Four Encounters with Sierpiński's Gasket[1]

Ian Stewart

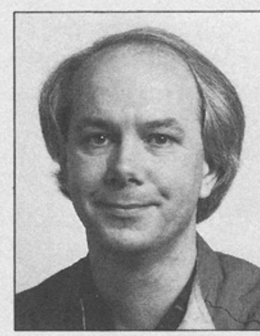
> Mathematicians would not be happy merely with simple, lusty configurations. Beyond these their curiosity extends to psychopathic patients, each of whom has an individual case history resembling no other; these are the pathological curves of mathematics.
>
> Edward Kasner and James Newman
> *Mathematics and the Imagination*

One of the most fascinating features of mathematics is the way in which the same idea crops up again and again in apparently unrelated areas. Over the last few years, I have been haunted by the object that Benoît Mandelbrot [1] has christened *Sierpiński's gasket*. It is the triangular fractal shown in Figure 1. Nowadays, fractals are respectable, and the sentiments expressed in the above quotation seem old-fashioned: it shows how much attitudes have changed. Most people's, anyway. Sierpiński's gasket arises naturally in many branches of mathematics, and my aim is to convince you that without it, mathematics would be the poorer. But first, a few words about the man himself.

Explorer of the Infinite

Wacław Sierpiński was born in Warsaw, Poland, on 14 March 1882. His father Konstanty Sierpiński was a doctor. My source, Kasimierz Kuratowski [2], Vol. 1, fails to record any details about his mother. Sierpiński studied under the number theorist G. Voronoï, and his early work was also in number theory, a topic to which he repeatedly returned in later life. He obtained a doctorate in 1906, and by 1909 he had moved to the University Jean Casimir in Lvov, becoming a professor there in 1910. The year 1909 is more significant, however, because in that year Sierpiński gave the first systematic lecture course ever taught on set theory. He published a book based on it in 1912, which was among the first texts on that subject. Sierpiński had found his subject, and the bulk of his subsequent research was in set theory and point set topology.

Ian Stewart



Ian Stewart is a professor at the University of Warwick, where he is director of its Interdisciplinary Mathematical Research Programme. He has held visiting positions in Germany (Tübingen), New Zealand (Auckland), and the United States (Storrs, CT and Houston, TX). He is currently Gresham Professor of Geometry (Gresham College, London). He works on the effects of symmetry on dynamics, including applications to pattern formation and chaos, taking a particular interest in problems that lie between pure and applied mathematics. He is the author of *Singularities and Groups in Bifurcation Theory* (with Martin Golubitsky and David Schaeffer) and *Catastrophe Theory and Its Applications* (with Tim Poston). Much of his writing is centred on popular science and includes *The Collapse of Chaos* (with Jack Cohen), *Fearful Symmetry: Is God a Geometer?* (with Martin Golubitsky), *Does God Play Dice?*, *The Problems of Mathematics*, *Another Fine Math You've Got Me Into*, *Game, Set & Math*, and *Concepts of Modern Mathematics*. He is the mathematics consultant for *New Scientist* and writes the bimonthly "Mathematical Recreations" column of *Scientific American*. He has published science fiction in *Omni* and *Analog*, is an active member of SFWA (Science Fiction and Fantasy Writers of America), and is currently finishing a science fiction novel, *Jack Of All Trades*.

[1] This article is an expanded and somewhat rewritten version of the Lonseth Lecture given at Oregon State University, Corvallis on 14 May 1991, and the London Mathematical Society Popular Lecture given at Sheffield University on 17 June 1991 and Imperial College London on 28 June 1991.

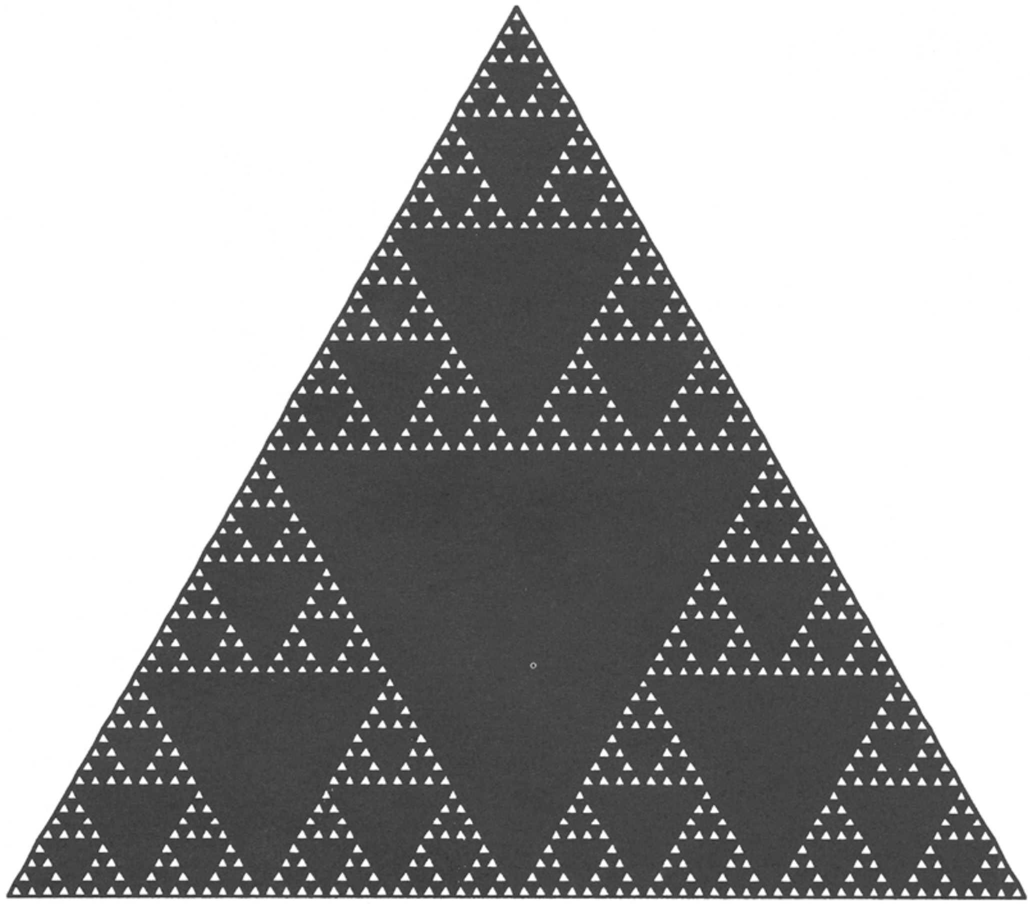


Figure 1. The Sierpiński gasket.

During the First World War, Sierpiński was interned in Russia, first at Viatka and then in Moscow. There he worked with the Russian mathematician N. Lusin on projective sets and real functions. Their first joint paper appeared in 1917 and the last in 1929. At the end of the war, Sierpiński returned to Lvov, but almost immediately became a professor at the now reestablished University of Warsaw. The place was a hotbed of Polish mathematics, specialising in set theory and foundational matters, with people such as Zygmunt Janiszewski, Stefan Mazurkiewicz, and Jan Łukasiewicz. Together with Sierpiński, the first two started their own journal, *Fundamenta Mathematicae,* which exists to this day. Lvov, too, became a major centre of Polish mathematics under Stefan Banach — see Ciesielski [3, 4] for the atmosphere of this period, including the story of the famous "Scottish Café" in Lvov.

Between the wars, Sierpiński's talents flourished. He was always prolific: his collected works include 720 papers published between 1906 and 1968, 106 expository articles, 50 books (plus 7 at the level of secondary education), and 12 mimeographed sets of lecture notes. The start of World War II found him still in Warsaw, where he continued his scientific work as best he could, teaching clandestine courses at the university to small audiences. After the uprising of 1944, he was deported by the Germans to the region around Kraków. In 1945, he briefly lectured at the Jagiellonian University of Kraków, before returning once more to Warsaw. In 1958, he wrote a major monograph, *Cardinal and Ordinal Numbers.* He remained very active in administrative matters and received a number of important prizes and other honours from the Polish government. He died in Warsaw on 21 October 1969. Following Sierpiński's wishes, his grave bears just two words (in Polish): *Explorer of the Infinite.*

Encounter 1: Sierpiński's Encounter with Sierpiński's Gasket (Wacław Sierpiński, 1915)

My claim is that (well before Koch, Peano, and Sierpiński) the tower of Gustav Eiffel built in Paris deliberately incorporates the idea of a fractal curve full of branch points.

Benoît Mandelbrot
The Fractal Geometry of Nature

The gasket made its first appearance in an article only three and a half pages long [5]. (Though, being published in *Comptes Rendus,* it couldn't have been very much longer without infringing the rule brought in — it is said — to prevent Augustin-Louis Cauchy from filling every issue with vast screeds.) A more detailed treatment followed a year later [6]; see also Ref. 2. The gasket's role was to provide an example of "a curve simultaneously Cantorian and Jordanian, of which every point is a point of ramification"; less formally, a curve that crosses itself at every point. A *point of ramification* of a curve C is a point p such that there exist three subsets of C, all continua, of which any pair intersect only at p.

Sierpiński's own diagram of the construction of this curve is shown in Figure 2(a). He first establishes that it is a Cantor curve (a continuum that is not dense in the plane). By a careful study of the process by which the various triangles and subtriangles are constructed, he then proves that every point *other than the three vertices of the original triangle* is a point of ramification. These three vertices are clearly not points of ramification, but

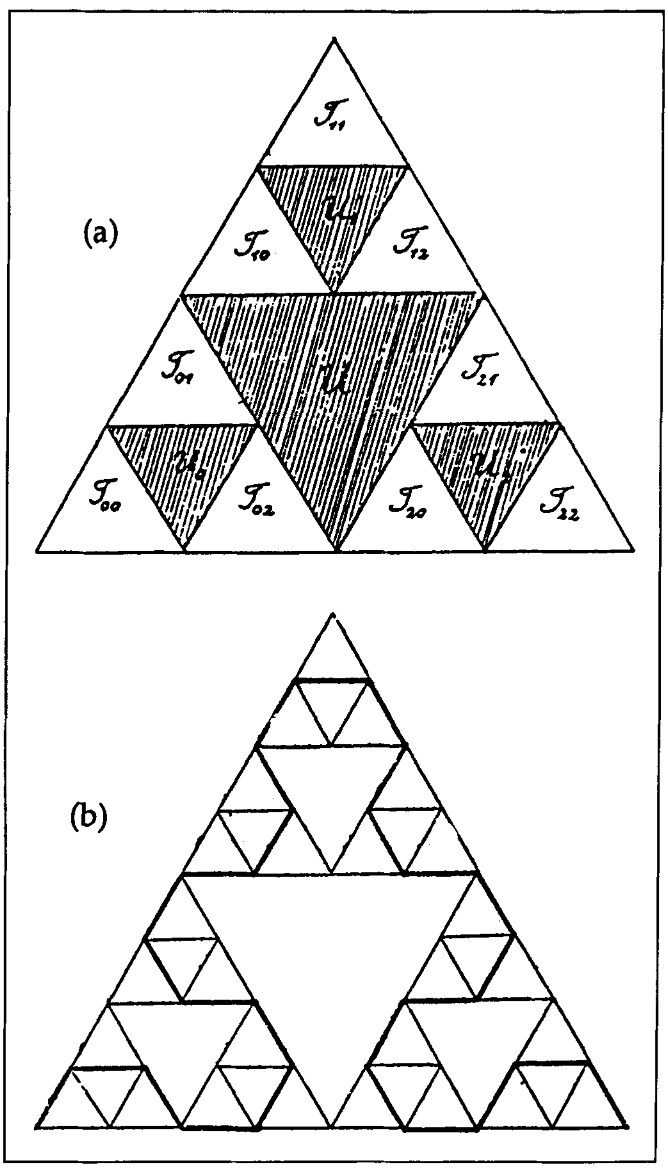


Figure 2. Sierpiński's version.

before dealing with them, Sierpiński offers Figure 2(b) as a sketch proof that his set is also a Jordan curve. Finally, he observes that if six copies of his triangle are arranged to form a regular hexagon, then the result is a Cantor and Jordan curve for which *every* point is a point of ramification.

Sierpiński's curve is of course a fractal, though that word was not coined until 1975 by Mandelbrot [1], who also, in jest, introduced the term *Sierpiński gasket.* At about the same time, Sierpiński invented several other celebrated fractals, including his space-filling curve [7] and the Sierpiński carpet [8]. He also invented several functions with fractal properties: a function [9] that has zero derivative almost everywhere, yet climbs monotonically from 0 to 1 (a forerunner of the "devil's staircase" [1]), and a function f such that $f(f(x)) = x$ whose graph is dense in the plane [10]. (You might like to try to construct such a function; see below for Sierpiński's solution, a typical example of his ingenuity.) Because the gasket is assembled from three copies, each half the size, its fractal (or Hausdorff–Besicovitch) dimension is $\log 3/\log 2 = 1.5849\ldots$. See Ref. 11 for details. It has a three-dimensional relative, to which Mandelbrot [1] gives the less inspired name "a fractal skewed web" (Fig. 3), but which I prefer to call the *Sierpiński cheese.* Curiously, this has fractal dimension $\log 4/\log 2 = 2$, the same as that of an ordinary Euclidean plane. Observe that the section cut away at each stage is *not* an inverted tetrahedron, which is why tetrahedra — contrary to Aristotle — do not tile space.

Solution To get a function f such that $f(f(x)) = x$ with a dense graph, define $f(a + b\sqrt{2}) = b + a\sqrt{2}$ for rational a and b; otherwise define $f(x) = x$.

Sierpiński invented a fractal 60 years before the word existed. Mandelbrot — with some justification — suggests that Eiffel invented the moral equivalent of the Sierpiński gasket 26 years before Sierpiński did. A year later, in 1890, another Frenchman characterised a combinatorial incarnation of the Sierpiński gasket:

Encounter 2: Pascal's Encounter with Sierpiński's Gasket (Edouard Lucas, 1890)

> I have yet to see a problem, however complicated, which when you looked at it in the right way, did not become still more complicated.
>
> Poul Anderson

As the section title shows, attributions for this encounter are tricky: The material goes back so far into the collective mathematical consciousness that it is difficult to award credit to any specific person. Instead, we record

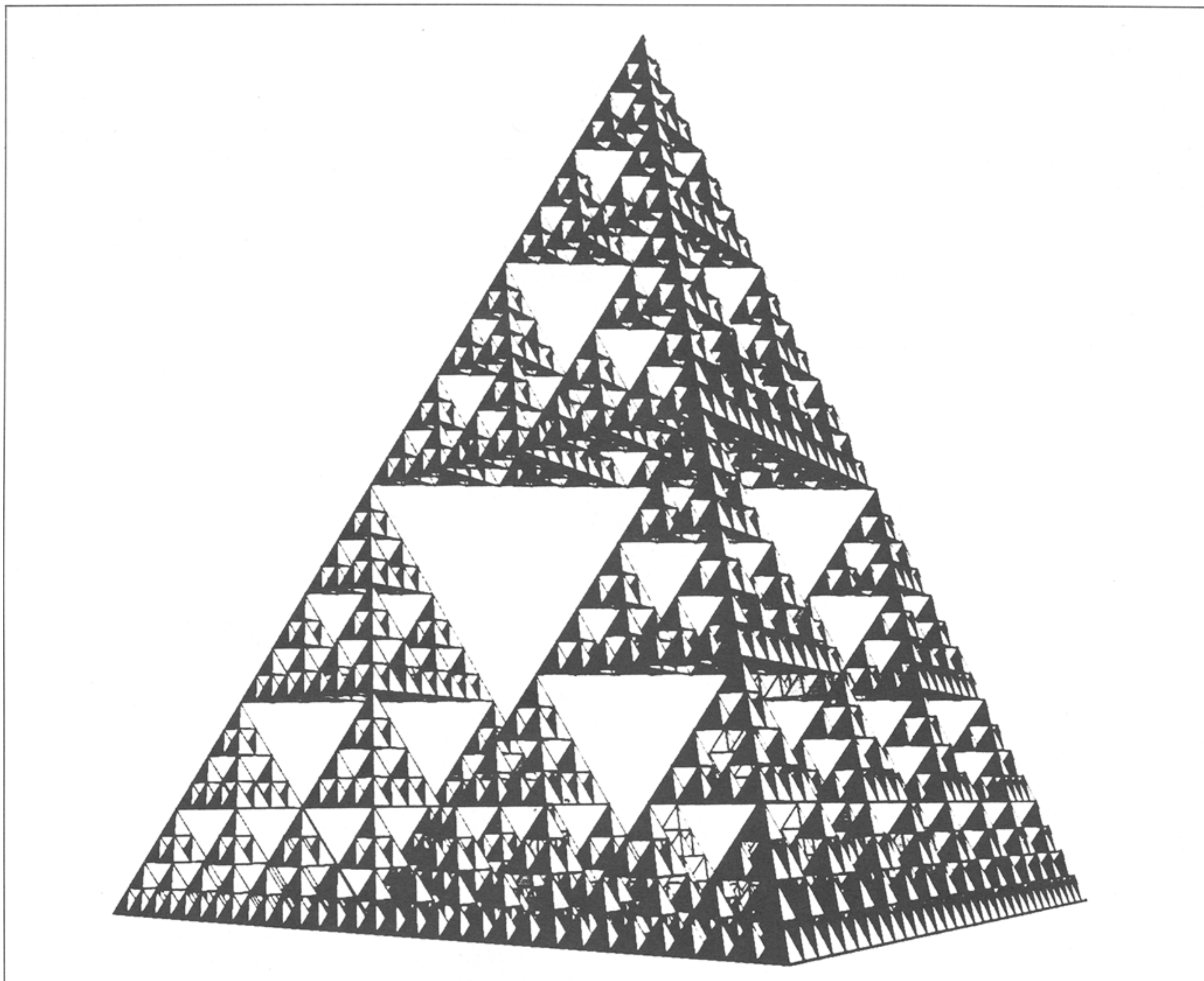


Figure 3. The Sierpiński cheese.

some milestones. Pascal gets credit for the encounter because it is his name that is attached to the triangular array of binomial coefficients $\binom{n}{r}$. Like most attributions from the distant past (and many from the near present), it is utterly wrong — for example, the triangle appears on the title page of an early 16th-century arithmetic by Petrus Apianus; it can be found in a Chinese mathematics book of 1303; and, indeed, it has been traced back at least to Omar Khayyám around 1100, who almost certainly got it from earlier Arabic or Chinese sources. Michael Stifel introduced the term *binomial coefficient* around 1500. The explicit formula $n!/r!(n-r)!$ was given by Isaac Newton and permitted the nonrecursive computation of binomial coefficients. In its interpretation as the number of ways to choose r items from a set of n, this expression (though not in that notation) was known to Bhāskara (b. 1114).

What is the parity of $\binom{n}{r}$? That is, what is its value (mod 2)? It is easy to experiment with a computer because it suffices to implement the rule of formation of the triangle mod 2. The result, Figure 4, is striking and surprising.

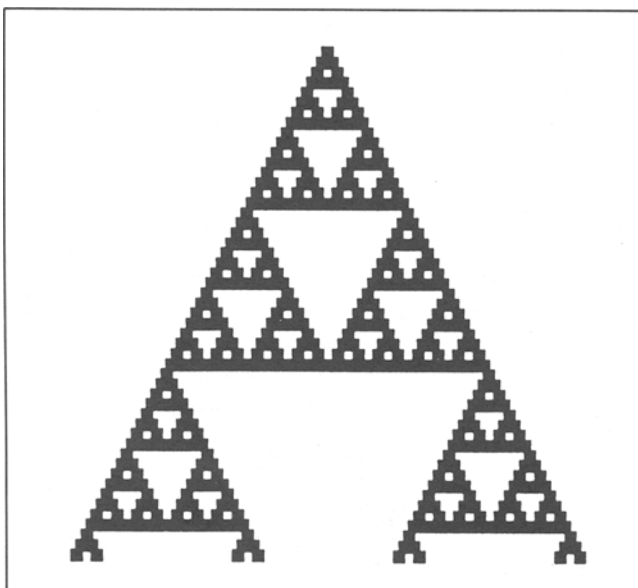


Figure 4. Pascal mod 2.

The odd binomial coefficients form a discrete variant of the Sierpiński gasket. Indeed, by suitably rescaling segments of the figure and taking an appropriate limit, we may consider the parity-coloured Pascal triangle to be a second manifestation of the gasket.

The ultimate explanation for this pattern is a theorem attributed by Dickson [12] to Edouard Lucas in 1890:

THEOREM 1: *Let p be a prime. Write n and r in p-ary notation:* $n = n_k \cdots n_0$, $r = r_k \cdots r_0$, *where the* n_j *and* r_j *are* $0, 1, \ldots, p-1$. *Then*

$$\binom{n}{r} \equiv \prod_{j=0}^{k} \binom{n_j}{r_j} \pmod{p}.$$

We make the standard convention that if r does not lie in the range $0 \le r \le n$, then the value of $\binom{n}{r}$ is 0. (An attractive convention which differs from this if n or r is negative may be found in Ref. 13.) Taking $p = 2$ and observing that

$$\binom{0}{0} = \binom{1}{0} = \binom{1}{1} = 1, \quad \binom{0}{1} = 0,$$

it follows that $\binom{n}{r}$ is even if and only if some binary digit of n is 0 while the corresponding digit of r is 1. Similar statements can be made for any prime p.

Lucas's theorem explains the recursive Sierpiński pattern of Figure 4. It implies that when n and r are expressed in binary as above, the parities of

$$\binom{n}{r} = \binom{n_k \cdots n_0}{r_k \cdots r_0},$$
$$\binom{2^{k+1}+n}{r} = \binom{1n_k \cdots n_0}{0r_k \cdots r_0},$$
$$\binom{2^{k+1}+n}{2^{k+1}+r} = \binom{1n_k \cdots n_0}{1r_k \cdots r_0}$$

are the same, whereas

$$\binom{n}{2^{k+1}+r} = \binom{0n_k \cdots n_0}{1r_k \cdots r_0}$$

is always even.

The patterns for $\binom{n}{r}$ (mod k), some of which are shown in Figure 5, are at least as pretty as that for $k = 2$. They have been discussed in this journal by Sved [14]; see also Refs. 15 and 16.

When $k = p^s$ is a prime power, there is a generalization of Lucas's theorem, due to Kazandzidis [17], which explains the patterns in the same manner as Lucas's theorem does and relates them to the base k expansions of n and r. For other values of k, it seems to be necessary to write k as a product of prime powers k_j and consider expansions to all bases k_j.

As the gasket is a fractal of dimension less than 2, its area (two-dimensional Hausdorff measure) is zero. It follows that almost all binomial coefficients are even. Singmaster [18] takes this observation further, proving that for any m, almost all binomial coefficients are divisible by m.

A refinement of Lucas's theorem was proved by Glaisher [19]. Similar results were known to Kummer [20].

THEOREM 2: *Let p be prime. The largest power of p that divides* $\binom{n}{r}$ *is equal to the number of carries in the* (mod p) *addition of* r *and* $n - r$.

A proof is given in Ref. 21. Related articles include Refs. 22–26.

Encounter 3: Hinz's Encounter with Sierpiński's Gasket (Andreas Hinz, 1990)

In the great temple at Benares, beneath the dome which marks the centre of the world, rests a brass plate in which are fixed three diamond needles, each a cubit high and as thick as the body of a bee. On one of these needles, at the creation, God placed sixty-four discs of pure gold, the largest disc resting on the brass plate, and the others getting smaller and smaller up to the top one. This is the Tower of Bramah. Day and night unceasingly the priests transfer the discs from one diamond needle to another according to the fixed and immutable laws of Bramah, which require that the priest on duty must not move more than one disc at a time and that he must place this disc on a needle so that there is no smaller disc below it. When the sixty-four discs shall have been thus transferred from the needle on which at the creation God placed them to one of the other needles, tower, temple, and Brahmins alike will crumble into dust, and with a thunderclap the world will vanish.

M. de Parville
La Nature, 1884

Edouard Lucas also seems to have been haunted—albeit unwittingly—by Sierpiński's gasket. In 1883, he marketed the puzzle known as the *Tower of Hanoi*, under the pseudonym N. Claus [27]. It is similar to de Parville's romantic Tower of Brahma but uses eight (or fewer) discs made of more mundane materials; moreover, instead of needles, it employs pins. It is an old friend of recreational mathematicians and computer scientists.

Number the pins 1, 2, and 3: Assume all discs start on pin 1 and must end on pin 2. It is well known that the solution has a recursive structure, explained, for example, in Ref. 28. I call it the *army method*. This particular army has a large number of privates, who have been trained to solve 1-disc Hanoi but nothing more ambitious:

- "Move the disc to pin 2 while keeping it in order of size—yes Bloggs, I know you can't change the order of one disc, but I wouldn't be surprised if you 'orrible lot found a way to screw even *that* up."

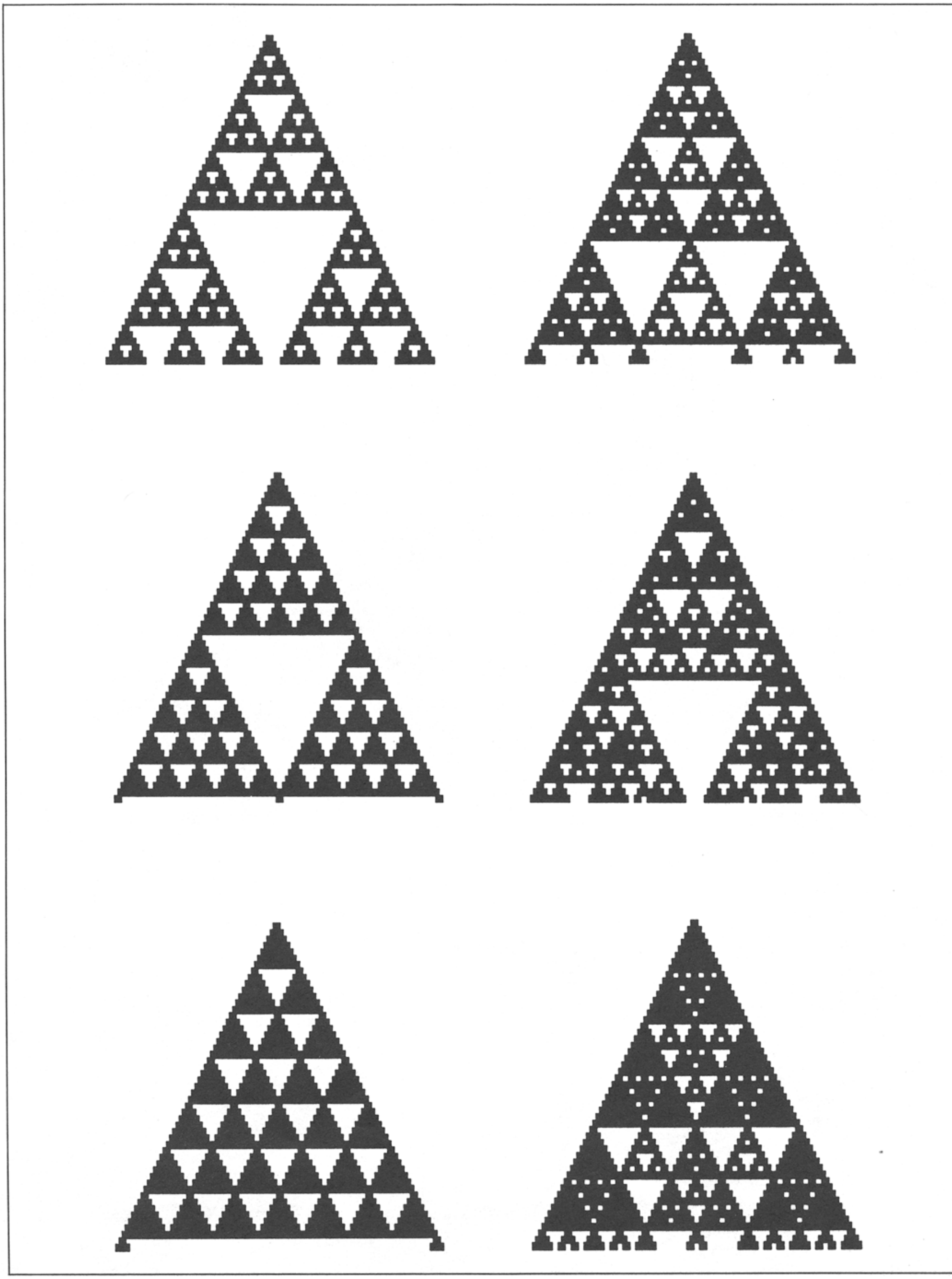


Figure 5. Pascal mod *k*.

The corporals can solve 2-disc Hanoi:

- "Private Bloggs: Move the smaller disc to pin 3 as in 1-disc Hanoi. Private Aggs: move the larger disc to pin 2. Private Bloggs: Move the smaller disc to pin 2 as in 1-disc Hanoi."

The sergeants can solve 3-disc Hanoi:

- "Corporal Cloggs: Move the top two discs to pin 3 as in 2-disc Hanoi. Private Aggs: Move the larger disc to pin 2. Corporal Cloggs: Move the top two discs to pin 2 as in 2-disc Hanoi."

The lieutenants can solve four-disc Hanoi:

- "Sergeant Doggs: Move the top three discs to pin 3 as in 3-disc Hanoi. Private Aggs: Move the larger disc to pin 2. Sergeant Doggs: Move the top three discs to pin 2 as in 3-disc Hanoi."

Private Aggs, of course, is a specialist. You should be able to write down how Captain Eggs, Major Foggs, Colonel Goggs, and n-star Generals Hoggs, and so forth, solve their own levels of the puzzle. To solve n-disc Hanoi for $n \geq 8$ takes an $(n-7)$-star General. I once tried to convince Yorkshire Television to enact this method for 5-disc Hanoi, using real soldiers. Unfortunately, they decided it would put undue strain on the audience's attention span.

That solves the puzzle: What else is there to say? Rather a lot, actually. You can pose many more subtle problems—such as how to move from any given position to any other in the most efficient manner—none of which are within the capabilities of my monstrous regiment.

Sometimes it helps to think geometrically. With any puzzle of this general type (moving objects, finite number of positions) we can associate a graph. Its nodes are the positions, its edges the moves between them. The graph H_n for n-disc Hanoi was introduced by Scorer et al. [29], rediscovered by Er [30], re-rediscovered by Lu [31], and $(\text{re})^3$discovered by me [15]. It is in the nature of the topic that most people working on it don't know the literature, so $(\text{re})^n$discovery for all $n \geq 4$ will occur with probability 1 in the long run.

What does H_n look like? For definiteness, consider H_3, which describes the positions and moves in three-disc Hanoi. To represent a position, number the three discs as 1, 2, and 3, with 1 being the smallest and 3 the largest. Number the pins 1, 2, and 3 from left to right. Suppose, for example, that disc 1 is on pin 2, disc 2 on pin 1, and disc 3 on pin 2. Then we have completely determined the position, because the rules imply that disc 3 must be *underneath* disc 1. Thus, we can encode this information in the sequence 212, the three digits, in turn, representing the pins for discs 1, 2, and 3. Therefore, each position in 3-disc Hanoi corresponds to a sequence of three digits, each being 1, 2, or 3. There are $3^3 = 27$ positions (because each disc can be on any pin, independently of the others).

What are the permitted moves? The smallest disc on a given pin must be at the top, so it corresponds to the *first* appearance of the number of that pin in the sequence. If we move that disc, we must move it to the top of the pile on some other pin, that is, we must change the number so that it becomes the first appearance of some other number. For example, in position 212, suppose we wish to move disc 1. This is on pin 2 and corresponds to the first occurrence of 2 in the sequence. Suppose we change this first 2 to 1. Then this is (trivially!) the first occurrence

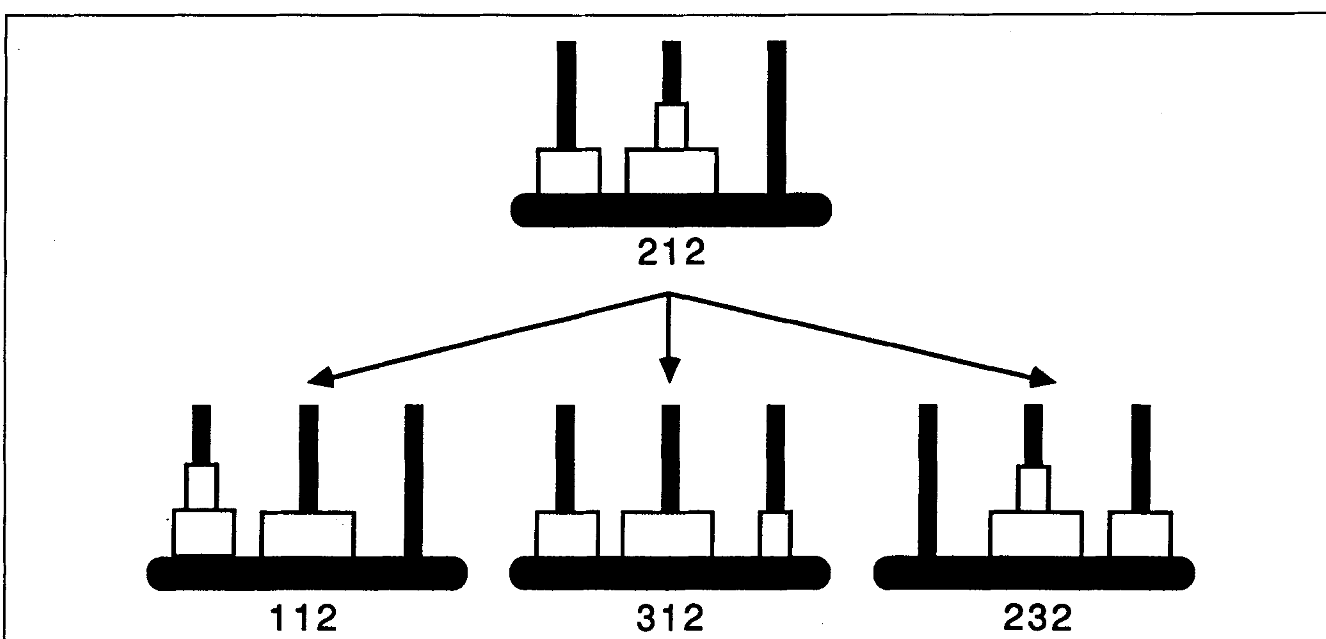


Figure 6. Moves in 3-disc Hanoi from position 212.

of the digit 1; so the move from 212 to 112 is legal; so is 212 to 312 because the first occurrence of 3 is in the first place in the sequence. We may also move disc 2, because the first occurrence of the symbol 1 is in the second place in the sequence. But we cannot change it to 2, because 2 already appears earlier, in the first place. A change to 3 is, however, legal. So we may change 212 to 232 (but not to 222). Finally, disc 3 cannot be moved, because the third digit in the sequence is a 2, and this is *not* the first occurrence of that digit.

To sum up: From position 212 we can make legal moves to 112, 312, and 232, and only these (Fig. 6). Proceeding in this way, we list all 27 positions and all possible moves by following the above rules. The graph H_3 can then be constructed: The result (after some rearrangement for elegance) is Figure 7.

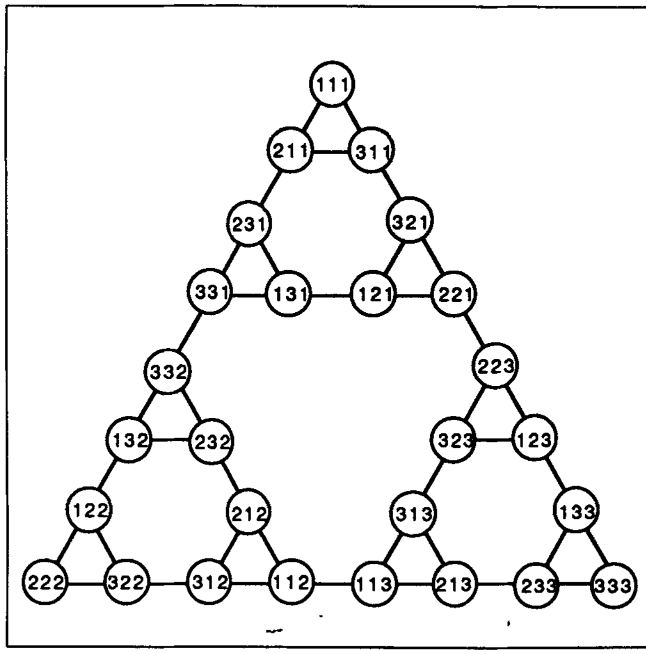


Figure 7. The graph H_3 of 3-disc Hanoi.

Something that pretty can't be coincidence! H_3 consists of three copies of a smaller graph, linked by three single edges to form a triangle. But each smaller graph, in turn, has a similar triple structure. Why does everything appear in threes, and why are the pieces linked in this manner?

In fact, the graph H_2 looks exactly like the top third of Figure 7. Even the labels on the vertices are the same, except that the final 1 is deleted. It is, of course, easy to see this without working out the graph again. You can play 2-disc Hanoi with three discs: just ignore disc 3. Suppose disc 3 stays on pin 1. Then we are playing 3-disc Hanoi but restricting attention to those three-digit sequences that end in 1, such as 131 or 221. These are precisely the sequences in the top third of the figure. Similarly, 3-disc Hanoi with disc 3 fixed on pin 2 corresponds to the lower left third, and 3-disc Hanoi with disc 3 fixed on pin 3 corresponds to the lower right third. It works for the same reason that the army method does.

This explains why we see three copies of the 2-disc Hanoi graph in the 3-disc graph. A little further thought shows that these three subgraphs are joined by just three single edges in the full puzzle. For, in order to join up the subgraphs, we must move disc 3. When can we do this? Only when one pin is empty, one contains disc 3, and the other contains all the rest! Then we can move disc 3 to the empty pin, creating an empty pin where it came from and leaving the other discs untouched. There are six such positions, and the possible moves join them in pairs.

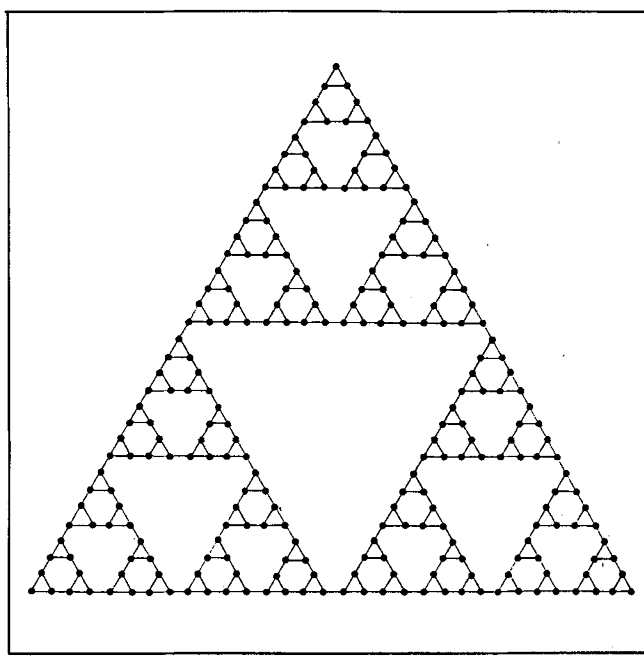


Figure 8. The graph H_5 of 5-disc Hanoi.

The same argument works for any number of discs, so H_{n+1} consists of three copies of H_n linked at the corners. For example, Figure 8 shows H_5. As the number of discs becomes larger and larger, the graph looks more and more like the Sierpiński gasket.

You can use the graph to answer all sorts of questions about the puzzle. For example, it follows inductively that the graph is connected, so you can always move from any position to any other. The minimum path from the usual starting position to the usual finishing position runs straight along one edge of the graph, so (again by induction) has length $2^n - 1$. This result, long assumed in the form "the largest disc moves only once," was first proved by Wood [32].

The resemblance of H_n to the Sierpiński gasket has a curious application. Not long after Ref. 15 appeared, I attended the International Congress of Mathematicians in Kyoto, and a German mathematician named Andreas Hinz introduced himself. He had been trying to calculate the average distance between two points in a Sierpiński gasket of unit side, encountered difficulties, and asked two experts. Here's what they said.

The Considered View of Expert 1:
It's very difficult.

The Considered View of Expert 2:
It's trivial, and the answer is 8/15.

Here is Expert 2's proof. The idea is first to find the average distance a to some particular corner, and then to use that to find the average distance d between two arbitrary points.

From Figure 9, it follows immediately that

$$a = \frac{1}{3}\left[\frac{a}{2} + 2\left(\frac{a}{2} + \frac{1}{2}\right)\right],$$

from which $a = \frac{2}{3}$. Now consider two points: They are either in the same subtriangle, as in Figure 10(a), or not, as in Figure 10(b). The respective probabilities are $\frac{1}{3}$ and $\frac{2}{3}$. In the latter case, the shortest path between them goes through the common vertex. Therefore,

$$d = \frac{1}{3}\frac{d}{2} + \frac{2}{3}\left(2\frac{a}{2}\right),$$

from which $d = 8/15$.

Happy? You shouldn't be. Expert 2's proof is fallacious. In the second case, the shortest path sometimes goes through two connecting vertices. An example in H_3, pointed out by Lu [31], is shown in Figure 11. The identical mistake—assuming that "the largest disc moves at

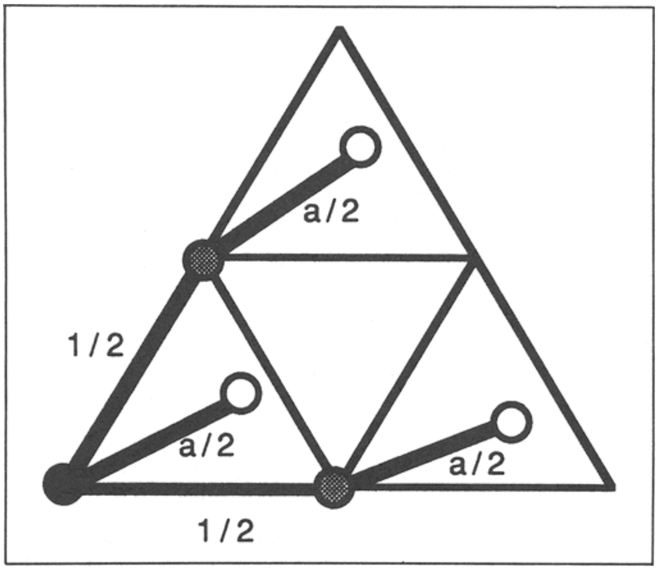


Figure 9. Proof that $\alpha = \frac{2}{3}$.

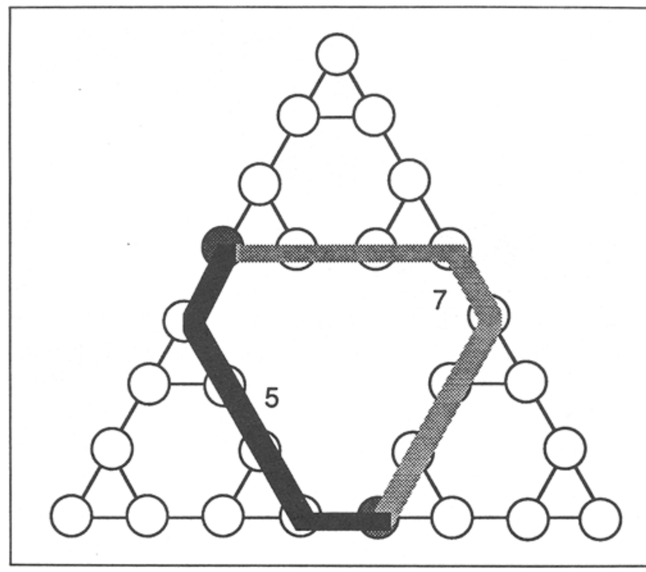


Figure 11. Lu's counterexample.

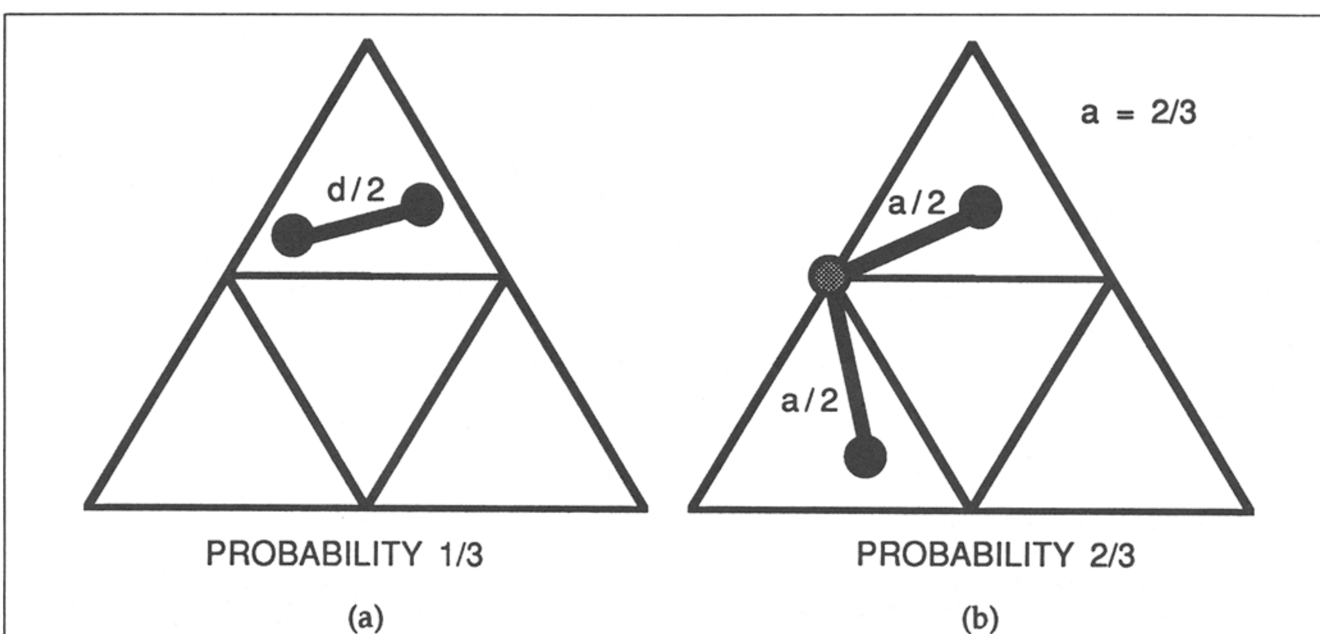


Figure 10. Proof (?) that d = 8/15.

most once" when moving between *any* two positions by the most efficient route — occurs many times in the literature on the Tower of Hanoi. See, for example, Refs. 30 and 33. Psychologists have used the tower of Hanoi as an experimental testbed for human decision-making, for example, Ref. 34; and on occasion the same mistake has crept in — for example, in Ref. 35.

Unfortunately, even when the nature of the fallacy is grasped, it seems hard to incorporate this third case into the analysis in the same manner as Figures 9 and 10, and the story becomes far more complicated. Hinz [36, 37], and independently Chan [38], give a formula for the average number of moves between positions in the Tower of Hanoi. In fact, the *total* number of moves (using shortest paths) between all possible pairs of positions is

$$\frac{466}{885}18^n - \frac{1}{3}9^n - \frac{3}{5}3^n + \left(\frac{12}{59} + \frac{18}{1003}\sqrt{17}\right)\left[\frac{1}{2}\left(5+\sqrt{17}\right)\right]^n + \left(\frac{12}{59} - \frac{18}{1003}\sqrt{17}\right)\left[\frac{1}{2}\left(5-\sqrt{17}\right)\right]^n.$$

Thus, the average distance between two positions is asymptotic to $(466/885)2^n$.

Hinz hadn't realised there was any connection with the Sierpiński gasket; but having seen Ref. 15, he realised that the limit as $n \to \infty$ of his result for n-disc Hanoi proves that the average distance between two points in a unit Sierpiński gasket is 466/885 precisely (just normalize to make the diameter of the graph 1, by dividing by $2^n - 1$). This is some 2% smaller than the value suggested by Expert 2. Who says recreational mathematics has no serious payoff? At the moment, this approach via the tower of Hanoi is the only known method for finding the answer.

For the statistically minded, Hinz also proved that the variance of the distance between two random points is precisely 904 808 318/14 448 151 575.

Encounter 4: Barnsley's Encounter with Sierpiński's Gasket (Michael Barnsley, 1988)

> A fractal set generally contains infinitely many points whose organization is so complicated that it is not possible to describe the set by specifying directly where each point lies. Instead the set may be defined by "the relations between the pieces". It is rather like describing the solar system by quoting the law of gravitation and stating the initial conditions. Everything follows from that. It appears always better to describe in terms of relationships.
>
> Michael Barnsley
> *Fractals Everywhere*

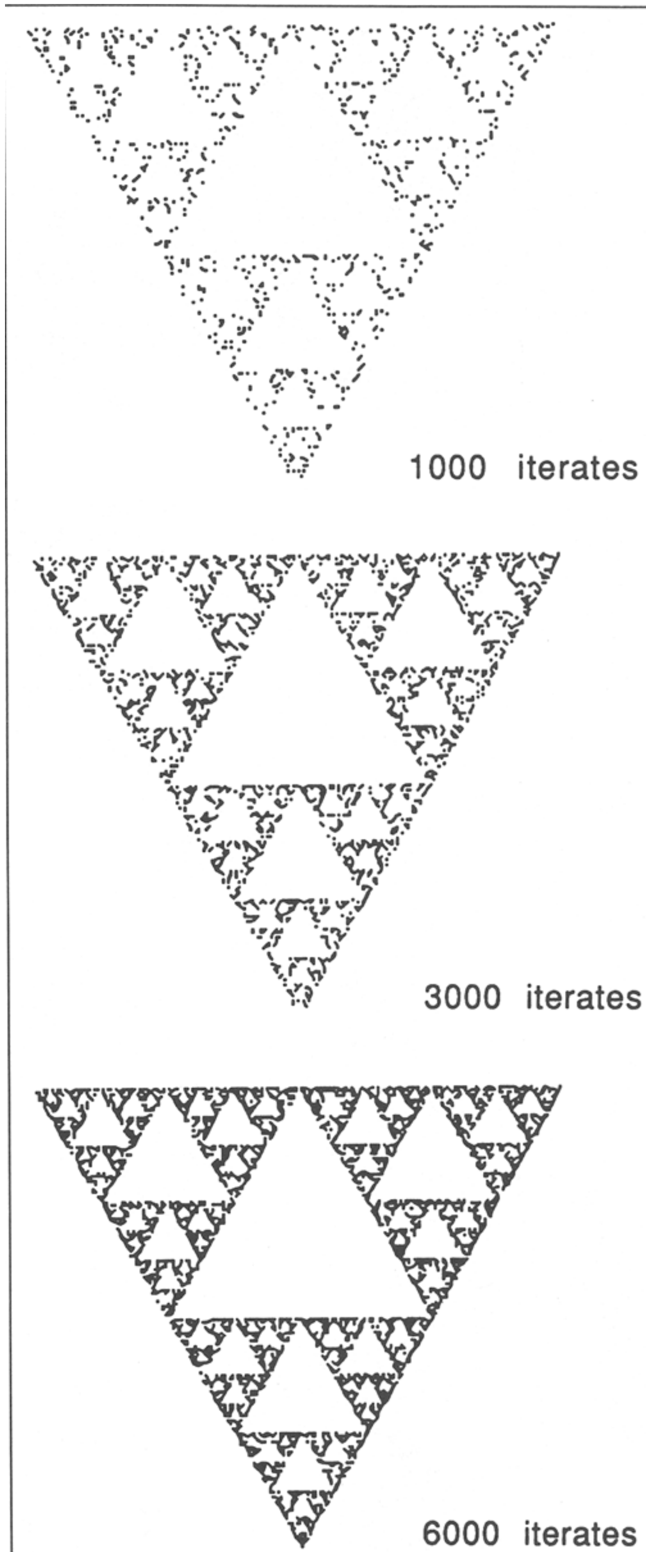


Figure 12. The chaos game after 1000, 3000, and 6000 iterations.

In his celebrated textbook [39], Michael Barnsley introduces the *chaos game*. Mark three points in the plane, say at the vertices of an equilateral triangle. Obtain a three-sided coin for which heads, tails, and edge have the same probability, namely $\frac{1}{3}$. Label the three vertices of the triangle correspondingly. Now play the following game. Start with a randomly chosen point x_0 in the plane. Toss the coin and move the point halfway towards the corresponding vertex, getting a new point x_1. Repeat this procedure, always generating x_{n+1} from x_n by tossing the coin and moving x_n halfway towards the appropriate vertex. What do you see?

You might expect the result to be some uniform cloud of points in the plane. But having read the title of this article, you might suspect that this is not so, and a shrewd guess would be ... well, Figure 12 shows that you're right. You get a Sierpiński gasket, which becomes more and more sharply defined the more iterates you use. (The first 50 iterates are omitted, for these constitute dynamical "transients" that spoil the perfection of the figure.)

This seems a very odd shape to generate by a random procedure, although we shall see shortly that it is entirely natural. Barnsley defines a generalization: an *iterated function system* or IFS. This is a finite set of affine maps from the plane to itself. Affine maps are specified by six parameters:

$$F(x, y) = (ax + by + e,\ cx + dy + f).$$

The *contractivity factor* of such a map is defined to be

$$s = |ad - bc|.$$

If $s < 1$, then F shrinks areas by a factor s. (If $s > 1$, it expands them by a factor s.) Suppose that $\mathcal{F} = \{F_n\}$ is an IFS, where F_n has contractivity factor s_n, and suppose that $s_n < 1$ for all n. Define a set A to be *invariant* under $\mathcal{F}$ if

$$A = \bigcup_{i=1}^{n} F_i(A).$$

For example, suppose maps $F_i : \mathbb{R} \to \mathbb{R}, (i = 1, 2)$ are defined by

$$F_1(x) = \frac{x}{3},$$

$$F_2(x) = \frac{x+2}{3}.$$

Then the standard middle-third Cantor set is invariant under $\mathcal{F}$.

THEOREM 3: *Under the above conditions (in particular with all $s_i < 1$), there exists a unique nonempty invariant set for $\mathcal{F}$.*

Proof:. Let $\mathcal{H}$ be the set of all subsets of $\mathbb{R}^2$ with the Hausdorff metric. Then $\mathcal{F}$ defines a contraction mapping on $\mathcal{H}$ with contractivity factor $s = \max(s_n)$. This has a unique fixed point. See Ref. 11 or 39 for details.

In view of the proof, we denote the invariant set by Fix($\mathcal{F}$). Typically, Fix($\mathcal{F}$) is a fractal; and by definition, it is *self-affine*, that is, the union of affine copies of itself.

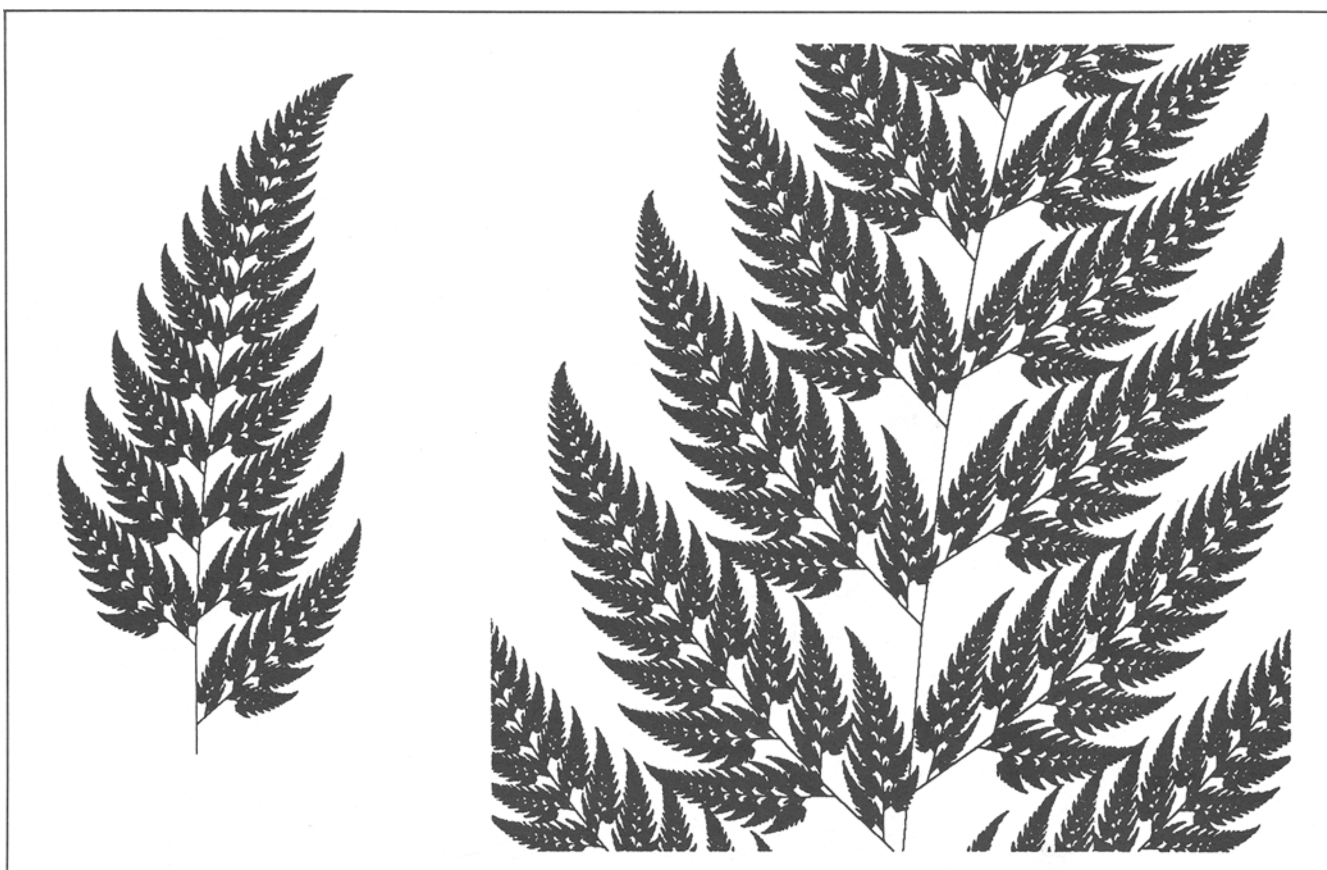


Figure 13. The black spleenwort fern.

With certain technical hypotheses (the images of Fix($\mathcal{F}$) under the f_i should not overlap "too much"), the fractal dimension d of Fix($\mathcal{F}$) is the unique value such that

$$s_1^d + \cdots + s_n^d = 1.$$

The Sierpiński gasket is obviously equal to Fix($\mathcal{F}$) when $\mathcal{F} = \{F_1, F_2, F_3\}$ and

$$F_1(x, y) = \left(\frac{x}{2}, \frac{y}{2}\right),$$
$$F_2(x, y) = \left(\frac{x+1}{2}, \frac{y}{2}\right),$$
$$F_3(x, y) = \left(\frac{x}{2} + \frac{1}{4}, \frac{y}{2} + \frac{\sqrt{3}}{4}\right).$$

These three transformations correspond precisely to "move halfway towards the vertices of an equilateral triangle." It is now clear that the point set defined by playing the chaos game is almost surely a very close approximation to the invariant set for the corresponding IFS, and that's why we see the Sierpiński gasket. Barnsley [39] and Falconer [11] contain the proof, plus extensive generalizations.

This observation has a curious and potentially important consequence. Suppose you want to send a colleague a picture of the Sierpiński gasket. You could draw one and run it through a fax machine. This will scan the page in raster fashion and send several hundred thousand numbers along the telephone lines, from which another fax machine can reconstruct the picture. On the other hand, if the recipient has a computer that can play the chaos game, all you need do is send the numbers that define the iterated function system — six per affine map, 18 altogether. This represents a considerable saving in data to be transmitted.

As it happens, a great many natural objects have fractal structure, and so can be given a "compressed" description as invariant sets of iterated function systems. So can many *non*fractals, such as a solid square (play the chaos game with four points and a four-sided coin). The traditional example is the black spleenwort fern (Figure 13). Although you might not often want to transmit a black spleenwort fern, most pictures are made up out of pieces that have the same kind of fractal structure, and an extension of the notion of an IFS can be applied to them: see Refs. 40–43 and 48. Initially the method was greeted with some skepticism, but it is a perfectly practical one: see Refs. 44 and 48. Commercial software to implement the process is available — at commercial prices. The whole story suggests a new view of complexity — or at least, encourages a view more akin to algorithmic information theory [45] — namely, it is the complexity of the *process* that produces an object that is important, not the apparent complexity of the object itself. *Pre*scription, not *de*scription, is the key. It is a point of view with substantial implications for evolutionary and developmental biology; see Ref. 46.

Encounters, Encounters, ...

The variety of situations in which we encounter the Sierpiński gasket is considerable. Indeed, there are many *more* such encounters scattered throughout the mathematical literature: One I was told of recently is the graph of positions for hexaflexagons. These ingenious mathematical toys are described in Ref. 47.

Why do we meet the gasket in so many different places? The underlying theme in all four encounters is recursion: The Sierpiński gasket is the incarnation of recursive geometry. Indeed, it is probably the simplest genuinely two-dimensional recursive geometric object, just as the Cantor set is the simplest one-dimensional one. (I mean that the gasket lives in the plane — I'm not talking about its *fractal* dimension.)

Even given this rationalisation, it is still rather odd that it appears in so many guises.

References

1. Benoît Mandelbrot, *The Fractal Geometry of Nature*, San Francisco: Freeman (1977).
2. W. Sierpiński, *Oeuvres Choisies* (2 vols.), Warsaw: Pańswowe Wydwnicto Naukowe, (1975).
3. Krzysztof Ciesielski, Lost legends of Lvov I: The Scottish Café, *Mathematical Intelligencer* 9(4) (1987), 36–37.
4. Krzysztof Ciesielski, Lost legends of Lvov II: Banach's grave, *Mathematical Intelligencer* 10(1) (1988), 50–51.
5. W. Sierpiński, Sur une courbe dont tout point est un point de ramification, *Compt. Rendus Acad. Sci. Paris* 160 (1915), 302–305.
6. W. Sierpiński, On a curve every point of which is a point of ramification, *Prace Mat. Fiz.* 27 (1916), 77–86 [Polish].
7. W. Sierpiński, Sur une nouvelle courbe continue qui remplit toute une aire plane, *Bull Int. Acad. Sci. Cracovie A* (1912), 462–478.
8. W. Sierpiński, On a Cantorian curve which contains a bijective and continuous image of any given curve, *Mat. Sb.* 30 (1916), 267–287 [Russian].
9. W. Sierpiński, Un exemple élémentaire d'une fonction croissante qui a presque partout une derivée nulle, *Giornale Mat. Battaglini (3)* 6 (1916), 314–334.
10. W. Sierpiński, On a reversible function whose image is dense in the plane, *Wektor* 3 (1914), 289–291 [Polish].
11. Kenneth Falconer, *Fractal Geometry*, New York: Wiley (1990).
12. L. E. Dickson, *History of the Theory of Numbers*, Vol. 1, New York: Chelsea (1952).
13. P. Hilton and J. Pedersen, Extending the binomial coefficients to preserve symmetry and pattern, *Computers Math. Appl.* 17 (1989), 89–102; reprinted in *Symmetry 2 — Unifying Human Understanding* (I. Hargittai ed.), Oxford: Pergamon Press (1989).
14. Marta Sved, Divisibility — with visibility, *Mathematical Intelligencer* 10(2) (1988), 56–64.
15. Ian Stewart, Le lion, le lama et la laitue, *Pour la Science* 142 (1989), 102–107.
16. Ian Stewart, *Game, Set, and Math*, Oxford: Basil Blackwell, (1989) [reprint: Harmondsworth: Penguin Books (1991)].
17. G. S. Kazandzidis, Congruences on the binomial coefficients, *Bull Soc. Math. Grèce (NS)* 9 (1968), 1–12.

18. David Singmaster, Notes on binomial coefficients III—Any integer divides almost all binomial coefficients, *J. London Math. Soc. (2)* 8 (1974), 555–560.
19. J. W. L. Glaisher, On the residue with respect to p^{n+1} of a binomial-theorem coefficient divisible by p^n, *Quart. J. Pure Appl. Math.* 30 (1899), 349–360.
20. E. E. Kummer, Über die Ergänzungssätze zu den allgemeinen Reciprocitäsgesetzen, *J. Reine Angew. Math.* 44 (1852), 93–146.
21. David Singmaster, Notes on binomial coefficients I—A generalization of Lucas' congruence, *J. London Math. Soc. (2)* 8 (1974), 545–548.
22. A. W. F. Edwards, Patterns and primes in Bernoulli's triangle, *Math. Spectrum* 23 (1991), 105–109.
23. Siegfried Rösch, Expedition in Unerforschtes Zahlenland, *Neues Universarum* 79 (1962), 93–98.
24. Siegfried Rösch, Neues vom Pascal-Dreieck, *Bild der Wiss.* (Sept. 1965), 758–762.
25. David Singmaster, Notes on binomial coefficients II—The least n such that p^e divides an r-nomial coefficient of rank n, *J. London Math. Soc. (2)* 8 (1974), 549–554.
26. David Singmaster, Divisibility of binomial and multinomial coefficients by primes and prime powers, *A Collection of Manuscripts Related to the Fibonacci Sequence, 18th Anniversary Volume of the Fibonacci Association* (1980), 98–113.
27. N. Claus [= Edouard Lucas] La tour d'Hanoï, jeu de calcul, *Sci. Nature* 1 (1884), 127–128.
28. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete Mathematics*, New York: Addison-Wesley (1989).
29. R. S. Scorer, P. M. Grundy, and C. A. B. Smith, Some binary games, *Math. Gaz.* 28 (1944), 96–103.
30. M. C. Er, A general algorithm for finding a shortest path between two n-configurations, *Inform. Sci.* 42 (1987), 137–141.
31. Lu Xuemiao, Towers of Hanoi graphs, *Int. J. Comput. Math.* 19 (1986), 23–38.
32. D. Wood, The towers of Brahma and Hanoi revisited, *J. Recreational Math.* 14 (1981–82), 17–24.
33. D. Wood, Adjudicating a towers of Hanoi contest, *J. Recreational Math.* 14 (1981–82), 199–207.
34. F. Klix, J. Neumann, A. Seeber, and H. Sydow, Die algorithmische Beschreibung des Lösungsprinzips einer Denkanforderung, *Z. Psychol.* 168 (1963), 123–141.
35. H. Sydow, Zur metrischen Erfassung von subjektiven Problemzuständen und zu deren Veränderung im Denkenprozeß I, *Z. Psychol.* 177 (1970), 145–198.
36. Andreas M. Hinz, The tower of Hanoi, *L'Enseignement Math.* 35 (1989), 289–321.
37. Andreas M. Hinz, Shortest path between regular states of the tower of Hanoi, *Inform. Sci.*, to appear.
38. Chan Hat-Tung, A statistical analysis of the towers of Hanoi problem, *Int. J. Comput. Math.* 28 (1989), 57–65.
39. Michael Barnsley, *Fractals Everywhere*, Boston: Academic Press (1993).
40. Michael Barnsley, A Better way to compress images, *BYTE*, January 1988.
41. Michael Barnsley and A. E. Jacquin, Application of recurrent iterated function systems to images, *SPIE* 1001 (1988), 122–131.
42. A. E. Jacquin, *A Fractal Theory of Iterated Markov Operators with Applications to Digital Image Coding*, Ph.D. Thesis, Georgia Institute of Technology (1989).
43. A. E. Jacquin, A novel fractal block-coding technique for digital images, *ICASSP '90* (1990).
44. Jon Waite and Mark Beaumont, An introduction to block based fractal image coding, preprint, British Telecom Research Station, Ipswich (1991).
45. Gregory J. Chaitin, *Algorithmic Information Theory*, Cambridge: Cambridge University Press, (1987).
46. Jack Cohen and Ian Stewart, The information in your hand, *Mathematical Intelligencer* 13(3) (1991), 12–15.
47. Martin Gardner, *Mathematical Puzzles and Diversions from Scientific American*, London: Bell (1961).
48. Michael F. Barnsley and Lyman P. Hurd, *Fractal Image Compression*, Wellesley MA: A. K. Peters (1993).

Mathematics Institute
University of Warwick
Coventry CV4 7AL
United Kingdom

Williams, continued from p. 34.

Whether Daniel Gumb proved the theorem for himself or was simply so impressed by the simplicity of the proof that he decided to record it in stone is not clear. It is still easily recognisable despite 250 years of weathering. This proof is, of course, fairly well known. For example, it is mentioned in Ref. 3 as "another proof."

The cave is not very easy to find. Figure 2 shows it in relation to the Cheesewring (and the author) with the diagram easily recognisable on the roof. To those visiting this part of Cornwall, and possibly considering the astronomical significance of the Hurlers, the location of Daniel Gumb's cave and proof is left as an exercise.

References

1. S. Baring-Gould, *Cornish Characters and Strange Events*, Bodley Head (1909).
2. W. H. Paynter, Daniel Gumb, The Cornish cave-man mathematician, *Old Cornwall: Journal of the Federation of Old Cornwall Societies*, II (4) (1932).
3. C. Godfrey and A. W. Siddons, *Elementary Geometry*, 4th ed., Cambridge: Cambridge University Press (1962).

Faculty of Mathematical Studies
University of Southampton
Southampton SO17 1BJ, England

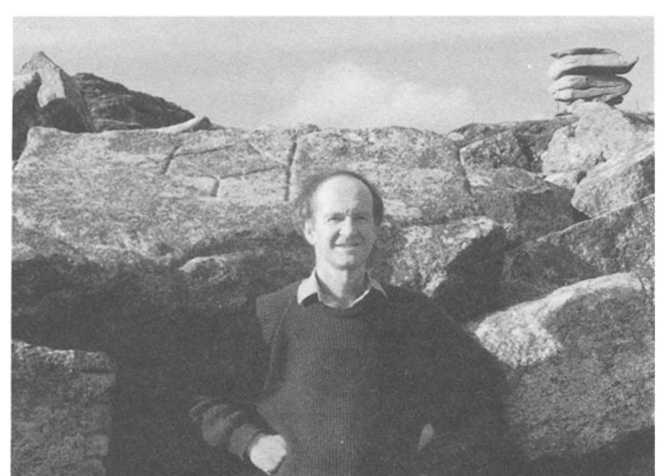


Figure 2