PENROSE TILES TO TRAPDOOR CIPHERS

...AND THE RETURN OF DR. MATRIX

Martin Gardner

Revised Edition

The Mathematical Association of America
A fascinating aspect of the history of mathematics is the way that the definitions of names for classes of mathematical objects are continually revised. The process usually goes like this: The objects are given a name, $x$, and defined in a rough way that conforms to intuition and usage. Then someone discovers an exceptional object that meets the definition but clearly is not what everyone has in mind when he calls an object $x$. A new and more precise definition is then proposed that either includes the exceptional object or excludes it. The new definition "works" as long as no new exceptions arise. If they do, the definition has to be revised again, and the process may continue indefinitely.

If the exceptions are strongly counter to intuition, they are sometimes called monsters. The adjective *pathological* is often attached to them. In this chapter we consider the word "curve," describe a few monsters that have forced redefinitions of the term and introduce a
frightening new monster captured last year by William Gosper, a brilliant computer scientist now with Symbolics, Inc., in Mountain View, California. Readers of my books have met Gosper before in connection with the cellular-automata game Life. It was Gosper who constructed the "glider gun" that made it possible to "universalize" Life's cellular space. (See the three chapters on Life in my Wheels, Life, and Other Mathematical Amusements.)

Ancient Greek mathematicians had several definitions for curves. One was that they are the intersection of two surfaces. The conic-section curves, for instance, are generated when a cone is cut at certain angles by a plane. Another was that they are the locus of a moving point. A circle is traced by a rotating compass leg, an ellipse by a moving stylus that is stretching a closed loop of string around two pins, and so on for other curves generated by more complicated mechanisms.

Seventeenth-century analytic geometry made possible a more precise definition. Familiar curves became the diagrams of algebraic equations. Could a plane curve be defined as the locus of points on the Cartesian plane that satisfy any two-variable equation? No, because the diagrams of some equations emerge as disconnected points or lines, and no one wanted to call such diagrams curves. Calculus suggested a way out. The word "curve" was limited to a graph whose points are a continuous function of an equation.

It seems intuitively obvious that if a curve is a continuous function, it should be possible to differentiate the function or, what amounts to the same thing, to draw a tangent to any point on the curve. In the second half of the nineteenth century, however, mathematicians began to find all kinds of monster curves that had no unique tangent at any point. One of the most disturbing of such monsters was described in 1890 by the Italian mathematician and logician Giuseppe Peano. He showed how a single point, moving continuously over a square, could (in a finite time) pass at least once through every point on the square and its boundary! (Actually any such curve must go through an infinity of points at least three times.) At the limit, the curve becomes a solid square. Peano's curve is a legitimate diagram of a continuous function. Yet nowhere on it can a unique tangent be drawn because at no instant can we specify the direction in which a point is moving.

David Hilbert proposed a simple way to generate a Peano curve with two end points. The first four steps of his recursive procedure should be clear from the pictures in Figure 20. At the limit the curve begins and ends at the square's top corners. The four steps in Figure 21 show how Waclaw Sierpinski generated a closed Peano curve.
In both versions think of the successive graphs as approximations approaching the graph of the limit curve. This limit curve in each version is infinitely long and completely fills the square even though each approximation misses an uncountable infinity of points both of whose coordinates are irrational. (In general the limit of a sequence of approximation curves may go through many points that are not on any of the approximations.) Sierpinski’s curve bounds an area 5/12 that of the square. Well, not exactly. The constructions approach this fraction as a limit, but the curve itself, the diagram of the limiting function, abolishes the distinction between inside and outside!
Peano curves were a profound shock to mathematicians. Their paths seem to be one-dimensional, yet at the limit they occupy a two-dimensional area. Should they be called curves? To make things worse, Peano curves can be drawn just as easily to fill cubes and hypercubes.

Helge von Koch, a Swedish mathematician, proposed in 1904 another delightful monster now called the snowflake curve. We start with an equilateral triangle and apply the simple recursive construction shown in Figure 22 to generate a crinkly curve resembling a snowflake. At the limit it is infinite in length; indeed, the distance is infinite between any two arbitrary points on the curve! The area bounded by the curve is exactly $8/5$ that of the initial triangle. Like a Peano curve, its points have

Figure 21 Sierpinski's closed Peano curve
no unique tangents, which means that the curve's generating function, although continuous, has no derivative.

If the triangles are constructed inward instead of outward, one gets the anti-snowflake curve. Its perimeter is also infinite, and it bounds an infinity of disconnected regions with a total area equal to \(\frac{2}{5}\) that of the original triangle. One can start with regular polygons of more than three sides and erect similar polygons on the middle third of each side. A square, with the added squares projecting outward, produces the cross-stitch curve of infinite length that bounds an area twice the original square. (See my *Sixth Book of Mathematical Games from Scientific American*, Chapter 22.) If the added squares go inward, they produce the anti-cross-stitch, an infinite curve that bounds no area. Similar constructions, starting with polygons of more than four sides, produce curves that self-intersect.

A 3-space analogue of the snowflake is constructed by dividing each face of a regular tetrahedron into four equilateral triangles, erecting a smaller tetrahedron on the central triangle and continuing the procedure indefinitely. Will the final result be a finite solid with a surface of infinite area? No, the astonishing answer (Gosper assures me) is that at the limit the surface becomes a perfect cube!

![The first four orders of Helge von Koch's snowflake](image)
We can generalize further by dividing the sides of a regular polygon into more than three parts. For example, divide each side of an equilateral triangle into five parts, erect smaller triangles on the second and fourth sections and repeat to the limit. For an ultimate generalization begin with any closed curve that can be divided into congruent segments, then alter the segments any way you like, provided the alteration is segmented so that the change can be repeated on the smaller segments and carried to the limit. Analogous constructions can be made on the surfaces of solids. Of course, the results may be messy, self-intersecting curves or surfaces of no special interest.

A book could be written about other kinds of pathological planar monsters. The Dutch topologist L. E. J. Brouwer published in 1910 a recursive construction for cutting a region into three subregions in such an insane way that at the limit all three subregions touch at every point (see “Geometry and Intuition,” by Hans Hahn; *Scientific American*, April 1954). Brouwer’s construction generalizes to divide a region into \( n \) subregions, all meeting at every point. A more recently discovered family of monsters, the dragon curves, were introduced in *Scientific American*’s Mathematical Games department in 1967 (reprinted in my *Mathematical Magic Show*, Knopf, 1977) and later analyzed by Chandler Davis and Donald Knuth in a 1970 article.

It is now my privilege to present Gosper’s new monster, a beautiful Peano curve that he calls the flowsnake. Its construction starts with a pattern of seven regular hexagons (see Figure 23). Eight vertexes are joined as shown by the gray line, made up of seven equal-length segments. The gray line is order 1 of the flowsnake. Order 2, shown in black, is obtained by replacing each gray segment with a similar twisted line of seven segments. Each segment of the black line is \( 1/\sqrt{7} \) the length of a gray segment; this proportion holds at every stage of the construction. The recursive procedure is continued to produce flowsnakes of higher orders. Figure 24 shows two computer drawings of flowsnakes of orders 3 and 4. By dividing the plane into black and white, with the bifurcating line passing through the flowsnake’s end points, we see how the curve cuts the plane into two regions that twist about in almost, but not quite, the same pattern.

The curve that diagrams the limit of the successive flowsnake functions passes through every point of its region at least once, completely filling the space. The curve is infinite and nondifferentiable. Like the straight line, it is self-similar in the sense that if you enlarge any portion of it, the pattern always looks the same. Snowflake curves have the same property.
In the light of these crazy curves, how do mathematicians currently define a curve? The scene is so crowded with monsters that no single definition covers all the objects to which the word “curve” is commonly applied. The topologist defines a curve as a set of points that are compact and connected and form a 1-dimensional continuum. To make the definition clear, however, a lengthy discourse on point-set topology would be required. The definition catches well-behaved curves that diagram functions with derivatives, but it misses some of the nondifferentiable monsters we have been considering. “Of course we have no physical snowflake curves,” Philip Morrison has written. “Nature gives no infinities, not even within molecular collisions. There is a cutoff at the ang-
Figure 24 Flowsnakes of order 3 (top) and order 4 (bottom)

Mandelbrot is a Polish-born (Warsaw, 1924) French mathematician who is currently an IBM Fellow at the Thomas J. Watson Research Center at Yorktown Heights, N.Y. Like Stanislaw Ulam and many other eminent Polish mathematicians, Mandelbrot has had a career involving a marvelous mixture of creative work in both pure and applied mathematics, notably in physics and economics. His teacher, the French mathematician Paul Pierre Lévy, made the first systematic study of statistically self-similar curves, but they were regarded as useless, bizarre curiosities until Mandelbrot recognized them as being a basic tool for analyzing an enormous variety of physical phenomena.

Mandelbrot’s book is filled with pictures of just such phenomena. Consider coastlines. Their butterfly-flight irregularity is statistically self-similar. A coastline looks the same from a high altitude as it does from a low one. It is meaningless to speak of a coastline’s “length” because it all depends on the precision of measurement. As Morrison puts it, “a coastline on maps at varying scales obeys a power law like the snowflake curve’s, from a scale of hundreds of kilometers down to one of perhaps meters, where geography stops and pebbles begin.”

The surface of the moon is another example. Remember your surprise on seeing the first closeup photographs of the moon made from a satellite in orbit around it? The moon’s pocked surface looked basically as it did in photographs made with telescopes on the earth. Only the crater sizes were different. The same random self-similarity is found on the surface of certain cheeses, in the scattering of stars in the sky, on the bark of trees, in the contours of mountains, in atmospheric turbulence, in auditory noise and in countless other natural patterns. The Brownian motion of suspended particles approximates a statistically self-similar curve that (at the limit) has infinite length and no tangents.

Let us go back to the flowsnake for a close look at its perimeter and at an amazing paradox. The perimeter can be constructed by a recursive procedure much simpler than the one used to get the flowsnake itself. Figure 25 shows how it works. Start with a regular hexagon, then replace each side with a zigzag line (gray) of three equal segments, each $1/\sqrt{7}$ the original side. The result is a nonconvex 18-gon. Since the zigzag line adds the same amount of area as it takes away, the 18-gon obviously has
the same area as the original hexagon. Repeat the construction on each of the 18 sides to produce a 54-gon, and imagine that the recursive procedure is continued to the limit. At each step the number of sides triples, but the area never changes. At the limit the area filled by the flowsnake is exactly the same as the area of the original hexagon.

The entire region has an astounding property. It can be dissected, as is shown in Figure 26, into seven subregions, each of which is an exact copy of the entire region.

Now for the paradox. What is the ratio of the area of a subregion to the entire region? Clearly it is $1/7$, since seven identical subregions make up the whole. But let us approach it from another angle, remembering that the areas of similar figures are proportional to the square of their linear dimensions. The outside perimeter consists of six segments, such as the segment from A to B, which is half the closed boundary of a subregion. Clearly the boundary of a subregion must be enlarged by a linear factor of 3 to fit the boundary of the entire region. But if this is true, the areas must be in a ratio of $(1/3)^2 = 1/9$. We seem to have proved that the ratio of the areas is both $1/7$ and $1/9$. As Gosper asked when he first sent the paradox, Voss ist los?

The answer lies in the peculiar, counterintuitive character of the pathological boundary. There is no fuzziness about the area of the region it bounds. It is indeed seven times the area of a subregion. Nor is the boundary fuzzy. It looks fuzzy, but it is nevertheless a precisely defined infinite set of points. It has, however, a strongly counterintuitive property. By what linear factor must the boundary of one of the seven subregions be enlarged to make it congruent with the outside overall boundary? One would suppose a factor of 3, but the actual factor is $\sqrt{7} = 2.645. . .$. It is impossible, of course, to print the boundary because at the limit its complexity is infinite. Only a few steps of its construction can be shown before the ink begins to smear.
A deep question now arises. What "dimension" should be assigned to the flowsnake's boundary? Like the snowflake, it lies in a strange twilight zone between one dimension and two dimensions. In 1919 a German point-set topologist, Felix Hausdorff, resolved the difficulty by giving fractional dimensions to such curves, or what Mandelbrot calls "fractal" dimensions. The term was invented by Mandelbrot about 1975. He based it on the Latin verb *frangere*, which means "to break," and its adjective form, *fractus*. The term suggests the broken, fragmented character of fractals, and also the fractional numbers that provide, as we shall see, a fractal's degree of shagginess. Fractal dimensions should not be confused with Hausdorff spaces—topological structures that mercifully we do not have to go into here.

The familiar Euclidean dimensions 0, 1, 2, 3, 4 . . . are sometimes called topological dimensions because their spaces are topologically

*Figure 26* A flowsnake paradox
distinct; that is, you can't transform one space into another by continuous topological deformation. A point has topological dimension 0. Smooth, well-behaved curves such as straight lines, circles, parabolas and so on have a Euclidean dimension of 1. Surfaces have dimension 2; solids, dimension 3; and hypersolids, higher integers.

To grasp how fractal dimensions are calculated, consider first a straight line segment. If magnified by a factor \( x \), the enlarged line can be cut into \( y \) copies of the original. The dimension of the line segment is the exponent of \( x \) that gives \( y \). In this case \( x = y \) because doubling the line segment produces two copies of the segment, tripling the line segment produces three copies and so on. We can express the scaling ratio by writing \( \log 2 / \log 2 = 1 \).

Magnify a square by doubling its edge. The enlarged square can be cut into four copies of the original. If you triple the edge, it can be cut into nine copies. Generally, if you magnify a plane figure by a linear factor \( x \), its area increases by a factor of \( x^2 \). The dimension of a square, therefore, is \( \log 4 / \log 2 = 2 \). If you double a cube's edge, the enlarged cube can be cut into eight copies of the original. Its dimension is \( \log 8 / \log 2 = 3 \). And so it goes for hypercubes in higher topological (Euclidean) spaces.

Let us now apply this technique to the snowflake. If you enlarge a portion of it by a linear factor of 3, it produces 4 copies of the original. At each construction step the ragged line is exactly \( 4/3 \) times the length of the previous line, although each straight segment is \( 1/3 \) the length of the previous segment. It is reasonable, therefore, to assign to the limit curve a Hausdorff dimension \( D \), or fractal dimension, that is \( \log 4 / \log 3 = 1.261859 \).

Calling these numbers "dimensions" is somewhat misleading. They are not Euclidean dimensions. It is best to think of them as measures of complexity, of the "degree of wiggliness," as Mandelbrot once put it. The complexity of Gosper's flowsnake boundary is a trifle lower than the snowflake. Figure 27, from Mandelbrot's 1977 book, is made by replacing lines of four units with lines of eight units to produce a squarish asymmetric snowflake with complexity \( \log 8 / \log 4 = 1.5 \). Its fractal dimension, therefore, is a bit higher than the flowsnake boundary. Because each construction step adds the same amount of area as it takes away, the limit curve bounds the same area as the original square.

We can express fractal dimensions with the general formula \( D = \log n / \log \sqrt[1/r]{ } \), where \( n \) is the number of self-similar parts that result when the original is magnified by factor \( r \) and \( D \) is the fractal dimension. Fractal
curves that are 1-dimensional in the Euclidean sense can have a fractal number that ranges from 1 to 2 if the curve is on the plane, but can go to higher numbers if the curve twists through higher Euclidean spaces. If a fractal curve passes through every point on the plane within its boundary, as does the flowsnake (the snake itself, not its boundary) and other Peano curves, then it achieves at the limit both a Euclidean dimension and a fractal dimension of 2. If it passes through every point in a solid, it achieves at the limit Euclidean and fractal dimensions of 3. Similarly, a fractal surface of Euclidean dimension 2 can have a fractal dimension that ranges from 2 to 3 as long as it is confined to 3-space, but can go higher if it twists through higher Euclidean spaces, and so on for fractal structures of higher topological dimensions.

Can fractals have complexity numbers less than 1? The answer is yes, although the structures are not topologically equivalent to line segments. For example, remove the center third of a line segment, then remove the central third of each of the two remaining segments and continue this process of third removing to infinity. The result is what Mandelbrot likes to call “Cantor dust.” In the literature they are called Cantor discontinua or Cantor sets, after Georg Cantor, who studied them. When you apply Mandelbrot’s formula for fractals, you get a number between 0 and 1. In this case, the number is $D = \log \frac{2}{3} = 0.6309$. Other “cutout” procedures give other numbers. Mandelbrot calls them “subfractals” to distinguish them from fractals with numbers higher than their Euclidean dimension.

In the abstract world of pure mathematics, fractal structures such as we have considered are called “ordered fractals.” In the real world there
are, of course, no ordered fractals. Structures such as coastlines, trees, rivers, clouds, blood vessels, lightning bolts, paths of particles in Brownian motion and thousands of other fractal-like phenomena are imperfect models that are (within certain upper and lower bounds) fractals in a statistical sense. They are self-similar in that they preserve a statistical similarity that is independent of scaling. One must average the fractal numbers at different scalings, and of course to do this one must make empirical studies. Such fractals are called random fractals or statistical fractals. Coastlines, for instance, have fractal dimensions that change from coast to coast. Investigations show that they fall within a range of 1.15 to 1.25, the second number measuring the complexity of the west coast of England.

The surfaces of mountains are splendid models of random surface fractals. Mandelbrot and his associates, notably Richard F. Voss at IBM, have for the past few years been writing computer programs that generate artificial mountain ranges, clouds and imaginary planets with artificial oceans and continents. Plate 2 shows one of Voss’s most striking computer displays: an earthlike planet that never was, as viewed from an equally artificial moon. Artificial clouds have also been produced by computer programs based on fractal formulas. Trees have proved to be more difficult to mimic, but Michael Barnsley and his colleagues at the Georgia Institute of Technology have found ways to mimic leaves and ferns (see Ivars Peterson’s 1987 article). These computer graphic displays have led to the generation of strange artificial landscapes for imaginary planets in science fiction films, starting with Star Trek II: The Wrath of Khan.

Cantor sets can be constructed in any Euclidean space to make spongelike structures of “dust” with fractal numbers less than the number of the space. Mandelbrot has produced random fractals of Cantor dust in 3-space that model to a surprising degree the apparent distribution of stars in the universe. The hierarchy of star clusters, superclusters and super-superclusters suggests that perhaps the entire universe is, within limits, close to the structure of a random fractal.

**ADDENDUM**

Although this chapter has been considerably expanded and updated since it first appeared in Scientific American in 1976, some additional updating is called for.
In 1967, when Mandelbrot first wrote about fractals in his classic paper "How Long is the Coast of Britain?," he surely could not have anticipated the speed with which his work would trigger revolutions in both mathematics and physics. Not only have fractals become one of the most energetic research areas of topology, with high-speed computers serving as essential tools, they have also become fundamental aspects of a revolutionary new field of physics called chaos theory. The topic is so vast and developing so rapidly that I can do no more here than refer the reader to the first nontechnical book on the topic, James Gleick's splendid *Chaos: Making a New Science* (Viking, 1987).

Gleick's survey will introduce you to the most important fractals involved in what are called the "strange attractors" of chaos, to such beautiful fractals (on the complex plane) as Julia sets and to the incredible Mandelbrot set (M-set) that Mandelbrot found in 1980. It has been called the most mysterious object in geometry. As research continues, the definition of fractal has broadened to a point at which Mandelbrot has proposed that a final definition be postponed until things settle down. It is no longer necessary, for example, that ordered fractals preserve self-similarity at all scalings. All sorts of "nonlinear fractals" have been constructed, such as the affine fractals that show affine distortions at successive magnifications. The M-set, a sort of dictionary for all Julia sets (I won't try to explain what that means!), is not self-similar, except in a topological way, although it contains an infinite number of copies of itself. Every new level of enlargement reveals unpredictable surprises. For more than a year, as successive magnifications were made, it was not even known if the set is connected. Each enlargement disclosed isolated "islands" and particles of dust. Further enlargements would join these disconnected portions to the mainland, but new islands and dust would appear. It was finally proved in 1982 that the set is indeed connected at the limit, but it may be decades before its major properties are uncovered.

In an interview in *Omni* (June 1986) the British mathematical physicist Roger Penrose invoked the Mandelbrot set to support his Platonic approach to mathematics:

Have you ever seen those pictures produced by computers, the object known as the Mandelbrot set? It's as if you are traveling to some distant world. You turn on your sensing device and see this incredibly complicated configuration, with all sorts of structure to it, and you try to figure out what it is. You might think it is some extraordinary landscape or perhaps some kind of creature with lots of little babies all over the place, babies that are almost but not quite the same as the creature
itself. Very elaborate and impressive! Yet just from seeing the equations, nobody had the remotest conception that they would produce patterns of this nature. Now these landscapes aren’t conjured up out of someone’s imagination; everyone sees the same pattern. You’re exploring something with a computer, but it’s not dissimilar from exploring something with experimental apparatus.

If you know your way around the complex plane and are interested in exploring by computer the amazing jungles of the M-set, you should subscribe to *Amygdala*. This is a monthly newsletter that reports new discoveries about the M-set, more efficient computer techniques for investigating it and anything else related to the M-set that strikes the editor’s fancy. It has even introduced a new subgenre of science fiction called M-set SF, centered around such notions as that the M-set is a living entity inhabiting hyperspacetime and possessing paranormal powers. Rollo Silver is the newsletter’s founder and editor. He describes himself as “an ontological engineer who lives and works in the mountains of Northern New Mexico. Deprived of the company of his peers and half-crazed by isolation, he started *Amygdala* in self-defense in 1986.”

You can get a flyer containing samples of various issues and information on how to subscribe to the newsletter and to a color-slide supplement by writing to *Amygdala*, Box 219, San Cristobel, NM 87564. *Amygdala*, by the way, is Latin for almond. The title honors Mandelbrot, whose name in German and Yiddish means “almond bread.”

Mandelbrot’s *The Fractal Geometry of Nature* is surely one of the most beautiful, witty and exciting books about mathematics ever published. Its text and breathtaking graphics catch such marvelous monsters as the Devil’s staircase, Minkowski sausages, Gosper’s fudgeflake, Bernoulli clusters, Sierpinski carpets, Menger sponges, Fatou dust, squigs, dragons and all kinds of curds and cheeses.

Since 1987 Mandelbrot has been a professor of mathematics at Yale University. In 1980 he was given the F. Bernard Medal for Meritorious Service to Science by Columbia University, a prestigious award recommended every five years by the National Academy of Sciences. Since then he has received six other distinguished service awards and six honorary doctorates. There are sure to be more honors to come. He has been called the most versatile mathematician since John von Neumann and Norbert Wiener.
BIBLIOGRAPHY

Books and articles by Benoit Mandelbrot


Books about fractals


Nontechnical articles and interviews


Books on Fractals and Chaos Theory