pattern of the data, this yields a virtually perfect reconstruction of the underlying phase curve (the circled points on the highest parabola in the figure).

This procedure works, of course, because there is absolutely nothing random in the process. Each generated point in the time path slavishly follows the dictates of the underlying model and so must correspond to a point on the graph of the equation. This immediately suggests a naive test to determine whether a time series involves random or chaotic influences; if such a reconstruction of the underlying model yields a highly regular relationship, the time path can be presumed chaotic, not random.

However, this calculation has several pitfalls. First, the underlying system may have many variables and/or a complicated lag structure with an unknown number of periods. This underlying mechanism is no longer so "simple," and the kind of structure to look for is never obvious. Yet, this problem is no different from that in choosing the structure of a model for econometric estimation. Second, the available observations may not provide information on the variables of the underlying system but on some function of those variables. Thus, if the system is a difference equation in n variables (vector x_t), then we may only be in a position to observe a function of x_t, $y_t = h(x_t)$ where y_t is a single variable, perhaps an aggregated time series like GNP. The system, therefore, cannot be reduced to a single difference equation since each y_t is compatible with many values of the vector x_t.

Given the preceding considerations, the problem of distinguishing essentially deterministic dynamics from dynamics primarily governed by stochastic elements becomes difficult if not ambiguous. Sophisticated methods to deal with such problems have been sought. These include means to determine whether a given time series is generated by a *stable* and stochastic system or by one which is chaotic but deterministic. Here, the "dimension" of the set of points toward which the time path tends in the limit has proved a helpful criterion. To understand this, note that in the stable stochastic case the state of the system at a future date is a random variable whose future limit may be describable beforehand only by a frequency distribution. Consequently, in the random case only a continuum may be sufficient to contain all possible limit points of the time path. In contrast, a trajectory of a deterministic dynamical system may, as some of the preceding illustrations show, converge to a finite number of points (say, a stationary point or a cycle). Or, instead it may follow a chaotic path or converge to a chaotic set (a "strange attractor"). In the last two cases, the trajectory, while not constituted by a finite number of points, can nevertheless be distinguished from a continuum because, roughly speaking, the former contains "fewer points." (This will be explained presently in somewhat greater detail.)

Methods that seek to distinguish empirically whether the underlying mechanism is deterministic or not on the basis of finite (but large) and possibly aggregated sets of time series data are based on this distinction between the "dimension" of a "strange attractor" and the "dimension" of a stationary distribution generated by a stable stochastic dynamical system. The "dimension" of the latter, suitably defined, can be shown to be infinite. Several definitions of "dimension," which are appropriate for use in testing for a finite but large data set, have been provided by Takens (1985), Proccacia and Grassberger (1985) and others. For a description and some applications of these methods see Brock (1986), Brock and Dechert (1986), Scheinkman and LeBaron (1986), Brock and Sayers (1985), and Brock, Dechert and Scheinkman (1986).

The logic of the dimension approach to empirical testing of whether a time path is chaotic or random can be suggested with the aid of Figure 2. We have just seen how plotting of successive values of y_{t+1} against the corresponding y_t from the chaotic time path in Figure 6b gives us a series of points on the phase curve of Figure 2. Now, even if in the limit these points were to fill in the entire phase curve, they would still only form a one dimensional set—a curve in two dimensional space.

In contrast, had the time path been subjected to random influences, the exercise just carried out would obviously have yielded a set of points scattered about the parabola: at best, an area that can be covered by a continuous two-dimensional region. This suggests why chaotic behavior is associated with a set of points lower in dimension than is randomness, and indicates how dimension can, in principle, be used to distinguish one case from the other.

This leaves the difficult problem of distinguishing empirically any mixed case in which chaotic and random influences are *both* present. The deterministic structure of y_t may involve a time path with a limit of low dimension (possibly corresponding to a "strange attractor") and the deterministic part can be large relative to the magnitude of the variation in the series of independently distributed random shocks. The possibility of a "strange attractor" arising out of the deterministic part of the system, with some noise superimposed upon it, has been called "noisy chaos." Methods to identify cases where the noise component of the time path is "small" relative to the deterministic part are given by Ben-Mizrachi (1984) and also Brock and Dechert (1986).

These approaches are complicated by the fact that the data sets used in reality are necessarily finite. A linear *stochastic* difference equation system can appear to generate a "finite dimensional" attractor if its stochastic component is small enough, and can therefore suggest the conclusion that the underlying dynamic system is strictly deterministic. (For a discussion of the difficulties stemming from finiteness of the data samples, see Ramsey and Yun, 1987.)

A promising additional check tests whether the underlying system is, on the average, "stable." Such a test seeks to ascertain whether trajectories generated by a given relationship, but with different initial conditions, that are initially close together, remain close, as they would not do if they were chaotic. Methods have been designed to estimate the mean rate of divergence of such trajectories. A positive divergence rate is taken as evidence of the presence of a strange attractor rather than a stable but stochastic dynamical system, whose stability prevents marked divergence of the trajectories of its variables.

At present, methods to distinguish whether a time series has been generated by a stable linear stochastic system or a deterministic nonlinear system giving rise to chaotic dynamics (possibly also containing a negligible stochastic component) are still very new. Whether these methods can be developed further to overcome problems that arise with "small" data sets remains to be seen.

Though necessarily incomplete, this discussion should offer the reader some impression of the methods now being used in empirical studies of chaotic phenomena in economics.

■ *The authors are grateful to the C. V. Starr Center for Applied Economics for its assistance in preparation of this paper. We are also deeply indebted to William Brock, Dermot Gately, Alannah Orrison, Richard Quandt, Carl Shapiro and Joseph Stiglitz for their very substantial help, and to Vacharee Devakula for her skillful contribution to the graphs.*

References

▶**Akerlof, C. A. and J. Stiglitz,** "Capital, Wages and Structural Unemployment," *The Economic Journal*, 1969, *79*, 769–281.

▶**Baumol, W.J.,** "Speculation, Profitability, and Stability," *Review of Economics and Statistics*, August 1957, *39*, 263–71.

▶**Baumol, W.J.,** "Pitfalls on Contracyclical Policies: Some Tools and Results," *Review of Economics and Statistics*, *Feb*. 1961, *43*, 21–6.

Baumol, W. J., "Unpredictability, Pseudorandomness and Military-Civilian Budget Interactions," *Revista Internazionale di Scienze Economiche e Commerciali*, April 1986, *XXXIII:4*, 297–318.

▶**Baumol, W. J. and E. N. Wolff,** "Feedback from Productivity Growth to R & D," *Scandinavian Journal of Economics*, 1983, *85:2*, 147–57.

Baumol, W. J. and R. E. Quandt, "Chaos Models and their Implications for Forecasting," *Eastern Economic Journal*, Jan.–Mar. 1985, *11*, 3–15.

▶**Benhabib, J.,** "Adaptive Monetary Policy and Rational Expectations," *Journal of Economic Theory*, 1980, *23*, 261–266.

▶**Benhabib, J. and R. H. Day,** "Erratic Accumulation," *Economics Letters*, 1980a, *6*, 113–117.

▶**Benhabib, J. and R. H. Day,** "Rational Choice and Erratic Behavior," *Review of Economic Studies*, 1981, *48*, 459–472.

▶**Benhabib, J. and R. H. Day,** "A Characterization of Erratic Dynamics in the Overlapping Generations Model," *Journal of Economic Dynamics and Control*, 1982, *4*, 37–55.

▶**Benhabib, J. and K. Nishimura,** "The Hopf Bifurcation and the Existence and Stability of Closed Orbits in Multisector Models of Optimum

Economic Growth," *Journal of Economic Theory*, 1979, *21*, 421–444.

Benhabib, J. and K. Nishimura, "Competitive Equilibrium Cycles," *Journal of Economic Theory*, 1985, *35*, 284–306.

Benhabib, J. and K. Nishimura, "Stochastic Equilibrium Oscillations," *International Economic Review*, forthcoming.

Ben-Mizrachi, A., et al., "Characterization of Experimental (noisy) Strange Attractors," *Physical Review A*, 1984, *29:3*, 975–977.

Boldrin, M. and L. Montrucchio, "The Emergence of Dynamic Complexities in Models of Optimal Growth: The Role of Impatience," *Journal of Economic Theory*, 1986, *40*, 26–39.

Brock, W. A., "Distinguishing Random and Deterministic System: Abridged Version," *Journal of Economic Theory*, 1986, *40*, 168–195.

Brock, W. A. and W. D. Dechert, "Theorems on Distinguishing Deterministic from Random Systems," preliminary draft, 1986.

Brock, W. A. and C. L. Sayers, "Is the Business Cycle Characterized by Deterministic Chaos?" Department of Economics, University of Wisconsin at Madison, September 1985.

Brock, W. A., W. D. Dechert and J. A. Scheinkman, "A Test for Independence Based on Correlation Dimension," University of Wisconsin at Madison, 1986, mimeo.

Collet, P. and J. Eckmann, *Iterated Maps on the Interval as Dynamical Systems*. Basel: Birkhauser, 1980.

Crutchfield, James, et al., "Geometry From a Time Series," *Physical Review Letters*, 1980, *47*, 712.

Dana, R. A. and P. Malgrange, "The Dynamics of a Discrete Version of a Growth Cycle Model." In Ancot, J. P. (ed.), *Analyzing the Structure of Econometric Models*. Amsterdam: M. Nijhoff, 1984.

Day, R. H., "Irregular Growth Cycles," *American Economic Review*, 1982, *72*, 406–414.

Day, R. H., "The Emergence of Chaos from Classical Economic Growth," *Quarterly Journal of Economics*, 1983, *98*, 201–213.

Day, R. H. and W. J. Shafer, "Keynesian Chaos," Department of Economics, University of Southern California, 1983.

Deneckere, R. and S. Pelikan, "Competitive Chaos," *Journal of Economic Theory*, 1986, *40*, 13–25.

Diamond, P., "Chaotic Behavior of Systems of Difference Equations," *International Journal of Systems Science*, 1976, *7*, 953–956.

Frisch, Ragnar, "Propagation Problems and Impulse Problems in Dynamic Economics," *Economic Essays in Honor of Gustav Cassel, London*, 1933.

Gale, D., "Pure Exchange Equilibrium of Dynamic Economic Model," *Journal of Economic Theory*, 1973, *6*, 12–36.

Grandmont, J. M., "On Endogenous Competitive Business Cycles," *Econometrica*, September 1985, *5*, 995–1045.

Grandmont, Jean Michel, "Periodic and Aperiodic Behavior in Discrete, One-dimensional, Dynamical Systems." In Hildenbrand, W. and A Mas-Collel, eds, *Contributions to Mathematical Economics*. New York: North Holland, 1986.

Grassberger, P. and I. Procaccia, "Measuring the Strangeness of Strange Attractors," *Physica*, 1983, *9D*, 189–208.

Guckenheimer, J. and P. Holmes, Nonlinear Oscillations, Dynamical Systems and Bifurcation of Vector Fields. *New York*: *Springer-Verlag*, 1983.

Jacobson, M., "Absolutely Continuous Invariant Measures for One-Parameter Families of One-Dimensional Maps," MDULComm. Math. Phys.MDNM, 81, 1981, 39–88.

Jensen, R. U. and Urban, R., "Chaotic Price Behavior in a Nonlinear Cobweb Model," mimeo, Yale University, 1982.

Li, T. Y. and J. A. Yorke, "Period Three Implies Chaos," *American Mathematical Monthly*, 1975, *82*, 985–992.

Llibre, J., "Sufficient Conditions for a Continuous Map of R^n into Itself to Have m-periodic Points for all $m > 0$." In Atteia, M., D. Dancel, and I. Cernowiski (eds.), *Nonlinear Problems of Analysis in Geometry and Mechanics*, New York: Pitman, 1981.

Lundberk, Erik, *Studies in the Theory of Economic Expansion*, London: P. S. King & Son Ltd., 1937.

Marotto, F. R., "Snapback Repellers Imply Chaos in Rn," *Journal of Mathematical Analysis and Applications*, 1978, *63*, 199–223.

May, R. M., "Simple Mathematical Models with very Complicated Dynamics," *Nature*, 1976, *261*, 459–467.

May, Robert M. and John, R. Beddington, "Nonlinear Difference Equations: Stable Points, Stable Cycles, Chaos," unpublished manuscript, 1975.

Phillips, A. W., "Stabilization Policy in a Closed Economy," *Economic Journal*, 1954, *64*, 290–323.

Phillips, A. W., "Stabilization Policy and the Time-Form of Lagged Responses," *Economic Journal*, 1957, *67*, 265–277.

Poincaré, Henri, "Sur les Courbes Définies par les Equations Différentielles," *C. R. Acad. Sci. Paris*, 1880, *90*, 673–75.

Poincaré, Henri, *Les Méthodes Nouvelles de la Mécanique Celecte I*, *II*, *III*. Paris: Gauthier-Villars, 1892; New York, Dover, 1957.

Ramsey, J. and Hsiao-Jane Yun, "The Statistical Properties of Dimension Calculations Using Small Data Sets," mimeo, New York University 1987.

Rand, D., "Exotic Phenomena in Games and Duopoly Models," *Journal of Mathematical Economics*, 1978, *5*, 173–184.

Samuelson, P. A., "Interactions Between the Multiplier Analysis and the Principle of Acceleration," *Review of Economics and Statistics*, May 1939, *21*, 75–78.

Sarkovskii, A. N., "Coexistence of Cycles of a Continuous Map of a Line into Itself," *Ukranichkii Matematicheskii Zhurnal*, 1964, *16*, 61–71.

Scheinkman, J. and B. Le Baron, "Nonlinear Dynamics and Stock Returns," Department of Economics, University of Chicago, 1986.

Shaw, R., "Strange Attractors, Chaotic Behavior, and Information Flow," *Zhurnal Naturforsch*, 1981, *36A*, 80.

Smale, Steven, "Differentiable Dynamical Systems," *The Bulletin of the American Mathematical Society*, 1967, *74*, 747–817.

Stefan, P., "A Theorem of Sarkovskii on the Existence of Periodic Orbits of Continuous Endomorphism of the Real Line," *Communications in Mathematical Physics*, 1977, *54*, 237–248.

Stiglitz, J., "Recurrence of Techniques in a Dynamic Economy." In Mirrlees, J. A. and N. H. Stern (eds.), *Models of Economic Growth*. London: MacMillan Press, 1973, 138–162.

Stutzer, M., "Chaotic Dynamics and Bifurcations in a Macro Model," *Journal of Economic Dynamics and Control*, 1980, *2*, 353–376.

Takens, F., "Distinguishing Deterministic and Random Systems." In Rosenblatt, G. G. Iooss, and D. Joseph (eds.), *Nonlinear Dynamics and Turbulence*. Boston: Pitman, 1985, 315–333.

Woodford, M., "Imperfect Financial Intermediator and Complex Dynamics." In Barnett, W., J. Geweke and K. Shell, *Chaos, Sunspots, Bubbles and Nonlinearity*. New York: Cambridge University Press, forthcoming.