Self-Organization and Dissipative Structures

Applications in the Physical and Social Sciences

Edited by William C. Schieve and Peter M. Allen

University of Texas Press, Austin, 1982

4. The Many Faces of Scaling: Fractals, Geometry of Nature, and Economics

Benoit B. Mandelbrot

This paper brings together diverse topics from several areas of mathematics and science. These areas are reputed to be totally unrelated to each other, yet my investigations show them to be akin in one way at least: all of them gain greatly from the application of the same very versatile device or principle, scaling. The links that result from this commonality of mathematical structure are striking and surprising.

Science became acquainted with scaling through Richardson's picture of turbulence, but I believe I was the first to apply this idea elsewhere, for example in linguistics (1951 on), economics (1959 on), noise theory (1963 on), neurophysiology (1964 on), hydrology (1965 on). Starting with investigations of geomorphology (1967 on) and turbulence (1967 on), my work has become increasingly geometric. The scaling methods in the analytic physics of critical collective phenomena (Kadanoff's work of 1967 and renormalization groups) arose quite independently but now find numerous uses for my scaling geometry. Being expository, this paper freely incorporates paragraphs from my book, *Fractals: Form, Chance, and Dimension*, and my other publications. See also my more recent book, *The Fractal Geometry of Nature*.

GEOMETRY OF THE IRREGULAR

The best is to begin with comments on the bottom half of the combined figs. 4.1 and 4.2. My point is that fig. 4.1 does not represent what I hope you think it might represent; with due apologies to Mr. Baedeker and his heirs and competitors, this is not a landscape on the earth, the moon, or any other planet, but an example of fractal surface contrived and computer generated deliberately to mimic a landscape. The program, the product of Richard F. Voss, implements the mathematical model I propose for the Earth's relief.

Benoit B. Mandelbrot is with the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.

Figs. 4.1 and 4.2. Graphics courtesy of Benoit B. Mandelbrot © 1977 from the back jacket of his book FRACTALS: FORM, CHANCE, AND DIMENSION, W. H. Freeman and Company, San Francisco, 1977.

The two significant parameters are a real number D lying between 2 and 3 and the seed of a pseudorandom subroutine. This D is the *fractal dimension* of the surface. In fig. 4.1, in order to ensure the resemblance that I hope you have perceived, the suitable value had to be $D = 2.2500$. Had we picked a different D, the resulting surface would have been different in form.

What is really meant by *form*? The question is vital in many sciences, but it is far from having been answered satisfactorily. If we were to trust our mathematical friends blindly, we would interpret *form* as that which is studied by topology. Such would have been the case if topology were actually devoted to the task suggested by the etymology of its name. Unfortunately, the contents of this branch of mathematics is such that in the present case we would be sorely disappointed. Indeed, at least in principle, the surface in fig. 4.1 can be obtained from a square without a tear, using a one-to-one continuous transformation, and this property *defines* it as being topologically a square. The surfaces corresponding to the same algorithm but different values of D are also squares from the topological viewpoint; nevertheless their form depends greatly upon the value of D. When D is close to 2, they are smoother in detail and less flat overall than when D is close to 3 as seen on pages 210 to 215 in my book, *Fractals: Form, Chance, and Dimension*. (The back of the book's jacket reproduces the present combined figs. 4.1 and 4.2.)

Fig. 4.2 offers a further example of the limitations of topology. Again, its point is that it does not represent what I hope you think it might represent: it is not a line version of the photograph of a planet, but the projection (upon a tangential plane) of the surface of a sphere on which a curve was drawn by a computer, again the work of Richard Voss using an algorithm I supplied. Topologically, this curve can be obtained by deforming a collection of a few circles by a one-to-one continuous transformation. Therefore this drawing is topologically a collection of circles, but this feature accounts for only a small part of the whole truth. There ought to exist some mathematical way of expressing the obvious: between a collection of circles and the curve you see here there is, again, a profound difference in form. Furthermore, a second and more careful look at fig. 4.2 leads you to conclude that it is not quite Earth-like. The curves you see are markedly too wiggly, too complicated to represent the contours of Earth's continents.

The fact that very irregular shapes are often encountered in Nature requires little elaboration. They never tire of exciting the layman's imagination, but science has failed to tackle them. Does it follow from the inappropriateness of topology that the degree of wiggliness must remain an intuitive notion, that is, a notion inaccessible to mathematical description? The answer is a resounding no. It turns out that mathematicians had long ago taken steps in a direction I later saw pointed towards a geometry of irregularity. For example, the degree of local irregularity of a surface or of any other set is measured by its Hausdorff-Besicovitch dimension D. This notion, however, is only rarely called upon in mathematics, and it failed completely to draw the attention of scientists. Applications in the sciences, and the very idea that (in the case of certain geometric shapes) D can also be used to quantify an aspect of form, did not come out until my papers that eventually led to my book *Fractals*.

Originally, the above D was geared towards the study of sets that had been designed as standards of irregularity, most notably by George Cantor and Giuseppe Peano. Some among you may be aware of the reputations of Cantor and Peano, of being unsurpassed champions of mathematics as a form of art for art's sake. If you know them in this light, you must be surprised at hearing their names in the context of natural science, so let me elaborate. The great mathematicians active during the period 1875–1922, which witnessed a deep crisis in pure mathematics, introduced many sets intended solely to prove that, in comparison to the concepts of the old mathematics, those of modern mathematics were of increased generality. Such sets, however, were never meant to be applied in science. In fact, both the creators and their followers were unanimous in considering them as "pathological" and as "monsters." I claim that they were mistaken, that these specific sets can also be given another and very different sort of use, as models of the irregularity and fragmentation in nature. Far from being pathological, they possess features that turn out to be by far closer to certain aspects of nature than the geometry on which all of us have been brought up, that of Euclid. Their usefulness is linked to the fact that they have something deep in common with Brownian motion. Together, these various sets are examples of a class of sets that I have proposed to call *fractals*, and they are the topic of the book to which I have already alluded.

To summarize this book here would be foolish, but I would like to take this opportunity to sketch some leading ideas, and to point out the profound conceptual link that exists between my current work, which is devoted to the mathematics of fractals and to the geometry of nature, and my earlier work on the temporal behavior of commodity and security prices. The link between my topics is the fact that all three rely in an essential fashion upon the notion of self similarity and on other forms of scaling.

THE SCALING PRINCIPLES OF MATHEMATICAL AND NATURAL GEOMETRY

The largest and most complicated of the five diagrams that make up fig. 4.3 is the composite of two wondrous and many-sided broken lines, each devoid of self-contact. They are drawn by a computer instructed by Sigmund W. Han-

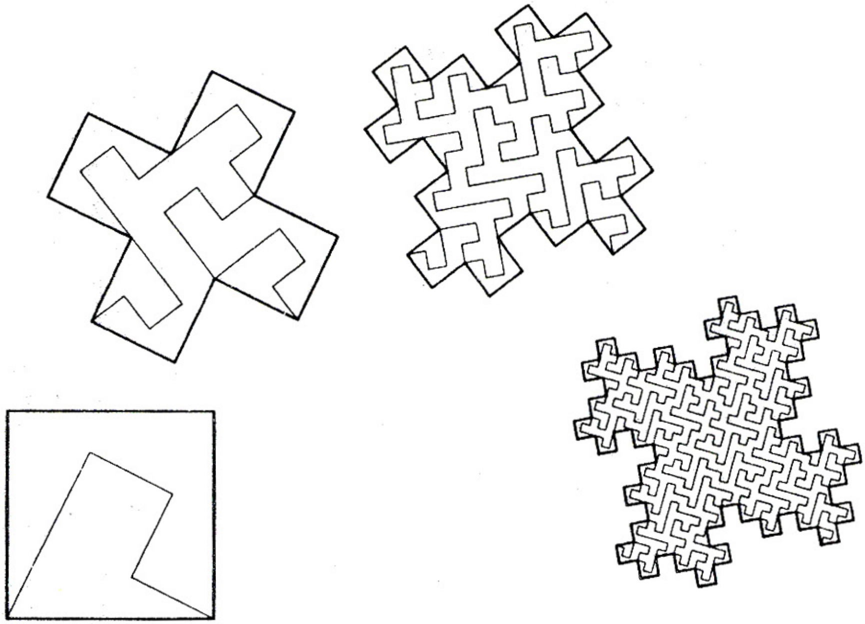

Fig. 4.3

delman. I am tempted to call them *teragons*, to take advantage of the original basic meanings of the Greek word *teras*, *teratos*, "a wonder or a monster," and of the use of *tera* in the metric system as the prefix to designate the very large number 10^{12}. One of these broken lines is so violently folded upon itself as to give the impression that it attempts a monstrous task for a curve: to fill the interior of the second teragon. This impression was intended and is quite justified, since indeed the teragon in question is an advanced stage of the construction of a space-filling curve. By way of contrast, the second curve can be called a wrapping.

The first space-filling curve was discovered by Peano in 1890 and many others followed in the next twenty-five years.

The present new illustration is inspired by Helge von Koch, who had the pioneering idea (first implemented in his snowflake curve) to seek curves with the property of being precisely as complicated in the small as in the large. His motivation was purely mathematical: he was seeking curves without tangent, that is, such that the direction of a cord joining two points has no limit as these points converge to each other. To achieve this goal, the simplest was to demand that this cord fluctuate exactly as much in the small as in the large.

It is easy to restate this goal from mathematical into physical terms. Koch wanted the fine details seen under the microscope, whatever the magnification, to be the same (scale aside) as the gross features seen by the naked eye. Using a different vocabulary, he wanted the fine details seen on a very precise map to be the same as the gross features seen on a rough map. Some such concrete interpretation of Koch's procedure has often inspired me in empirical work. Hence the following guiding principles:

A) SCALING PRINCIPLE OF NATURAL GEOMETRY. *To assume that small and large features are identical except for scale is often a useful approximation in science.*

B) SCALING PRINCIPLE OF MATHEMATICAL GEOMETRY. *To limit oneself to sets wherein small and large features are identical except for scale is often a convenient procedure in geometry.*

Part of my work consists in viewing B as having provided a collection of answers without questions and in setting them to work on the questions without answers summarized under A.

The only fairly wide justification for A is that any sum of many effects satisfying a "central limit theorem" is scaling. This statement is, of course, too loose to be provable, yet a prudent addition of natural assumptions makes it into provable theorems or plausible conjectures. But this central limit argument is seldom persuasive by itself.

EXAMPLES OF FRACTAL SHAPES

To implement the goal we have stated in his way and in mine, Koch proceeds step by step. Select an initiator set and a generator set, the former often an interval or an open or closed polygon. The first construction stage replaces each side of the initiator by an appropriately rescaled and displaced version of the generator. Then a second stage repeats the same construction with the polygon obtained at the first stage, and so on.

The early stages of the constructions shown on fig. 4.3 are illustrated by the four small diagrams of this figure, to be followed clockwise from left center, in order of increasing complication. The initiators are a unit square for the wrapping, and a side of this square for the filling. The generator of the filling is an irregular open equal-sided pentagon. This pentagon does it best to fill the square. (Indeed, one perceives an underlying square lattice of lines $1/\sqrt{5}$ apart, and our original filling passes through every lattice vertex contained in the original wrapping.) In the next stage of the construction, each side of the pentagon is replaced by an image of its whole reduced in the ratio of $1/\sqrt{5}$. The result no longer fits within the square, but it fills uniformly the crosslike shape obtained by replacing each side of the square by the wrapping generator, which has $N=3$ sides of length $r=1/\sqrt{5}$. The same two constructions are then repeated ad infinitum in parallel. A designer who zooms in as the construction proceeds will constantly witness the same density of filling, but one who stays put sees a curve that fills increasingly uniformly a varying wrapping whose complexity keeps increasing.

The Peano curves which mathematicians designed during the heroic period up to 1922 all filled a square or a triangle, but the present construct—like the dragon curve of Heightway and Harter and a curve due to Gosper—involves more imaginative boundaries.

Fig. 4.4 carries the construction of the two curves of fig. 4.3 one step further and presents the result in a different light. The filling is now interpreted as the cumulative shoreline of several juxtaposed river networks, and the wrapping as the combination of a drainage divide surrounding these networks and of a portion of seashore. (To build up the network, one proceeds step by step: (1) Each dead-end square in the basic underlying lattice—defined as such that three sides belong to the filling teragon—is replaced by a short stream from its center beyond the open side. After that the dead-ends

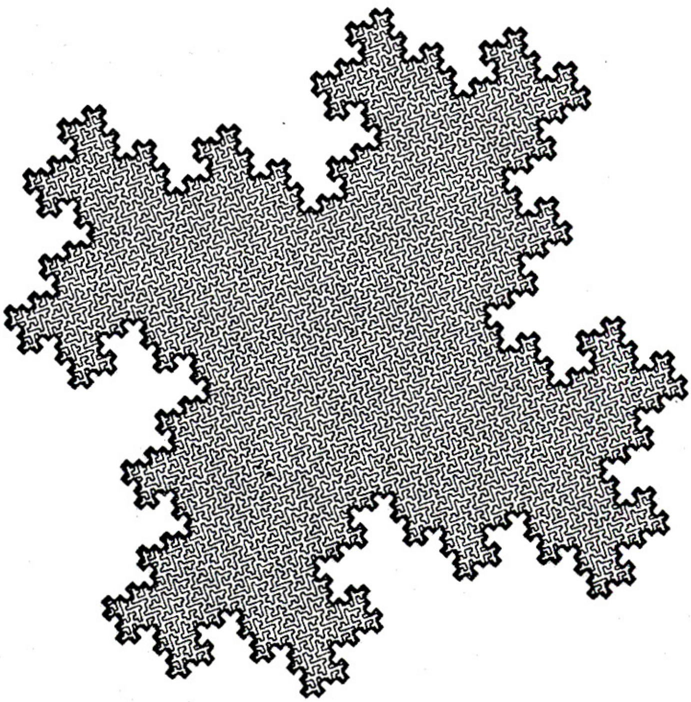

Fig. 4.4

are cut off. (2) Then one proceeds in the same fashion with the straightened teragon. (3) And so on until the filling is exhausted.) In this light, the wrapping becomes reinterpreted as the juxtaposed network's external drainage divide.

To use an old sophomoric line, after you think of it imaginatively, carefully, and at great length, it becomes quite obvious that a river network's shore gives an idea of the structure of a plane-filling curve. The converse is also true, except of course that the present network is much too squarish to be realistic. Much better-looking ones are given in my book, but the basic idea is present here. You may say that the present network resembles a bonsai tree.

To sum up, the mathematicians who tell us that Peano curves are totally nonintuitive are mistaken. To be charitable, let us students of nature accuse them of having tried to prevent us from knowing of a beautiful new tool and applaud their skill in succeeding in holding to it exclusively for so long.

THE NOTION OF FRACTAL DIMENSION

It is easy to see that each stage of a Koch construction multiplies a polygon's length by a fixed factor $Nr>1$; hence the limit curves obtained by pursuing the constructions of fig. 4.3 ad infinitum are of infinite length. Furthermore, it is tempting to say that the filling is "much more infinite" than its wrapping, because its length tends to infinity more rapidly. This intuitive feeling is expressed mathematically by the notion of fractal dimension, to which we have alluded repeatedly; the notion was formed by Hausdorff and was perfected by Besicovitch. The explanation of the underlying idea begins with the very simplest shapes: line segments, rectangles in the plane, and the like.

Because a straight line's Euclidean dimension is 1, it follows for every integer γ that the "whole" made up of the segment of straight line $0 \leq x < X$ may be "paved over" (each point being covered once and only once) by $N=\gamma$ "parts." These parts are segments of the form $(k-1)X/\gamma \leq x < kX/\gamma$, where k goes from 1 to γ. Each part can be deduced from the whole by a similarity of ratio $r(N)=1/N$. Likewise, because a plane's Euclidian dimension is 2, it follows that, whatever the value of γ, the "whole" made up of a rectangle $0 \leq x < X$; $0 \leq \gamma < Y$ can be "paved over" exactly by $N=\gamma^2$ parts. These parts are rectangles defined by $(k-1)X/\gamma \leq x < kX/\gamma$ and $(h-1)X/\gamma \leq y < hY/\gamma$, wherein k and h go from 1 to γ. Each part can now be deduced from the whole by a similarity of ratio $r(N)=1/\gamma=1/N^{1/2}$. Finally, in spaces whose Euclidean dimension is $E>3$, a D-dimensional parallelepiped can be defined for any $D \leq E$, and we see that in all the classical cases, the dimension satisfies

$$D=-\log N/\log r(N)=\log N/\log (1/r) .$$

Now observe that the exponent of self-similarity continues to have formal meaning for some shapes which are neither a segment nor a square. The main requirement is scaling: the whole may be split up into N parts deducible from it by self-similarity having the ratio r (followed by displacement or by symmetry). Such is precisely the case with the limits of our teragons in fig. 4.3. For the wrapping, we see that $N=3$ and $r=1/\surd 5$, hence

$$D=\log 3/\log\surd 5=\log 9/\log 5=1.3652$$

For the filling, we see that $N=5$ and $r=1/\surd 5$, hence

$$D=\log 5/\log\surd 5=2\ .$$

Thus, the impression that the filling is "more infinite" than its wrapping is confirmed and quantified by the inequality between their dimensions. The impression that the filling really fills a plane domain is confirmed and quantified by its dimension being $D=2$.

The preceding argument may seem overly specialized, so it may be comforting to know (a) that fractal dimension can be defined using alternative methods of greater generality and full rigor and (b) that the result behaves like the old-fashioned one in many other ways. For example, consider the notion of *measure*. If a set is self-similar and measure is taken properly, then the portion of this set that is contained in a sphere of radius R is of measure R^D.

This leads us to a mathematical definition I gave: a *fractal set* is defined as being a set such that the above D—or the dimension yielded by the more general and intrinsic methods of Hausdorff and Besicovitch—is greater than the topological dimension—or the intuitive dimension. The wrapping in fig. 4.3 is a curve of topological dimension 1; hence it is a fractal curve. The prototypical stochastic fractal is Brownian motion in a plane or a higher space: indeed it is topologically a curve of dimension 1, but it is fractally of dimension 2. Its coordinate functions, such as $X(t)$, are curves of topological dimension 1 and fractal dimension 3/2. Paul Levy has generalized the Brownian function to functions from a multidimensional "time," say (x,y), to a scalar Z. If Z is interpreted as an altitude, such a function represents a landscape; it is of fractal dimension $D=5/2$. Finally, one can preserve scaling while changing its specific rule (that is, its D) by performing a certain operation (which no longer deserves to be viewed as exotic) called Riemann-Liouville fractional integration. If one starts with a Brownian $Z(x,y)$ and the order of integration is 1/4, one obtains the surface shown in fig. 4.1, which in technical terms is described as a fractional Brownian surface of dimension $D=2.2500$. It brings us back to our point of departure, and we can proceed to other tasks.

ON TO ECONOMICS

To sum up, by showing that the kin of the fabled Cantor and Peano monsters are indispensable in modeling the geometry of nature, I hope to have triggered an unexpected encounter between the mathematicians enamored of art for art's sake and those who celebrate nature by trying to imitate it. The reunion was soon joined by a different and unexpected crowd, a large one. While I was identifying scaling structures in the fractal geometry of nature, physicists concerned with a variety of collective phenomena were independently identifying scaling structures in analytic physics, that is, the mainstream physics which expresses its results in formulas rather than in shapes. We soon met, and scaling geometry was promptly injected into the study of collective phenomena.

Where does this leave economics? All too many economic models merely translate into a different language some of the mathematical material of physics, but there are at least two exceptions, one old and one recent.

The old exception hinges on the historical fact that the first scholar to describe the Brownian motion process of independent Gaussian increments was not a physicist. He was Louis Bachelier, whose 1900 Ph.D. was granted (reluctantly) by a committee of mathematicians, but who viewed himself primarily as a student of economic risk. The economic modelers of the 1900's remained unaware of Bachelier's existence, and their heirs in the 1960's rediscovered the Brownian model in books of physics; but the history of ideas will remember that physicists were second to this front, five unquestionable years after a mathematical economist.

Around 1960 we witnessed a second instance of a principle of model making being used in physics several unquestionable years after it had been used in economics. To further the resemblance with the Bachelier episode, this principle did not originate with a professional economist but with the author of these lines—and (as has already been said) the physicists reinvented it when needed, without my help. A distinct difference from the Bachelier episode is that my work did not remain isolated, but merged into my later work on fractals and the geometry of nature. Hence, when the inevitable comes and an economist decides to look at collective phenomena of physics for inspiration in economic modeling, he will find that part of the work has already been done without reference to physics.

One often hears that economic models tend to involve so many parameters that an exacting empirical confirmation is impossible even in principle. In fact, the data that concern commodity and security markets are so plentiful one can drown in them. Bachelier had the brilliant idea that the first task for the student of economic risk is to examine phenomena for which data could readily be collected.

Thus, when Bachelier's day finally came in economics (sixty years after his Ph.D. and fifteen after his death), one could actually test whether he was right or wrong. Many a putative model is such (in structure or number of parameters) that no amount of evidence could prove it to be wrong, and Karl Popper tells us that such nonfalsifiable constructs are not acceptable as scientific models.

Sad to report, Bachelier's Brownian motion model is contradicted by the evidence. On the other hand, a more general idea of Bachelier's, called "efficient market hypothesis," proves remarkably fruitful; in particular, it implies that no mechanical trading procedure can be a winning one.

THE DISCONTINUITY OF PRICES

The simplest anti-Bachelier argument is based on the following experimental observation, which is unsophisticated but fundamental. Brownian motion's sample functions are continuous. Prices on competitive markets, however, are not continuous. The principal (or perhaps only) reasons for assuming continuity are that it was a marvellous success in mechanics and that diverse exogenous quantities and rates that enter into economics but are defined in purely physical terms must necessarily be reasonably close to being continuous. Prices, however, are different. The typical mechanism of price formation involves both knowledge of the present and anticipation of the future. Even in the cases where the exogenous physical components of a price are constrained to vary continuously, anticipations can and often do change drastically in a flash. A physical signal of negligible energy and duration, "the stroke of a pen," may provoke a brutal change of anticipations. When this happens without institutional constraint and inertia to complicate matters, a price determined on the basis of anticipation can crash to zero or soar out of sight; it can do anything.

Can this observation be of any predictive value? To show that it is, let me retell the history of the rise, fall, and burial of a briefly famous method of trading using "filters" (Alexander 1961). A p percent filter may be defined as a device that monitors price continuously, records all the local maxima and minima, gives a buy signal when price first reaches a local minimum plus exactly p percent, and gives a sell signal when a price first reaches a local maximum minus exactly p percent. Actually, Alexander monitored the sequence of daily highs and lows of the index. He assumed that if an earlier day's low plus p percent was exceeded for the first time by the high on day d, then at some time during the day d, the instantaneous price was exactly equal to said low plus p percent—thus provoking a buy signal. The process was similar for sell signals. The assumption that a price record can be handled like a continuous function seemed to Alexander too obvious even to be mentioned. The empirical conclusion is that to follow a filter's buy or sell signal is very much more beneficial than to "buy and hold."

I found (Mandelbrot 1963, p. 417) that there is a gap in this argument. Indeed, even on days when price variation seems reasonably continuous, its continuity is the result of deliberate action by a market specialist performing his assigned function of matching buyers and sellers and of ensuring the continuity of the market by buying or selling from his holdings. The specialist creates bargains reserved to friends, while most customers have to buy at the next higher price. In either case, the advantages of filter trading, as computed according to Alexander, are overstated. Since my theoretical and experimental studies (to be described momentarily) make me hold that price is discontinuous except for institutional lags or "inertias" that can have only a limited effect, I predicted that a recheck would show the advantage of filter trading over buy and hold trading to be largely or wholly spurious. Alexander 1964 found my prediction to be correct, and the method of filters to be no better than "buy and hold." I had killed this method. Fama and Blume, 1966, carried out a thorough "post-mortem" check using individual price series; the method of filters is now buried for good.

Winning martingales resemble perpetual motion machines. It is to the credit of Bachelier's efficient market hypothesis that it had predicted well in advance that filters should not work, but to the discredit of Bachelier's Brownian motion model that it could not explain why filters seemed to work. Therefore it is to the credit of my specific models that they permit an analysis and pinpoint the flaws present in diverse paths to sure wealth.

THE SCALING PRINCIPLE OF ECONOMICS

The failure of Brownian motion as a model of price variation elicited two very different responses. On the one hand, there is a plethora of ad hoc statistical "fixes." When faced with a statistical test that rejects the Brownian hypothesis, try one modification after another until the test is fooled. A popular fix is censorship, hypocritically labeled "rejection of outliers." It consists of distinguishing between ordinary small price changes and large price changes that defeat Alexander's filters. This traditional fix is effective in analyzing astronomic observations, but this is not an appropriate precedent. Large astronomic errors are traceable to equipment problems, can be corrected, and are not of intrinsic interest. Large price changes, on the contrary, are at the heart of the hopes or fears of participants in competitive markets. A second popular fix is to postulate that data are a mixture: if X is not Gaussian, maybe it is a mixture of Gaussian variables. Yet another fix is transformation: if X is positive and non-Gaussian, maybe $\log X$ is Gaussian; if X is symmetric and non-Gaussian, maybe $\tan^{-1}X$ will fool the test. Yet another fix consists in proclaiming that price follows a Brownian motion whose parameters vary uncontrollably. By design, a fix can never be falsified, hence (remember Karl Popper's views) it cannot be a scientific model.

At the opposite of the fixes stands my own work. The basic principle was first used to tackle income distributions, and it applies to diverse data of economics, but is best expressed in the context of prices.

SCALING PRINCIPLE OF PRICE CHANGE. *When* X(t) *is a price, log* X(t) *has the property that its increment over an arbitrary time lag* d, *log* X(t+d) − *log* X(t), *has a distribution independent of the lag* d, *except for a scale factor.*

Before exploring this principle's consequences, let us run through a checklist of properties.

A scientific principle must yield predictions that can be checked against the evidence. This one does so, and the fit is very good.

It is important for scientific principles to be reducible to other theoretical considerations in their fields. In this instance, the only explanatory arguments, Mandelbrot 1966, 1971, view it as the consequence of the scaling property of exogenous physical variables. These arguments are much less well established than the result they purport to justify.

Finally, even when no actual explanation is available, it is pleasant if a scientific principle does not actually clash with earlier presuppositions. The present scaling principle seems innocent enough—if only because the question of scaling had not previously been raised, so that contrary opinions could not be expressed. All that scaling seems to say is that in competitive markets no time lag is really more special than any other, that is, the obvious special features of the day and the week (and the year in case of agricultural commodities) are compensated or arbitraged away. While all the usual fixes of the Brownian motion involve privileged time scales, my principle affirms that there is no sufficient reason to assume that any time scale is more privileged than any other.

However, we want the actual implementation of the scaling principle to be distinct from the standard result, which of course is Brownian motion. To achieve this, it is necessary to clash violently with earlier statistical suppositions, by postulating that price changes have an infinite variance. This is a possibility that every other investigation of the topic excludes from the outset. Before my papers, it seemed a perfectly innocuous step to write "denote the variance by V." To demonstrate that infinite variance has significant and desirable consequences was a major step I took around 1960. Later, it inspired my investigations of fractal geometry, where it is essential to allow curves to have infinite length and surfaces to have infinite area, while the usual assumptions combined with scaling bring us a trivial conclusion: back to standard Euclidean lines, planes, and so forth.

THE INFINITE VARIANCE SYNDROME

One must question the finiteness of the population variance of log price because of diverse problems encountered by the use of typical values, which is the least sophisticated level of descriptive statistics but is far from being harmless. In order to summarize tables of frequencies, it is customary to use sample averages to measure location and sample root mean squares to measure dispersion. The former ordinarily raises little question in economics, as sample averages tend to vary little between samples. Mean squares of prices, to the contrary, prove extraordinarily elusive:

(A) Values corresponding to different long subsamples of the same price change series differ to the extent of having different orders of magnitude.

(B) As sample size gradually increases, the mean square fails to stabilize. It goes up and down, with an overall tendency to increase.

(C) The mean square tends to be influenced predominantly by a few of the squares being averaged, sometimes a single one.

These properties suggest that the unknown theoretical second moment is very large. When this moment is finite but huge, sample moments will eventually converge to it, but so slowly that the limit matters very little in practice. The alternative possibility that I put forward is that the theoretical mean square is infinite. Between very large and infinite, there is no difference one could detect through available sample moments. Also, of course, the fact that a variable X has an infinite variance in no way denies that X is finite with a probability equal to 1. For example, the famous (or infamous) Cauchy variable, whose probability density is $1/\pi(1+x^2)$, is almost surely finite but has an infinite variance and an infinite expectation. Thus, the choice between variables with extremely large finite and infinite variances can be decided on theoretical grounds: the latter is the more convenient to handle and we shall see it can accommodate the desirable scaling property.

Now let us move from moments to the next more sophisticated level of statistics. Mere curve fitting is notoriously controversial, in fact unavoidably so, since the quality of fit is a compromise between subjective simplicity of the fitting function and the largely arbitrary procedure used to measure the discrepancy between theory and data. Moreover, the study of price increments must deal at the same time with days, weeks, months, and (sometimes) years. One has no right to fix them separately, and there is no reason to expect the best-fitted monthly change to be the same as the sum of best-fitted daily changes. In all such cases, one must choose between fitting each category of data independently using distributions that are likely to be discordant and fitting all the categories simultaneously using a set of distributions that form an organized system. I view an organized system of fit as valuable. It is commonly used to buttress the Gaussian distribution, but in fact it is only an argument in favor of scaling.

Let us now review Bachelier's three steps towards his unacceptable Brownian motion model. They were (a) the efficient market idea, which is excellent; (b) the idea that successive price changes are independent with vanishing expectation, which is a very good approximation; and (c) the idea that price change variance is finite, which seems so obvious that is not even stated explicitly. From (b) and (c), application of the only central limit theorem known in 1900 led Bachelier to conclude that price follows a Brownian motion, thus proving the scaling property as a corollary.

STABLE DISTRIBUTIONS

The procedure in Mandelbrot 1963 is to use this last corollary as an assumption, replacing (c). In this context, a discussion of scaling begins by observing that, when G' and G'' are independent Gaussian random variables, with zero mean and mean squares equal to σ'^2 and σ''^2, the sum $G' + G''$ is also a Gaussian variable, with zero mean and a mean square equal to $\sigma'^2 + \sigma''^2$. In particular, the reduced Gaussian variable, defined as having zero mean and unit mean square, is a solution to the equation:

$$s'U + s''U = sU, \tag{S}$$

where the scale factor s is a function of the scale factors s' and s'', being given by the auxiliary relation:

$$s^2 = s'^2 + s''^2 \,. \tag{A_2}$$

Equation (S) has other solutions, however; for example, the Cauchy variable satisfies (S) combined with the auxiliary relation

$$s = s' + s'' \,. \tag{A_1}$$

More generally, given D satisfying $0<D<2$, equation (S) can be combined with the auxiliary relation

$$s^D = s'^D + s''^D \,. \tag{A_D}$$

The symmetric solutions of (S) and (A_D) have the characteristic function

$$\int_{-\infty}^{\infty} \exp(iuz)d\,Pr(U<u) = \exp(-\gamma|z|^D) ,$$

hence, the probability density

$$(1/\pi)\int_0^{\infty} \exp(-\gamma s^D)\cos(su)ds .$$

These densities are called *stable* (a dreadfully overworked term) or better *Lévy stable*. There is a discussion of them in volume II of Feller or in Lamperti 1966.

Tests of my optimistic conjecture, that price changes are symmetric stable, have proved it to be of very broad validity. The first tests (Mandelbrot 1963, 1967) applied to many commodity prices, to some interest rates, and to some old (nineteenth century) security prices; then Fama 1963 studied recent security prices and Roll 1970 studied other interest rates. Here we must be content with a single illustration, fig. 4.5, which combines doubly logarithmic graphs for large absolute cotton price relatives, together with the cumulated density function of the symmetrically stable distribution of exponent $D = 1.7$. The horizontal scale u of lines (1a), (1b), and (1c) is marked along the lower edge, and the horizontal scale u of lines (2a), (2b), and (2c) is marked along the upper edge. The vertical scale gives the frequencies of cases where the change of log Z exceeds the change in abscissa. The following three series of data are plotted:

(1a) $Fr[\log_e Z(t + \text{one day}) - \log_e Z(t) > u]$, (2$a$) $Fr[\log_e Z(t + \text{one day}) - \log_e Z(t) < -u]$, both for the daily closing prices of cotton in New York, 1900–1905 (communicated by the United States Department of Agriculture).

(1b) $Fr[\log_e Z(t + \text{one day}) - \log_e Z(t) > u]$, (2$b$) $Fr[\log_e Z(t + \text{one day}) - \log_e Z(t) < -u]$, both for an index of daily closing prices of cotton on various exchanges in the United States, 1944–1958 (communicated by Hendrik S. Houthakker).

(1c) $Fr[\log_e Z(t + \text{one month}) - \log_e Z(t) > u]$, (2$c$) $Fr[\log_e Z(t + \text{one month}) - \log_e Z(t) < -u]$, both for the closing prices of cotton on the fifteenth of each month in New York City, 1880–1940 (communicated by the United States Department of Agriculture).

The reader is advised to copy the horizontal axis and the theoretical distribution on a transparency and to move both horizontally. The theoretical curve will then be superimposed on either of the empirical graphs with slight discrepancies of general shape. This is precisely what my scaling criterion postulates. (The slight asymmetry can be handled too; it requires skew variants of the stable distribution.)

The curves (1a) and (1b), (2a) and (2b) would be identical if the pro-

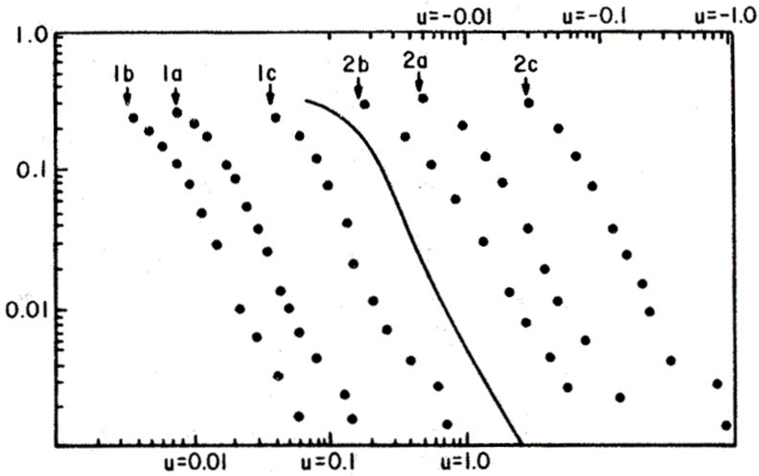

Fig. 4.5

cesses ruling cotton price change were stationary, but in fact they differ by a horizontal translation. Since translation on doubly logarithmic coordinates corresponds to a change of scale in natural coordinates, this discrepancy led me to concur in my 1963 paper with the economists' opinion that price change distributions around 1950 differed from their counterparts around 1900. I thought the distribution preserved the same shape but had a smaller scale.

More recently, however, I found that this concession to opinion was beyond necessity. I became aware that the data on which curves ($1a$) and ($2a$) were based had been read incorrectly (Mandelbrot 1972). Once this error is corrected, one is led to curves ($1a^*$) and ($2a^*$) that are nearly identical to the curves ($1b$) and ($2b$). In other words, the process that rules the changes in the price of cotton seems, in a first approximation, to have remained stationary over the very long period under study. There is no disputing the major changes in the value of currency and similar events. But such overall long trends are negligible in comparison with the fluctuations with which we deal here. One cannot deny that the data give at casual glance the impression of being grossly nonstationary, but this impression is the result of casual impressions formed against the background of a belief that the underlying process is Gaussian. My alternative to the nonstationary Gaussian process is a stationary non-Gaussian stable process. I believe the latter gives a much greater hope of eventually being related to a good economic theory and of yielding a realistic statistical algorithm.

To appreciate the nature of this achievement, it is vital to look carefully at the scales. In this instance, scaling implies that the distribution based on a record of daily price changes over a period of five years of average economic variability extrapolated to monthly price changes goes right through the data from the various recessions, the depression, and so forth. It accounts for all the most extreme events of nearly a century in the history of an essential and most volatile commodity. I do not believe there is any other comparably successful prediction in economics. It warrants being explored further.

OTHER FORMS OF SCALING

The impression that I devoted much of my life work to diverse facets of scaling is correct, even though I was late to recognize this fact myself and did not adopt this term wholeheartedly until a few years ago. The turning point was when I went from nongeometric examples, like those in the second part of this paper, to geometric examples, like those in the first part. Examples of both kinds abound in my 1977 book *Fractals* and in my 1982 book, *The Fractal Geometry of Nature*. Even my earliest work—which concerned so-called Zipf rank size rule of word frequencies—is now best understood and appreciated if it is not centered on the subject of statistical linguistics, but on the method of scaling.

Scaling is of course an ancient idea, thoroughly familiar to Leibniz; and the scientific application of scaling is the work of many hands. One facet has been familiar to zoologists since Julian Huxley under the term *allometry*. Another facet occurs in the urbanists' central place theory. In the theory of turbulence, scaling has been basic since the work of Lewis Fry Richardson.

CONCLUSION

It is an ancient observation that the variety of complexities in the real world is boundless, while the number of workable mathematical techniques to tame them is astonishingly small. Two most promising newcomers in that company are geometric scaling and the nonstandard geometric scaling shape, the fractals.

REFERENCES

Alexander, S. S. (1961) Price movements in speculative markets: trends on random walks. *Industrial Management Review of M.I.T.*, II, Part 2, 7–26. Reproduced in Cootner (1964) 199–218.

———. (1964) Price movements in speculative markets: No. 2. *Industrial Management Review of M.I.T.*, IV, Part 2, 25–46. Reproduced in Cootner (1964) 338–372.

Bachelier, L. (1900) Théorie de la spéculation. (Thesis for the doctorate in mathematical sciences, March 29, 1900.) *Annales de l'ecole normale supérieure*, série III, XVIII, 21–86. English translation in Cootner (1964).

Cootner, P. H. (1964) *The Random Character of Stock Market Prices*. Cambridge, Mass.: M.I.T. Press.

Fama, E. F. (1963) Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, XXXVI (October) 420–429. Reproduced in Cootner (1964).

———. (1965) The behavior of stock-market prices. *Journal of Business*, XXXVIII (January) 34–105. (Based on a Ph.D. thesis, University of Chicago: *The Distribution of Daily Differences of Stock Prices: a Test of Mandelbrot's Stable Paretian Hypothesis*.)

———, and Blume, M. (1966) Filter rules and stock-market trading. *Journal of Business*, XXXIX (January) 226–241.

Feller, W. (1968–1971) *An introduction to the Theory of Probability and Its Applications*. Vol. I (3d ed.) and Vol. II (2d ed.). New York: John Wiley & Sons.

Lamperti, J. (1966) *Probability, a Survey of the Mathematical Theory*. Reading, Mass.: W.A. Benjamin.

Mandelbrot, B. B. (1963) The variation of certain speculative prices. *Journal of Business*, XXXVI (October) 294–419. Reproduced in Cootner (1964) 307–332.

———. (1966) Forecasts of future prices, unbiased markets, and 'martingale' models. *Journal of Business*, XXXIX (January) 242–255 [important errata in a subsequent number of the journal].

———. (1967) The variation of some other speculative prices. *Journal of Business*, XL (October) 393–413.

———. (1971) When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, LIII (August) 225–236.

———. (1972) Correction of an error in "The variation of certain speculative prices." (1963) *Journal of Business*, XL (October) 542–543.

———. (1973a) Formes nouvelles du hasard dans les sciences. *Economie Appliquée*, XXVI, 307–319.

———. (1973b) Le syndrome de la variance infinie et ses rapports avec la discontinuité des prix. *Economie Appliquée*, XXVI, 321–348.

———. (1973c) Le problème de la réalité des cycles lents et et le syndrome de Joseph. *Economie Appliquée*, XXVI, 349–365.

———. (1975) *Les objets fractals: forme, hasard et dimension*. Paris and Montreal: Flammarion.

———. (1977) *Fractals: Form, Chance, and Dimension*. San Francisco and Reading: W. H. Freeman and Company.

———. (1978) Les objects fractals. *La Recherche* 9, No. 85 (January) 1–13.

———. (1982) *The Fractal Geometry of Nature*. San Francisco and Oxford: W. H. Freeman and Company.

Roll, R. (1970) *Behavior of Interest Rates: the Application of the Efficient Market Model to U.S. Treasury Bills*. New York: Basic Books. (Based on a Ph.D. thesis of the same title, University of Chicago, 1968.)