The 'Core' Concept and the Mathematical Mind

Chris King
Mathematics Department
University of Auckland

Abstract:
Abstract pure mathematics is often seen as an 'inverted pyramid', in which algebra and analysis stand at the focal point, without which students could not possibly have a firm grounding for graduate studies. This paper examines a variety of evidence from brain studies of mathematical cognition, from mathematics in early child development, from studies of the gatherer-hunter mind, from a variety of puzzles, games and other human activities, from theories emerging from physical cosmology, and from burgeoning mathematical resources on the internet that suggest, to the contrary, that mathematics is a cultural language more akin to a maze than a focally-based hierarchy; that topology, geometry and dynamics are fundamental to the human mathematical mind; and that an exclusive focus on algebra and analysis may rather explain an increasing rift between modern mathematics and the 'real world' of the human population.

1: Introduction

The idea of a fundamental duality between analysis as 'the science of continuity' and algebra as the 'art of discrete operations', forming the two pillars of conceptual mathematics raises a number of deep questions. At first sight it provides an appealing analogy to the idea of complementary left and right hemisphere brain activity, spanning order and chaos, typified by the specialization for descriptive language in the left hemisphere, particularly in males, and creative poetry and music in the right. However, rather than the 'Adam and Eve' primality of algebra and classical analysis, it is the disquieting entanglements of topology, like a jilted Lilith, which have been seen as ultimately standing in natural complementation to the 'divine' order of algebraic structures in the split brain. Neither does the algebra-analysis partnership stand on a solid axiomatic foundation, undercut, like the house built on sand, by the common, yet often discarded, bedrock of mathematical logic and set and category theory. However, the argument for algebra and analysis is not built so much on axiomatic primality as perceived 'heuristic necessity' - the idea that, without a firm grounding in these two abstract areas, a prospective graduate will lack essential knowledge and skills. Is this so, or is it a classical misconception?

Figure 1.1: Hmong appliqué illustrates a cultural tradition intuitively based on topology which illustrates the nested Jordan curve two colouring theorem of section 2 used to effect in early child maths readers. The left figure has 7 blue oceans and 19 white islands, the right has 7 red oceans and 11 white islands.

At stake is a model of 'reality' of cosmological proportions. Central to the 'belief system' of modern pure mathematics is the notion that mathematics is itself a cosmology of abstract archetypes, exemplified by Karl Popper's 'trinity'[1] in which the brain in the objective physical universe and the subjective conscious mind are complemented by a third 'dark force' - a 'cosmos' of conceivable abstract structures waiting to be 'discovered' by theoretical research. Nevertheless many of these structures, stemming from the very notion of an infinitesimal point singularity, abhorred by the wave-particle complementarity of the quantum universe, already have a slightly archaic classical tarnish in a world composed of entangled quantum waves, so are these archetypes really 'universal'? Comparison with cosmological theories of everything provides suggestive clues.

In contrast to the cosmological view of mathematics, is the 'cultural' view that mathematics is a human, relatively 'culture fair' language, crafted to enable the development of concepts, the discussion of hypotheses and the proof of theorems that would otherwise be impossible without it. Clearly mathematics, and with it the areas of algebra and analysis, are developed as linguistic branches of mathematics, containing their own definitions and propositions, which can logically be proved using the language of the objects, functions and operators in each of the fields. But if mathematics is primarily a human-invented language, the question arises as to whether a given branch of it has any unique claim to fundamentality and whether another quite different linguistic description might also lay equal claim to the territory, and whether, in the near future, or at the restaurant at the end of the universe, the claims algebra and analysis might make, even to heuristic fundamentality might appear quaint, myopic and archaic artifacts of a culture doomed to attrition by its own lack of adaptability to the circumstances around it.

We can see signs of this dilemma in the relation between analysis and topology, where competing ideas of continuity, which in analysis are based on measurement and an almost irresolvable strategic standoff between epsilon and delta, meet their nemesis in the fall of the metric space empire to the superior topological concept of continuity, based on the apparently elusive notion of 'openness' opening the Pandora's box of all the tortuous knotted worms of continuity and connectedness and the many worm holes back to geometry writhing in the topological category and in the futility of measurement in a landscape permeated by non-measurable sets. While analytic epsilon-delta is still considered the foundation concept, this is done with the knowledge that, lurking in the back closet, hopefully kicked upstairs to the graduate program, is a more fearsome concept, coming right out of left field, or more likely the enigmatic right cortex, that lays waste to all ideas of measure.

In presenting these ideas to a human population there is a contest of credibility gaps. While mathematicians rail at the futility of the naïve concept of limit in first year courses, performed merely by making spot checks at a few neighbouring points getting closer to the value concerned, the epsilon-delta game has proven to be a stand-off so counter intuitive that it has, by degrees, been shuffled into its own cubby hole in specialized theoretical courses, only to reappear as the required phoenix of a 'core' major. By contrast, the topological idea of continuity and connectedness, while remaining mathematically 'fringe' does have immense immediate and practical appeal to the human consciousness and imagination. The alternative naïve idea of continuity - that which one can execute while drawing without lifting a pencil off the paper - goes right to the heart of topological ideas of continuity, and path connectedness, and is the foundation of both algebraic homotopy and manifold topology and is thus the basis of the Poincaré conjecture.

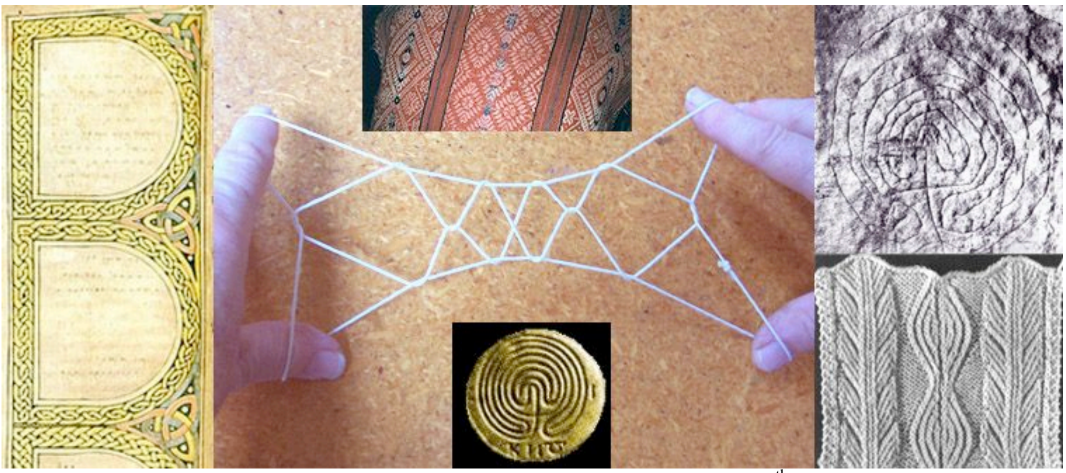


Figure 1.2: Left: Oldest example of Celtic frieze and knots from Durham Cathedral (7th cent). Center: Maori string figure Tahitinui[2], has been found to be topologically identical to the Native American Osage diamonds, illustrating the deep cross-cultural awareness of topology, as an intellectual skill associated with textiles, nets and knot-making. Chinese children also play string games. Top right: Neolithic meander maze resembles that on the coinage of Knossos Crete (lower inset) suggestive of the Minotaur's Labyrinth. Dura-Europos 'Arran'-style knitting (300BC-256AD)[3] and Andean woven textile (upper inset) illustrate the intrinsic topological complexity of all textiles made by weaving, knitting netting other methods.

Topological continuity is also a directly perceivable conservation concept that, in contrast to its relegation to the ivory towers of graduate mathematics, is directly understandable by children still learning to read, as successfully purveyed in some of the best mathematics education books for young children. For this reason topology courses have been a favorite with training teachers and those returning for additional mathematical education, for the very reason that they provide access to a fundamental form of mathematical reasoning wholly neglected in our emphasis firstly on statistics as an alternative to mathematics and secondly on algebra and calculus to the exclusion of topology and geometry in secondary mathematics and undergraduate core papers at the tertiary level, now prospectively becoming filtered by prerequisites even at the graduate level.

Topological continuity also has very plausible foundations in gatherer-hunter skills, honed over evolutionary time scales, in terms of tracing tortuous connected paths in the wilderness, between raging torrents, lairs of lions and dangerous precipices, to apply one's skills of geometry and dynamics to outwit the prey and retrace one's steps through the topological maze to finally bring home the meat, to trade for sexual favours in the 'amazing race' of genetic survival. With the foundation of Neolithic cultures, weaving and net-making added tremendously to the topological complexity of the human imagination, leading later to knitting and ultimately, with the flying jenny, to the industrial revolution.

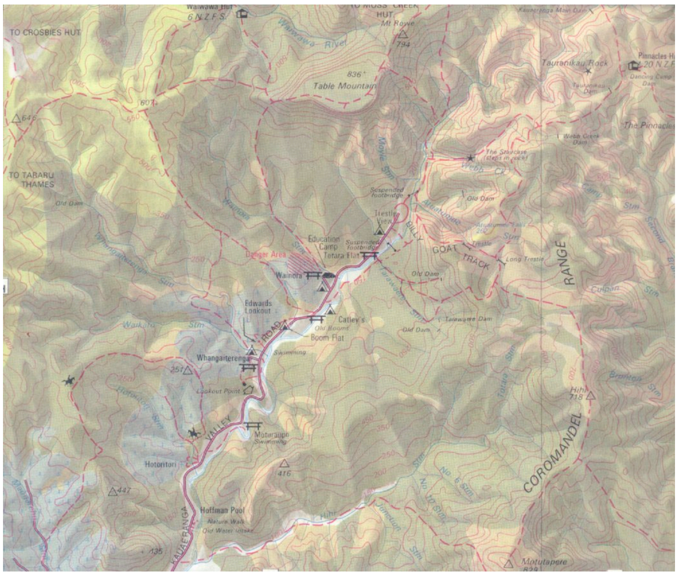


Figure 1.3: Map of the Kaueranga Valley illustrates how wilderness terrains give rise to a complex natural topology of path-connected routes (tracks and trails) partially obstructed by other topologically–connected features, such as water courses, ridges and escarpments, and patches of dense forest with steep impenetrable valleys. Learning the path-connected topologies of the wilderness is a gatherer-hunter task essential for survival.

To support this topological and discrete dynamic thesis, a stimulating investigation has been made of a variety of human puzzles and games, as a cultural 'imprint' of human mathematical imagination, including obviously topological wire, string and loop puzzles, Rubik cubes and their algebraic and geometrical variants, including irregular forms such as "Square-1", Erno Rubik's 3-ring puzzle, geometrical tiling puzzles, including the Soma cube and variants such as the "Lonpos Pyramid" and "Happy Cube", peg solitaire, the five peg puzzle, numeric puzzles such as magic squares, squaring the square and Sudoku, logical puzzles such as the 'Einstein' Zebra puzzle, and rule-guessing puzzles such as "Petals Around the Rose". In addition are included games of topological strategy, such as Go and the more recent fractal geometry variant "Blokus", based on pentaminoes. and topological tiling games such as Tantrix and Trax.

A disproportionate number of puzzles and games have a geometrical and/or topological basis, which belies the abstractness attributed to topology by mathematicians, and emphasizes the central place geometry and topology have in the human imagination. All puzzles and games, whether they are geometrical, logical or conceptual are fundamentally topological in nature, because they possess a path-connected route from an initial state to a final solution, or end-game. A puzzle is solvable if such a path-connected route exists, and humans, whether they are solving a logical puzzle, or a geometrical one, do so by having an intuitive, or deductive idea of territory within the 'state-space' of possibilities, and how to construct a connected path from beginning to end along the connected graph of nodes in this space, in which contingencies 'if … then' correspond to branches. All proving of mathematical theorems likewise has a basis in such 'topological logic'. In this sense a link is made between the ostensibly continuous subject of topology and the discrete area of graph theory, but regardless of this, a human is using, even in a highly abstract situation, the same topological skills that humans have relied upon for survival in the wild over millennia. In some puzzles and games the state space is surprisingly small, because the node transitions are complex, but in others the state space consists of a huge graph containing up to 10^{50} vertices, comparable even with the 10^{500} 'multiverses' now suggested by string theory. Nevertheless a human successfully solves these puzzles, which closely reflect the super-exponentiating np-complete $\frac{(n-1)!}{2}$ traveling salesman[4] contingencies of n real connected paths in the world at large, by developing a sense of connectivity within the territory.

An investigation has also been made of the current state of research into mathematical reasoning in brain studies, using functional magnetic resonance imaging and the electroencephalogram. To a certain extent these are limited by the relatively simplistic arithmetic and mental rotation tasks frequently assigned to subjects confined in a narrow noisy detector, but they do provide an informative view of the broader biological basis of mathematical reasoning lying beyond the axiomatic paradigm and show in a startling way how language and culture can affect mathematical processing in the brain, supporting the 'cultural' view that mathematics is a cultural adaption.

Models of human cognition, imagination and creativity go even further and are sometimes likened, not just to a generalized language but more to a Swiss army knife of evolutionary skills cobbled together, not from a universal foundation but in the subtle adaptive ingenuity of parallel genetic algorithms. In terms of brain science, another version of this fortuitous convergence of parallels is the idea that numerical reasoning is based on a convergence between three neural 'sensory' codes of numerical processing, involving analog magnitude, auditory verbal, and visual Arabic codes of representation. Unlike mental rotation and targeting, and the need to trace topological paths in the wilderness, numeracy itself has been found not to be a human universal, with the Amazonian Pirahã becoming famous for their lack of linguistic categories more specific than 'few', resulting in a cultural inability to distinguish even small numbers of objects one could count on one's fingers. In a way this is unsurprising since even 'advanced' cultures have trouble dealing with a digit span of over seven.

Cosmology also carries with it fundamental aspects of both topology and fractal dynamics, which we shall investigate to complete the perspective. To celebrate the new idea of intrinsic duality between theories of everything, in which supposed fundamental particles, such as the quark and lepton, exchange roles with perceived composites such as the magnetic monopole, in the next section we will make a case based on early learning mathematical texts, that mathematics could as well be founded, in human evolution and development, on twin concepts of the topology of connectedness and discrete dynamics.

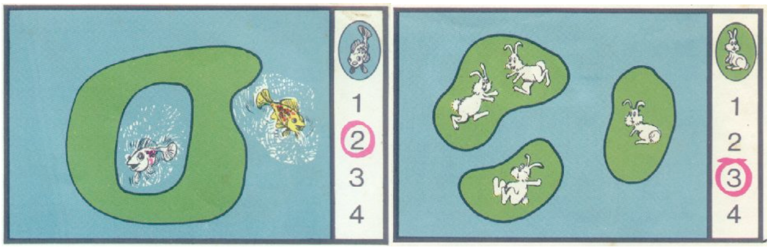
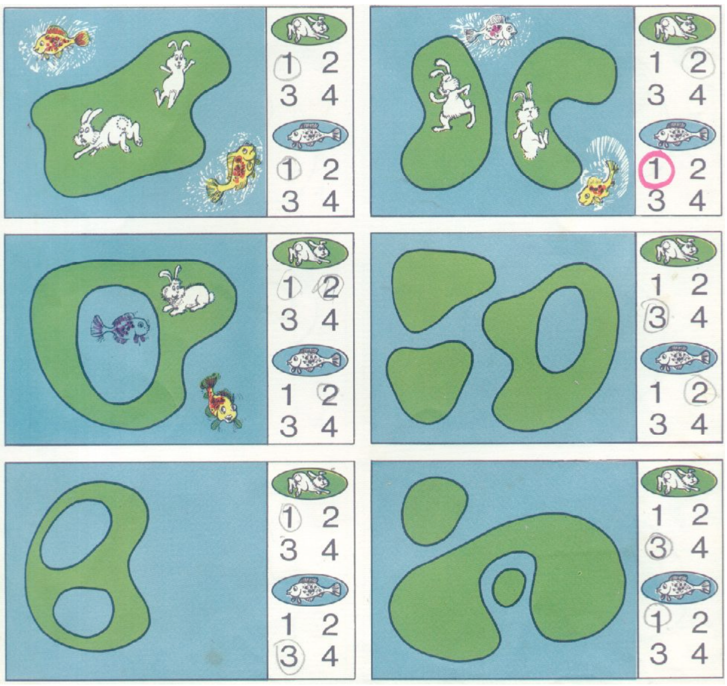


Figure 2.1: Clues for the ZOO islands-lakes puzzles.

2: Landmarks from Early Childhood and the Noosphere

Two mathematical education books, which made a lifetime impression on my own children, for their stimulating mathematical imaginativeness, are the Zoo series[5] and Inner Ring Maths[6]. These contain a variety of mathematical puzzles and problems involving sets, classification, sequence completion simple arithmetic numeracy, rotational and reflective symmetries and geometrical shapes.

Figure 2.2: The initial topological puzzles have both animal clues and a two colour coding of the regions.



(a) Topology of Lakes and Islands

The ZOO books are set out with a minimum of language statements so that even a child who cannot read can identify the tasks visually and solve the mathematical problems. Most interesting of these to my children were the two volumes dealing with topology and connected pathways.
The reader is first given the non-verbal clues shown in figure 2.1, which give instances indicating the problem is not how many rabbits or fish, but how many (possibly nested) islands and lakes there are. The reader is then led on an adventure into increasingly abstract and complex instances of the problem as shown in figures 2.2 and 2.3.

In a mathematical sense, the twin count of islands and lakes constitutes a topological invariant of a set of non-intersecting Jordan curves in the plane, each of which divides the region containing it in two in a way which always permits a two-colour coding of the whole rectangle. If the enveloping region is considered to be the ocean as is the case for (a flat) Earth, this results in an unambiguous classification into land and water, even when the only clues are the curves themselves, as in figure 2.3.

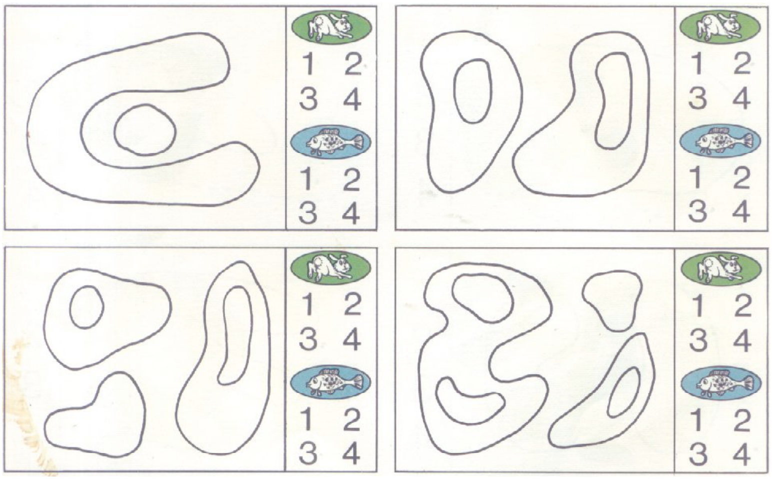


Figure 2.3: The puzzles have now become an example of using a topological invariant to make a two colour mapping of any rectangle containing a set of non-intersecting simple closed (connected) curves.

Several points follow. Firstly this is a concept of continuity topologically dividing a higher dimensional region that is immediately appreciated and understood by children young enough to not have proper reading skills or a verbally-based machinery to handle abstract concepts. Secondly is provides a mathematically meaningful expression of a theorem in abstract topology not generally taught until the third year of an undergraduate degree or graduate level. Thirdly it provides an example of a numeric topological invariant spanning algebra (arithmetic) and topology on a par with the Euler characteristic, Jones polynomial and homotopy group.

Likewise, as noted in the introduction, topological spaces and their knottings from the Moibius band to the Klein bottle and worm holes in space-time provide a rich and stimulating menagerie of mathematical challenges to our ideas of spatial connectivity and dimension which are well-known by many secondary students as the stuff of comic book and space fantasy on television largely because they can be appreciated directly by the human imagination even when they cannot be embedded in 3-D space.

Specific cultural traditions such as the appliqué work of the Hmong of Thailand, Laos, Vietnam and China figure 1.1 are direct intuitive expressions of the same nested Jordan curve two colouring property, showing in another way an 'innate' appreciation for the topological nature of these relationships between curves and surfaces.

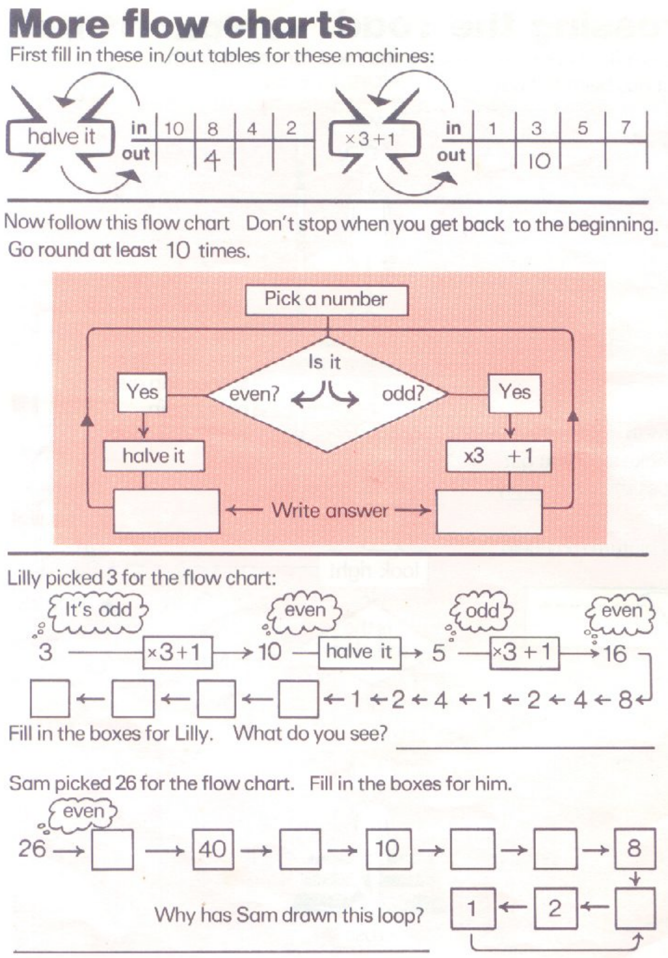


Figure 2.4: The 'Collatz' problem presented as a number cruncher for children in "Inner Ring Maths".

(b) The Discrete Dynamics of Choice

A second problem made the 'cover piece' of Inner Ring Mathematics the flow chart number cruncher illustrated in figure 2.5. Once again, this is a problem which can be appreciated as a recursive arithmetic decision-making process and even more so for its surprising variety and unpredictability by young children who have no knowledge of abstract algebra, yet, far from being trivial, it remains an unsolved problem in mathematics whether all numbers generate a sequence forming a discrete orbit, which is eventually periodic to the portrayed cyclic sequence $4 \rightarrow 2 \rightarrow 1$. Also called the 'Collatz' conjecture after Lothar Collatz, it is discussed in detail in Wolfram's Mathworld[7], in Wikipedia[8] and forms a key text editing example in Matlab as illustrated in figure 2.6. Paul Erdős said of the Collatz conjecture: "Mathematics is not yet ready for such problems" yet it is understood and appreciated by children learning to read. One clear thing this example illustrates is that discrete dynamics is as central to human mathematics as algebra and analysis are.

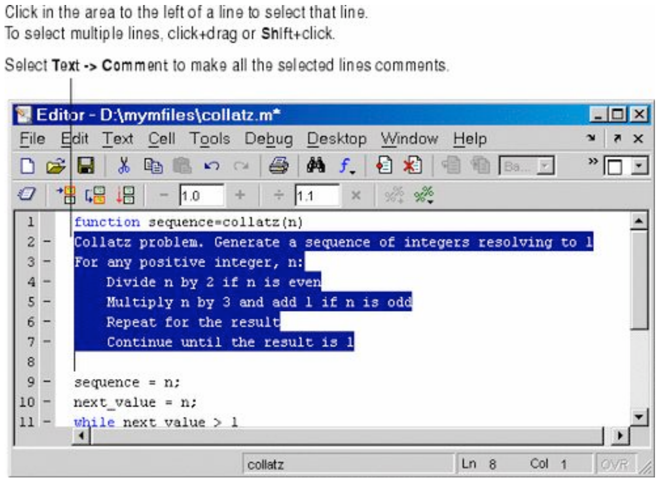


Figure 2.5: The Collatz problem iconic in Matlab's text editing tutorials

Matlab illustrates the nature of the problem well as a key example of a discrete dynamical system. Computations using the routine of figure 2.6 all display eventually periodic iterations to the 4>2>1 cycle, but with vastly varying orbit lengths and maxima.

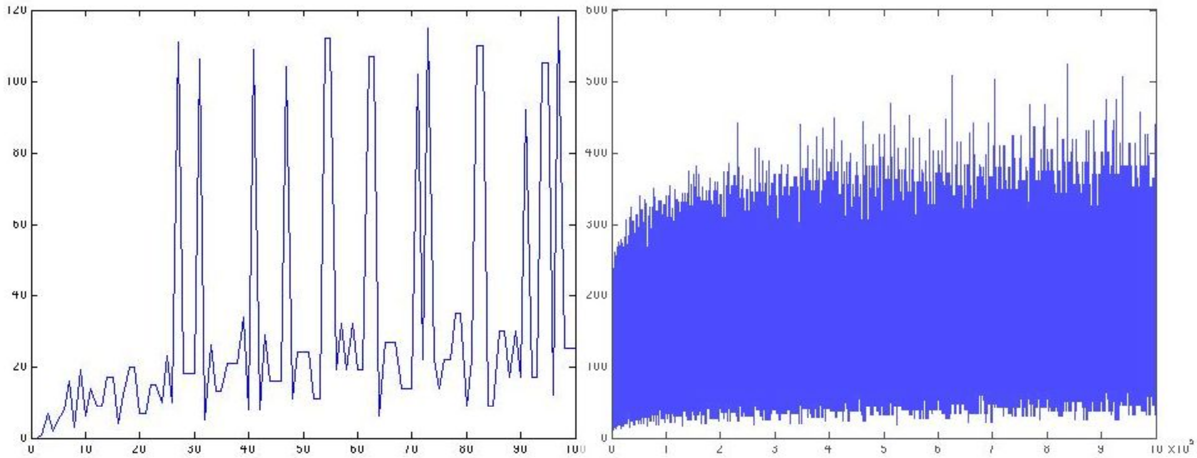


Figure 2.6: Matlab simulation of the 'Collatz' problem shows that although it trends to a logarithmic convergence, individual starting values have huge variation in the eventually-periodic orbit lengths of the iteration, rendering the problem unsolved to date, despite being able to be appreciated by young children.

The iteration is poised unstably between a decreasing and an increasing rule. The fact that some systems can tend to a variety of asymptotic limits is demonstrated by replacing the 3x+1 by 5x+1. Three distinct sequences cover the numbers 1-7, including two different eventual periodicities - of 7 and 12 and an unbounded orbit.

4	2	1	6	3	16	8							
5	26	13	66	33	166	83	416	208	104	52	26		
73	366	183	916	458	229	1146	573	2866	1433	7166	3583	17916	…

Table 1: Three orbit types of (5x+1)OR(x/2): Top period 7, middle period 12, bottom unbounded.

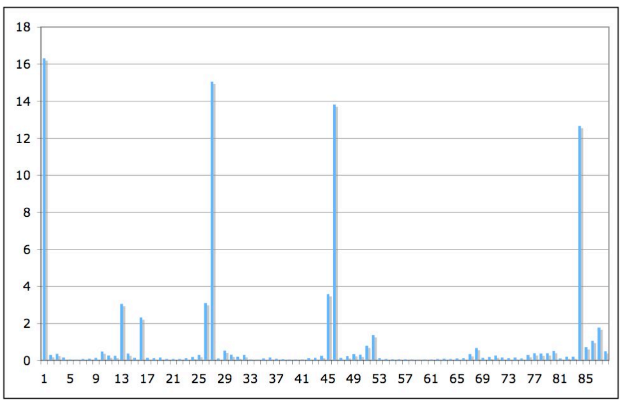


Figure 2.7: Divergences of the (3x+1)OR(x/2) iteration from [2.1] ($m(r)/r^2$) for successive path records.

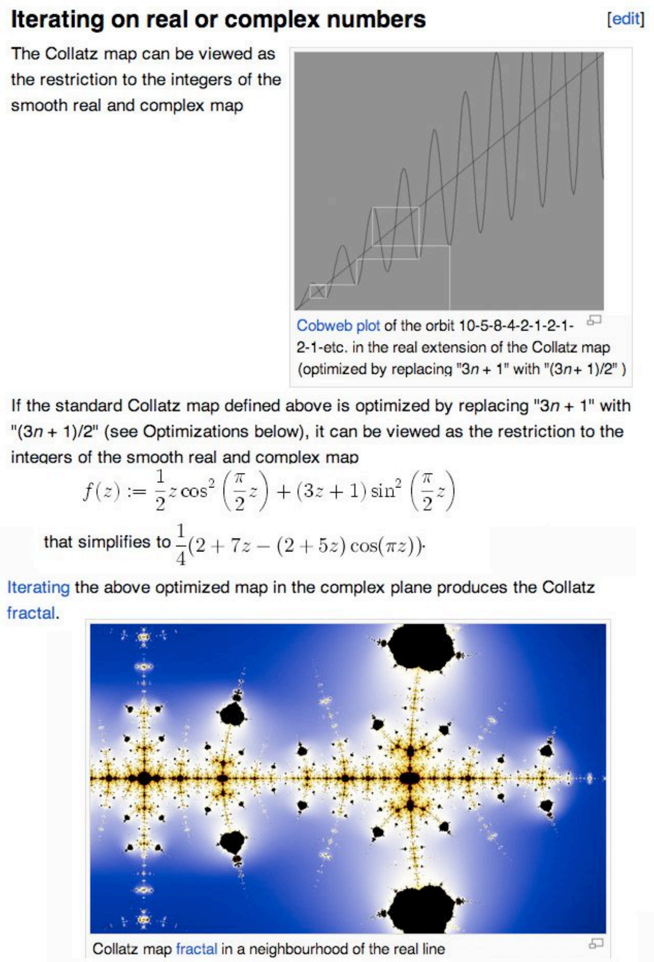
Computational paths records have been established for this iteration[9]. A current highest known, the 88th path record is 1,980976,057694,848447 which reached 64,024667,322193,133530,165877,294264,738020 before eventually entering the 4>2>1 cycle. The successive path records 2(2) 3(16), 7(52). 15(160) 27(9232) etc. vary erratically along:

$$\log_2 m(r) = 2\log_2 r \qquad [2.1]$$

Although the Collatz is really a problem in discrete dynamics it is still using arithmetic binary operations and is hence algebraic in nature. It also involves theoretical logic of computation in performing the iteration numerically by mental arithmetic in a manner immediately apparent to a child. It thus presents what is an unsolved abstract algebraic problem in immediately accessible terms of repeated discrete decision-making.

In evolutionary terms many social interactions are molded by a discrete sequence of prisoner's dilemma moves of cooperation and betrayal, which build up our understanding of trust and good character. The Wason test is a logical puzzle, in which a person is given a conditional 'rule', if P, then Q, together with four two-sided cards displaying information of the form P, not-P, Q, and not-Q. Subjects are instructed to turn over the cards necessary to determine whether the rule holds. The correct solution is to turn over the cards displaying P and not-Q to see whether their other sides contain not-Q and P respectively, because those, and only those, cards can violate the rule. is much better solved by humans when framed in terms of breaking a social contract, such as detecting cheating (NOT of legal age but IS drinking).

The case could thus be made that, in evolutionary terms, three of the most the most acutely important mathematical skills are topological and geometrical skills to do with hunting, generic classificatory skills to do with gathering and discrete dynamics to handle the Machiavellian intelligence of human groups.



Iterating on real or complex numbers [edit]

The Collatz map can be viewed as the restriction to the integers of the smooth real and complex map

Cobweb plot of the orbit 10-5-8-4-2-1-2-1-2-1-etc. in the real extension of the Collatz map (optimized by replacing "3n + 1" with "(3n + 1)/2")

If the standard Collatz map defined above is optimized by replacing "3n + 1" with "(3n + 1)/2" (see Optimizations below), it can be viewed as the restriction to the integers of the smooth real and complex map

$$f(z) := \frac{1}{2} z \cos^2\left(\frac{\pi}{2} z\right) + (3z+1)\sin^2\left(\frac{\pi}{2} z\right)$$

that simplifies to $\frac{1}{4}(2 + 7z - (2+5z)\cos(\pi z))$.

Iterating the above optimized map in the complex plane produces the Collatz fractal.

Collatz map fractal in a neighbourhood of the real line

Figure 5.11a Section of Wikipedia on the Collatz conjecture, showing its fractal extension to complex numbers.

(c) Mathematics as a Cultural Maze

Ian Stewart in "The Magical Maze"[10] has portrayed mathematics, not as a pipeline or axiomatic hierarchy, but as a maze. In his opening words: "Welcome to the maze. A logical maze, a magical maze. A maze of the mind." The maze is mathematics. The mind is yours. Let's see what happens when we put them together".

In contrast to the classical view of mathematics as an axiomatic and heuristic hierarchy in which algebra and analysis stand as a dual core, a new view of mathematics is emerging as a result of richer resources, on the world wide web and through packages such as Mathematica, Matlab and Maple that mathematics is more like a tangled bank, a maze or an interconnected graph of concepts and states, in which topics relevant to a particular quest can be rapidly explored and understood in a way which would require a long bureaucratic journey up an undergraduate pipeline and major finally passing specialized graduate courses to access. This view is emphasized by the fact that Mathworld and Wikipedia are themselves, unlike conventional text books and more like fully-referenced research articles, themselves conceptual mazes in which multiple live links carry us from topic to topic on the wings of research discovery, personal interest or sheer imagination, healing a rift typified by the comment – "So you are a mathematician! I gave that up in the 4th form!"

To illustrate how effective this maze-based view of mathematics is, let us go back to the Collatz problem that first appeared in a children's book and pick out two gems from Wikipedia (figure 5.11a) and Mathworld (figure 5.11b) concerning the nature of this problem. The two figures show different perspectives on the orbit of values from a given stating number. Neither of these connections would be expected in a mathematics text book, but are immediate extensions of the core concept which can be explored further by live links and linked references to the original research. Teilhard de Chardin's 'noosphere'[11] has thus become realized mathematically as a networked wiki. The Wikipedia entry is poignant, not just because it is edited and maintained real-time by the viewers, but because it carries the problem right into discrete dynamics and chaos theory, where the intractability of the problem naturally lies. It is fundamentally the maze properties of mathematics in terms of its logical relatedness connecting diverse areas, which give mathematics its power of explanation. The central properties of a language are not hierarchical definition, but maximal capacity to be used freely to articulate, in an optimal way, abstract semantic ideas. Effectively each language generates a maze, not only of words and phrases, but of expressible conjectures, comments and arguments, as well as metaphors and tales of the intrigues of human character, which it is up to our Machiavellian intelligence to put to the most flexible and advantageous use possible.

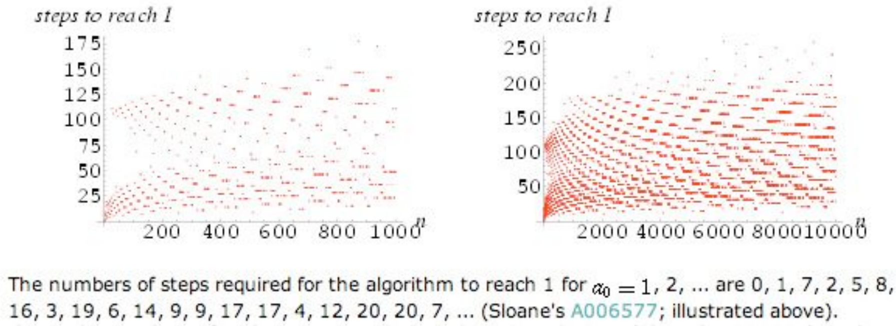


Figure 5.11b Section of Wolfram's Mathworld on the Collatz conjecture.

To make this point even clearer, mathematical research, to be efficient and innovative, by necessity, has to link often unrelated areas, such as group theory, graph theory and computational science, to solve real problems of complexity. In doing so it also has to find the most direct route to the kind of articulated expression we call a mathematics research paper. Despite the hierarchical nature of tertiary mathematics education, at the research level this necessarily means finding the shortest span across a the graph of mathematical ideas to get us from the problem to the solution, in much the same way a person solving the Rubik revenge vanquishes a state space containing 10^{50} vertices in a path to the solution of only perhaps 50 steps. It is thus a serious dichotomy that to achieve this level of networking efficiency, a heuristic assumption is made that we must teach mathematics in a bureaucratic hierarchy, even though this is largely in conflict with that very fascination with novelty and exploration as a discovery process the nature of mathematics as a conceptual maze provides.

3: The $\varepsilon - \delta$ Game, Topology and Two Small Clouds in Classical Analysis

The foundation of analytic continuity is the cryptic statement:

$$\forall \varepsilon > 0 \ \exists \delta > 0 : 0 < |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$

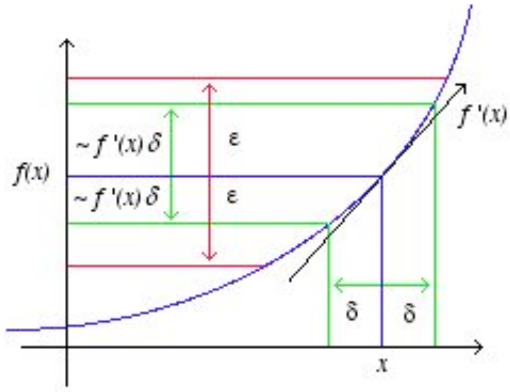


Figure 3.1: Solving the ε - δ game requires finding a $\delta(\varepsilon)$ for each ε. If this is not to become a stand-off, we need to know the local slope $f'(x)$ to fit the sandwiches.

This means that no matter how close we choose ε to be, there has to be a δ sandwich in the domain corresponding to it that will map into the range inside the ε sandwich. The counterintuitive aspect of this to a student used to dealing with practical problem solving is that unless we indulge a subtle form of 'cheating' this looks like an unresolvable standoff, because no matter how many δ values they pick demonstrating convergence, as we tend to *x*, an opponent can still pick an ever smaller and more meticulous ε demand, leading to perpetual impasse.

What we are using here is equivalent to local Lipschitz continuity[12]. A function is Lipschitz continuous if

$\exists K > 0 : \forall x, y \in D_f \ |f(x) - f(y)| < K|x - y|$ and is called a contraction if K<1. [3.2a]

Lipschitz continuity can also be defined locally $\forall x \in D_f, \exists \delta > 0 : f$ Lipschitz on $(x - \delta, x + \delta)$. [3.2b]

In effect ε - δ continuity is a two-player game, in which, to establish continuity, the domain player has to be always able to outmaneuver the range player, by finding a $\delta(\varepsilon)$ for each ε, effectively finding a functional relationship through a local inverse. As pictured in figure 3.1, to all intents and purposes, this depends not just on continuity but gauging the local slope of the function in the neighbourhood of *x* and hence using the local derivative $f'(x)$. But this is really a form of cheating at the game, because it really only works when the function is not just continuous, but differentiable, at least locally, and we are all taught that differentiability implies continuity, but not vice versa.

Example 3.1: Proving the continuity of $f(x) = x^2$, involves discovering the following functional relationship:

$$\left|(x+\delta)^2 - x^2\right| < \varepsilon \Leftrightarrow \left|2x\delta + \delta^2\right| < |4x\delta| < \varepsilon \Rightarrow \delta(\varepsilon) < \frac{\varepsilon}{2|2x|} = \frac{\varepsilon}{2|f'(x)|}, \ \delta < 2|x| \qquad [3.3]$$

What is actually happening here is that we are estimating a $\delta(\varepsilon)$ by comparing the ratio of the range ε and domain δ gaps, effectively using the derivative $f'(x)$, which correctly gives the varying $\varepsilon:\delta$ ratio for differentiable functions, which powers of x clearly are. A quick check of the formula for x^n gives:

$$\begin{aligned}&\left|(x+\delta)^n - x^n\right| < \varepsilon \Leftrightarrow \left|nx^{n-1}\delta + \ldots + \delta^n\right| < \left|nx^{n-1}\delta + \frac{n(n-1)}{2!}x^{n-1}\frac{\delta}{n^2} + \ldots + x^{n-1}\frac{\delta}{n^{2n-2}}\right| \\ &< \left|nx^{n-1}\delta + x^{n-1}\delta + \ldots + x^{n-1}\delta\right| = \left|2nx^{n-1}\delta\right| < \varepsilon \Rightarrow \delta(\varepsilon) < \frac{\varepsilon}{2\left|f'(x)\right|} \text{ for } \delta < \frac{|x|}{n^2}\end{aligned} \qquad [3.4]$$

The same argument can be extended to all standard functions, which are generally differentiable on their domains through their power series representation, which is effectively a fractal polynomial. This is particularly true for the trigonometric, hyperbolic and log functions, all of which are derived from the (complex) exponential:

$$\begin{aligned}&f(x) = e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!} = 1 + x + \ldots + \frac{x^n}{n!} + \ldots,\ f'(x) = \sum_{n=1}^{\infty}\frac{nx^{n-1}}{n!} = \sum_{n=1}^{\infty}\frac{x^{n-1}}{(n-1)!} = f(x) \\ &\left|e^{x+\delta} - e^x\right| = \sum_{n=0}^{\infty}\frac{\left|(x+\delta)^n - x^n\right|}{n!} < \sum_{n=1}^{\infty}\frac{\left|2nx^{n-1}\right|}{n!}\delta < 2\sum_{n=1}^{\infty}\frac{\left|x^{n-1}\right|}{(n-1)!}\delta = 2\left|f'(x)\right|\delta < \varepsilon \Rightarrow \delta < \frac{\varepsilon}{2\left|f'(x)\right|}\end{aligned} \qquad [3.5]$$

The question then naturally arises as to what we do when we come up against proving a non-differentiable function is continuous. The classic such function is the broad class of Weierstrass functions[13] typified by

$$w(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x),\ 0 < a < 1,\ b\ +ve\ odd,\ ab > 1 + \frac{3}{2}\pi \qquad [3.6]$$

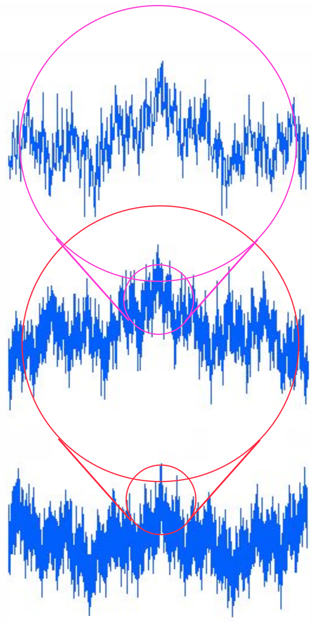


Figure 3.2 Weierstrass functions are fractal functions (Matlab simulation).

Weierstrass functions are fractals as illustrated in figure 3.2, fractality arising directly from their Fourier series representation. Their Hausdorff dimensions are closely bounded by $\frac{\log a}{\log b} + 2$ [14] [3.7]

Weierstrass functions can be proved to be nowhere differentiable, which is relatively obvious, since formal differentiation leads to divergence:

$$w'(x) = \sum_{n=0}^{\infty} -\pi(ab)^n \sin(b^n \pi x),\ (ab)^n > \left(1 + \frac{3}{2}\pi\right)^n \qquad [3.8]$$

The proof that they are continuous everywhere is immediate. Since the terms of the Fourier series are bounded by $\pm a^n$ and this has finite sum for $0 < a < 1$, convergence of the partial sums $w_n(x) = \sum_{n=0}^{n} a^n \cos(b^n \pi x)$ to the function is uniform. The uniform limit of continuous functions is continuous, and each partial sum is continuous:

$$\begin{aligned}&\forall \varepsilon > 0\ \exists \delta > 0,\ N \in \mathbb{N} : 0 < |x - y| < \delta,\ n > N \\ &\Rightarrow |w(x) - w(y)| < |w(x) - w_n(x)| + |w_n(x) - w_n(y)| + |w_n(y) - w(y)| < \varepsilon\end{aligned} \qquad [3.9]$$

Thus once we reach $w_n(x)$, we can switch from worrying about the derivative, or the local Lipschitz constant, because, no matter how fractally precipitous successive terms become, in generating the nowhere-differentiable uniform limit, having used the continuity of the differentiable partial sum $w_n(x)$, we can now ignore the graph of the function because of the ever-diminishing bound on the divergence caused by the additional fractal terms.

Uniform convergence guarantees that convergence is independent of x i.e.

$$\forall \varepsilon > 0 \forall x \in D_f\ \exists N \in \mathbb{N} : n > N \Rightarrow |f_n(x) - f(x)| < \varepsilon \Leftrightarrow \|f_n - f\|_\infty < \varepsilon,\ \|f\|_\infty = \sup_{x \in D_f}\{|f(x)|\} \qquad [3.10]$$

By contrast with Lipschitz continuity for real functions on [0,1], which are differentiable almost everywhere (i.e. except on a set of measure zero as defined below), in a topological sense, the set of nowhere-differentiable real-valued functions on [0,1] is dense in the vector space of all continuous real-valued functions on [0,1] with the topology U of uniform convergence derived from $\|\ \|_\infty$ of [3.10] that is $f_n \to f$ in $U \Leftrightarrow f_n \to f$ uniformly. The uniform norm and uniform convergence are here functioning in much the same way as the Hausdorff metric (the maximum distance either of two compact sets extend beyond the other) does for iterated function systems[15] based on (Lipschitz) contraction mappings forming a sequence in the space of compact sets in the plane, which converges uniformly to the fractal attractor of the system.

If we consider the Fourier series for a real function $f(x)$ on the interval:

$$f(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos(2k\pi x) + \sum_{k=1}^{\infty} b_k \sin(2k\pi x),$$
$$a_0 = \int_0^1 f(x)dx,\ a_k = \int_0^1 f(x)\cos(2k\pi x)dx,\ b_k = \int_0^1 f(x)\sin(2k\pi x)dx \qquad [3.11]$$

since both sin and cos are bounded by ±1, all that is required for uniform convergence is that the coefficients diminish in such a way that their successive partial sums are convergent. It is thus possible to construct a variety of Weierstrass functions, for example the polynomial type in Wolfram's Mathworld[16]:

$$f_a(x) = \sum_{k=1}^{\infty} \frac{\sin(\pi k^a x)}{\pi k^a},\ k = 2,3,4,\ldots \qquad [3.12]$$

Demonstrating fractal non-differentiable functions form a dense subset is also in principle straightforward, by forming a sequence using a Weierstrass function: ${}_m w \to f,\ \ {}_m w(x) = f(x) - \sum_{n=m}^{\infty} a^n \cos(b^n \pi x)$ [3.13]

The ${}_m w$ are nowhere differentiable, because they are a superposition of f and a rescaled Weierstrass function, and tend to f uniformly, just as the partial sums w_n tend to w, as each differs from its limit by the same set of terms.

Hence the space of continuous functions on [0,1] is densely permeated by fractal non-differentiable functions, and we would be better off working with the uniform topology and teaching students about topological continuity in a way which admits all the contortions it provides, including interesting functions in the real world, such as the fractal waves on the ocean, rather than limiting the arena of interest to the ideal archetypes we 'cheat' on proving continuity for, by relying on their differentiability to find $\delta(\varepsilon)$.

In a metric space (X,d) we replace the standard distance $d(x,y) = |x-y|$ in $\mathbb{R}$ with any real function on pairs of points which obeys non-negativity, symmetry and the triangle inequality. The statement of continuity for $f:(X,d)\to(Y,e)$ then becomes $\forall\varepsilon > 0\ \exists\delta > 0 : d(x,y) < \delta \Rightarrow e\big(f(x), f(y)\big) < \varepsilon$. [3.14]

If we define an open ball as the analogue of an open interval i.e. $B_\varepsilon(x) = \{y \in X : d(x,y) < \varepsilon\}$, we can then restate continuity in terms of open balls: $\forall\varepsilon > 0\,\exists\delta > 0 : y \in B_\delta(x) \Rightarrow f(y) \in B_\varepsilon(f(x))$, or

$$\forall B_\varepsilon(f(x))\,\exists B_\delta(x) : f(B_\delta(x)) \subseteq B_\varepsilon(f(x)). \qquad [3.15]$$

We can then define an open set O as one where every point has an open ball neighbourhood around it: $\forall x \in O\,\exists B_\varepsilon(x) \subseteq O$ and can straightforwardly prove that any open set is a union of open balls and hence any union of open sets is open. However simple examples such as $\bigcap_{i=1}^{\infty}\left(-\frac{1}{n}, 1+\frac{1}{n}\right) = [0,1]$ demonstrate that an intersection of open sets is necessarily open only if the intersection is over a finite collection. We thus have the basis for a topological space - a pair (X,τ) where X is a set and τ is a collection of open subsets of X defined by:

$$\text{(i) } O_i \in \tau, \forall i \in I \Rightarrow \bigcup_{i\in I} O_i \in \tau \quad \text{(ii) } O_1,\ldots,O_n \in \tau \Rightarrow \bigcap_{i=1}^{n} O_i \in \tau \qquad [3.16]$$

that is a collection of open subsets O_i of X is 'closed' under arbitrary unions but only finite intersections and their complements, closed sets are 'closed' under arbitrary intersections and finite unions.

This is a primary example of symmetry-breaking, the open sets throwing off their boundary points, and their complements, the closed sets, retaining them. Topology breaks the symmetry between union and intersection which characterizes set theory and logic in the form of a Boolean algebra in which we have two binary operations with laws which are mutually commutative, associative and distributive and have additional laws of absorption ($A \cap (A \cup B) = A$ and its dual) and complements ($A \cap \neg A = \varnothing$ and its dual) in which we can exchange $\cup$ and $\cap$ (or $\vee$ and $\wedge$ in logic) to gain dual statements like De Morgan's laws:

$$\neg(A \cup B) = \neg A \cap \neg B,\ \neg(A \cap B) = \neg A \cup \neg B. \qquad [3.17\ a,b,c]$$

When Lord Kelvin said there were two small clouds on the horizon of classical physics – namely the Michelson-Morley experiment confirming the invariance of the speed of light foreshadowing relativity and black body radiation foreshadowing quantization – these two statements harbingered a cultural revolution which spelt the end of classical physics. Analysis likewise has two small dark clouds, which in a similar sense may spell the nemesis of the classical paradigm.

The first is the non-equivalence of metric spaces under homeomorphism, the natural definition of continuous mapping equivalence. A homeomorphism is a function between spaces that is 1-1, onto, and continuous in both directions. However there are simple examples of spaces that are not metrically equivalent but are homeomorphic. Thus metric spaces cannot be the natural vehicle for continuity, but topological spaces are.

Example 3.2: The two metric spaces (X,d_1) and (X,d_2) $X=(0,1]$, $d_1(x,y)=|x-y|$, $d_2(x,y)=\left|\frac{1}{x}-\frac{1}{y}\right|$ are not metrically equivalent. We can see this at once, because many of the points in (X,d_2) are distance apart much greater than 1 tending to infinity, while all pairs in (X,d_1) are closer then 1 apart. There can thus be no rescaling of the finite metric to contain all the balls in the unbounded metric and hence no metric equivalence is possible between them.

Moreover (X,d_1) is not complete. It does not contain all its limit points, since the sequence $\left\{\frac{1}{n}\right\}\to 0\notin(0,1]$ is Cauchy (pairs of points in the sequence become arbitrarily close to one another), but the limit 0 is outside the space (0,1]. By contrast in (X,d_2), this sequence is not Cauchy, since $\left|\frac{1}{1/m}-\frac{1}{1/n}\right|=|m-n|$ does not tend to 0 for all $m,n\to\infty$, and this metric space is complete, being metrically isomorphic to $[1,\infty)$, as noted below.

However the two spaces are homeomorphic. $f(x)=\frac{1}{x}$ is a metric identity between (X,d_2) and $[1,\infty)$ with the standard metric $d_1(x,y)$ and this is an equivalence which is also necessarily a homeomorphism. But $f(x)=\frac{1}{x}$ is also a continuous bijection on $(0,\infty)$ using only the standard metric and is also its own continuous inverse so $[1,\infty)$ is also homeomorphic with (X,d_1). Hence (X,d_1) and (X,d_2) are homeomorphic.

It is thus natural to move from the metric ε - δ definition of continuity to the topological one:

Theorem 3.1: $f:X\to Y$ is a continuous function of metric spaces $\Leftrightarrow O\subseteq Y$ open $\Rightarrow f^{-1}(O)\subseteq X$ open.
proof:
($\Rightarrow$) Suppose *f* is continuous and $O\subseteq Y$ open. If $f^{-1}(O)=\varnothing$ then it is open, so assume the contrary. $x\in f^{-1}(O)\Rightarrow f(x)\in O$ open so $\exists B_\varepsilon(x)\subseteq O$. Hence by continuity $\exists\delta>0: d_1(x,y)<\delta\Rightarrow d_2\left(f(x),f(y)\right)<\varepsilon$ or $\exists B_\delta(x): f(B_\delta(x))\subseteq B_\varepsilon(f(x))\subseteq O$. But then $B_\delta(x)\subseteq f^{-1}(O)$ so we have found an open ball around any $x\in f^{-1}(O)$ making it open.
($\Leftarrow$) Suppose O open $\Rightarrow f^{-1}(O)$ open. Consider $B_\varepsilon(f(x))$. Since this is open, $f^{-1}(B_\varepsilon(f(x)))$ is also open and contains *x*. Hence there exists an open neighbourhood of *x* in $f^{-1}(B_\varepsilon(f(x)))$ i.e. $\exists B_\delta(x)\subseteq f^{-1}(B_\varepsilon(f(x)))$.
But this is the same thing as saying $\forall\varepsilon>0\ \exists\delta>0: d_1(x,y)<\delta\Rightarrow d_2\left(f(x),f(y)\right)<\varepsilon$ so *f* is continuous.

This definition becomes the basis of topological continuity. It might appear even more inaccessible than the ε - δ game because we have to deal with arbitrary open sets, but this is not so for elementary real functions, because it is sufficient to show inverse images of open intervals are open by the above theorem and to do this is no more difficult than the original Lipschitz type use of the derivative as a scaling factor in the ε - δ game, but it also has manifest advantages in making real the sense of topological discontinuity directly appreciated in breaking of a continuously drawn curve in space and the need to assign the boundary points of the breakage.

The second small cloud on the horizon of analysis comes not from the concept of metric, but of measure. Although the rationals are spread out densely on the number line so that between any two irrationals is a rational and vice versa, the rationals are countable, while the reals and hence the irrationals, have a strictly higher cardinality. For those interested, you can count the rationals by making a 2-D grid of all fractions and scanning the diagonals $\frac{1}{1}\to\frac{2}{1}\to\frac{1}{2}\to\frac{1}{3}\to\frac{2}{2}\to\frac{3}{1}$ etc. This is redundant, but shows we can put all rationals into a list.

To prove the reals are not countable, suppose you have a list of all elements of (0,1) as decimals $r_i=0.r_{i1}r_{i2}r_{i3}...$ We can always find an element $s=0.s_1s_2s_3...$ not on the list simply by making $s_i\neq r_{ii}$. This also shows in principle a way to make an identification between elements of $(0,1)$ and subsets of the natural numbers $\mathbb{N}$, by identifying the binary representation $b = 0.b_1 b_2 b_3 ...$ where $b_i = 0,1$ with the subset $B = \{i \in \mathbb{N} : b_i = 1\}$. Since the number of subsets of a set containing n elements is 2^n, this gives us the famous relation $2^{\aleph_0} = c$, between the countable cardinality $\aleph_0$ of $\mathbb{N}$ and $\mathbb{Q}$, and the uncountable cardinality c of $\mathbb{R}$.

Since the development of the Riemann integral, there has been a love affair with the idea of taming the rationals sufficiently to prove that a more general notion of integral (known as the Lebesgue integral) should successfully show that

$$\int_{-\infty}^{\infty} f(x)dx = 0,\ f(x) = \begin{cases} 1,\ x \in \mathbb{Q} \\ 0,\ x \notin \mathbb{Q} \end{cases}. \qquad [3.18]$$

Based on the Riemann idea of a limit of rectangles (figure 3.3), this integral does not exist because the maximum height of each rectangle is 1 and the minimum is 0, so no limit exists.

The solution to this dilemma comes in a different non-symmetry-breaking modification of Boolean algebras in which we consider instead a collection of subsets closed under countable union and complement, called a σ-algebra. By De Morgan's laws, this is also closed under countable intersection. The smallest σ-algebra over the reals containing the intervals is the algebra of Borel sets. Naturally it includes open and closed sets and the additional sets we get forming countable intersections and unions of these.

If we now expand to consider any set which differs from a Borel set by a null set, one which can be covered by a countable union of intervals the sum of whose lengths is less than any $\varepsilon > 0$, we arrive at Lebesgue measure[17].

In particular, we can define the outer measure of any subset B of $\mathbb{R}$: $\lambda^*(B) = \inf\{length(M) : M \supseteq B\}$
where M is a countable union of intervals, the sum of whose lengths is $length(M)$. [3.19]

A is then Lebesgue measurable if $\forall B \subseteq \mathbb{R},\ \lambda^*(B) = \lambda^*(B \cap A) + \lambda^*(B - A)$ [3.19]
i.e. if the measure of the inside and the outside add to that of the whole set in each case.

This makes things very easy for an integral over the rationals, because $\mathbb{Q}$ is a null set, since it is countable, and we can cover each enumerated rational $q_i,\ i \in \mathbb{N}$ by an interval of length $\frac{\varepsilon}{2^i}$ whose sum is $\varepsilon \sum_{i=1}^{\infty} \frac{1}{2^i} = \varepsilon$.

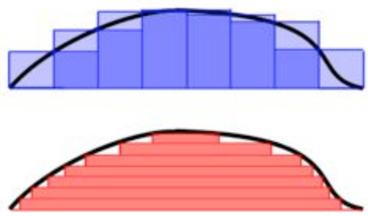


Figure 3.3: Rather than partitioning the domain, as in the Riemann integral (blue), the Lebesgue integral (red) works its limit by partitioning the range and adding the areas gained from multiplying the measure of the set for which the function is higher than a given level by that level height.

For the above function [3.18], which is either 1 or zero, this will just be the measure of $\mathbb{Q}$, which is 0.

However, while solving the measure problem for functions over 'virtually every set of interest we might encounter' there remains an infinite collection of residual non-measurable sets, called Vitali sets, which cast a pall shadow over the ideal of real numbers.

Example 3.3: Consider the rational equivalence class of real x: $[x] = \{y \in \mathbb{R} : x - y \in \mathbb{Q}\}$. This set of equivalence partitions $\mathbb{R}$ into disjoint subsets. We construct a Vitali set[18] $V \subseteq [0,1]$ by choosing one representative from each class, via the axiom of choice, an independent axiom, which allows such choices from arbitrary collections.

Now consider an enumeration $q_i,\ i \in \mathbb{N}$ of $\mathbb{Q} \cap [-1,1]$. From the definition of V the sets $V_i = V + q_i$ are pairwise disjoint since numbers differing by a rational are in the same equivalence class and only one representative of this was chosen. We can also show $[0,1] \subseteq \bigcup_k V_k \subseteq [-1,2]$. The first inclusion is because, for each x in $[0,1]$ if v is the representative for $[x]$ then $x - v = q_l$, some l and so $x \in V_l$. The second inclusion is clear from the maximum divergences from $[0,1]$ of 1.

This gives rise to a paradox, because Lebesgue measure is countably additive and translation invariant, so $1 \le \lambda^*\left(\bigcup_k V_k\right) = \sum_{k=1}^{\infty} \lambda^*\left(V_k\right) \le 3$, but each of the $\lambda^*\left(V_k\right)$ are identical by translation invariance, so we have a countable sum equals a finite number which is impossible since any such sum must be 0 if the terms are 0 or ∞ if the terms are finite and equal.

This is a problem that goes right to the heart of mathematics as a cultural language, which might have a very different description on another planet harbouring sentient life. Mathematicians have of course made a menagerie of number systems incorporating infinitesimals smaller than any real, including the hyperreals[19], the long line[20] and others, but nevertheless there is a serious problem about measure based on countability when we try to consider 'inverse countability' - factoring the reals into equivalence classes each containing a countable number of members, which makes the quest of measure a *will o' the wisp* of the mathematical will to order.

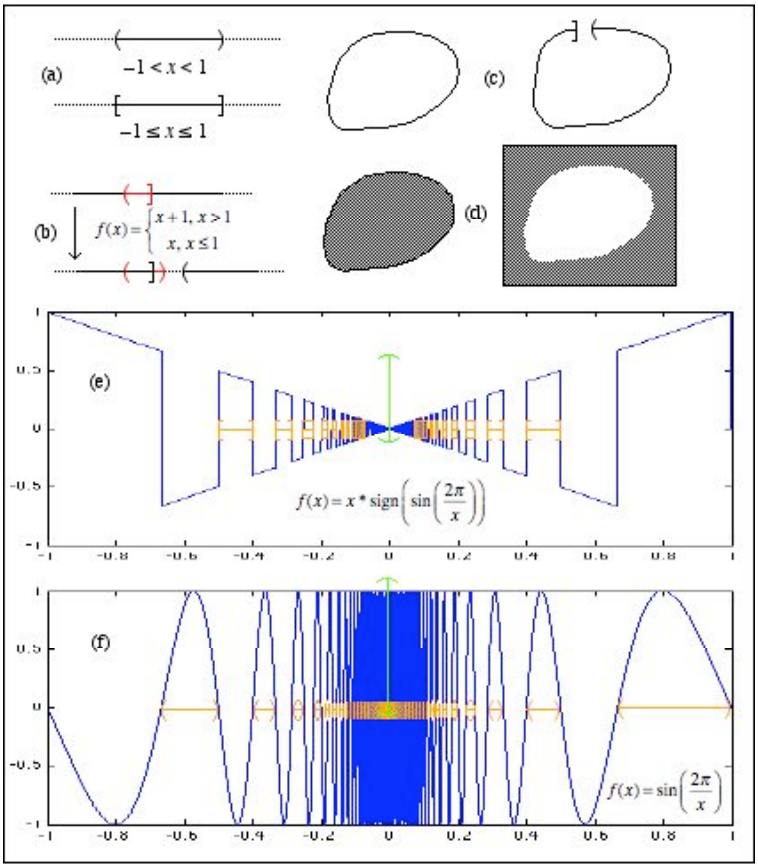


Figure 3.4: (a) Open and closed intervals and their variants are easily appreciated by school children. (b) A discontinuous mapping on the number line immediately reveals the problem of assigning boundary points. The inverse image of the open interval (red) has a boundary point as a result of the discontinuity. This topological idea of continuity based on openness and boundaries extends naturally to mappings of curves (c) and regions (d). The topological definition of continuity works as well as the ε - δ game for proving a function is continuous, as it can access the same Lipschitz arguments in the light of theorem 3.1, however it has an advantage in searching for discontinuities, because astute choices of open intervals in the range can highlight points of discontinuity as boundary points in the inverse image. The function in (e) has discontinuities at a sequence of values $\pm\frac{2}{n}$ tending to zero, but is continuous at 0. This is confirmed in $f^{-1}\left(-\varepsilon, \frac{1}{2}+\varepsilon\right)$, consisting of a finite union of closed intervals, demonstrating discontinuities at the end points, arbitrarily close to, but not including 0. The function in (f) is continuous on its domain $\mathbb{R}-\{0\}$, but if $f(0)$ is assigned to be 0, $f^{-1}(-\varepsilon, 1+\varepsilon)$ consists of a countable union of open intervals limiting to a single boundary point at 0 which is in the inverse image and is the one point of discontinuity on $\mathbb{R}$.

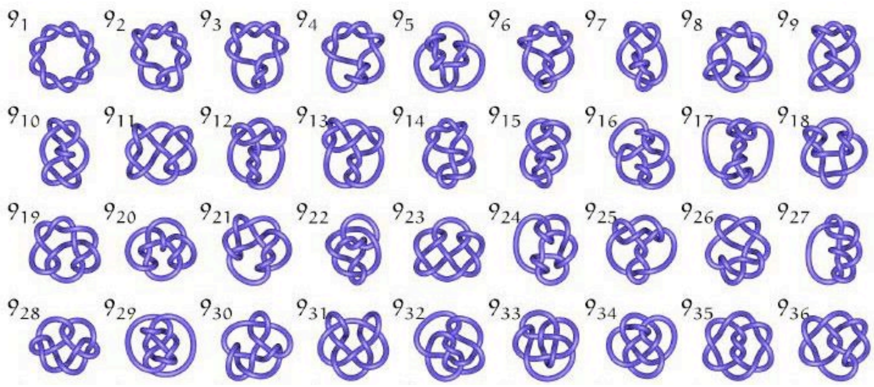


Figure 3.5 Knots of order 9 illustrate the realizable complexity of knot theory

An alternative to the credibility gaps of classical analysis is starting from realizable examples of continuity and path connectedness in which the continuity of curves is broken, requiring the assignment of the boundary points, leading into knots, manifolds and fractal topologies and how the idea of open set and topological space transcends the limitations of measurement in metric spaces, keeping the topological emphasis, while specializing to Lebesgue measure as particular studies require, rather than centering on classical analysis to the exclusion of both topology and geometry, when it is these latter areas that are most breathtaking to the imagination and still at the cutting edge of analysis, as the Poincaré conjecture[21] demonstrates.

4: Puzzles and Games as an Expression of Human Mathematical Imagination

One of the most obvious expressions of the human mathematical imagination in human culture is its presence in puzzles and games. Many of these and possibly the majority are geometrical and topological. There are of course reasons for this in that puzzles are frequently physical objects, but even when they are logical, conceptual, abstract or computational they still frequently use geometrical and topological ideas.

There is also a tendency for a given puzzle to bring together disparate areas of mathematics, implying that mathematics is best described as a tangled web, rather than a bureaucratic hierarchy of axiomatic systems. This is consistent with a new and very different view of mathematics as presented on the web in sites such as Math world and Wikipedia, in which mathematics is literally a maze of concepts related both by natural and logical affinity and by association, generalization and disparate linkages across widely differing fields to present complementary vies of a given phenomenon.

We present representative examples of puzzles and games to illustrate the diverse mathematical areas they bring into play and the types of mathematical reasoning in humans they highlight. Firstly let us examine three types of puzzle and game that specifically involve topological reasoning sometimes associated with geometrical thinking.

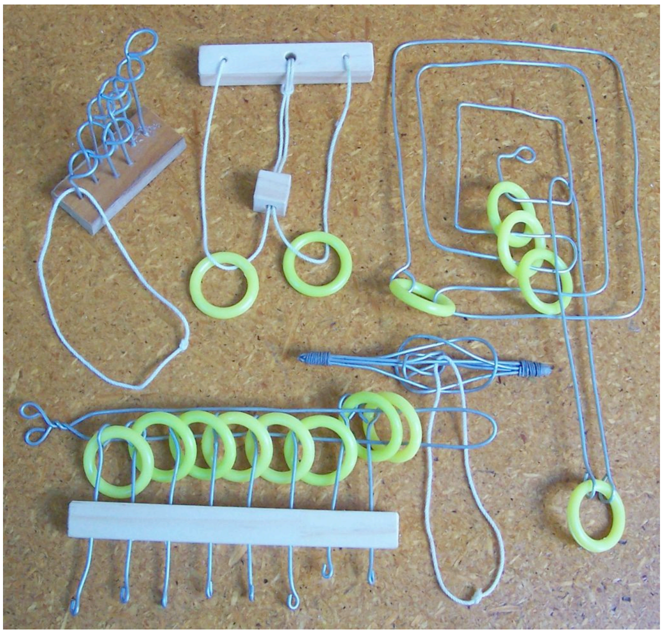


Figure 4.1: Five types of topological ring, wire and string puzzle

Example 4.1 a,b,c,d: Topological wire, loop and string puzzles.

A large class of puzzles use wire loops strings and rings to set up situations where the system is not in fact knotted or linked, which would make the puzzle impossible, but figuring out the unknotting moves is geometrically and/or topologically challenging.

The twin flight of loops (top left) is clearly unknotted, as each of the wire loops could be shrunk through those 'linked' over it. The string is thus not knotted and the puzzle is solvable. However visualizing the deformations of the string required is complex and involves an exponentiating number of topological moves, doubling for each additional step in the stairway, in the manner of a Towers of Hanoi problem. The lower right puzzle likewise has two nested pairs of unlinked loops, again requiring a recursive solution. The lower left puzzle consists of *n* linked rings and a long loop, which is initially linked over only the left hand *n*-th post. The (*k*+1)-th ring can be slid on and off the long loop only if only the *k*-th ring is linked over the loop. This give rise to the recursive relation for the number of moves to get to the *k*-th stage: $m_1 = 1,\ m_k = 2m_{k-1} + 1$ with solution $m_n = 2^n - 1$, giving 255 moves for this 8-stage puzzle. The top right-hand puzzle "Squaring Off" requires only four moves, corresponding to the four rings and the successive loops in the square, but the third is so counter-intuitive that many respondents have to ask for the solution. The centre top puzzle requires three simple moves to integrate the centre holes in the two pieces of wood and slip the string loop to the front. These puzzles thus combine three areas of mathematics, topology, geometry associated with the moves required to respect the fixed dimensions of the rings and wires, and recursive iterations governing the number of moves.

Figure 4.2: Left Trax has a qualifying white line and black loop. Centre: Tantrix pieces, including the forbidden pieces. Right: A Tantrix game in which the forbidden pieces are allowed for forced moves.

Example 4.2 a,b: Tantrix and Trax

Tantrix and Trax are games using regular square and hexagonal tilings but the strategy of both games depends on using the tilings to create topological loops and curves of maximum length. The two games share topological strategic curve building through a discrete process. Trax using invertible squares with crossed and uncrossed pathways plays cut-throat race for the first person to get a loop or line covering 8 rows or columns. Tantrix is more complex, having all combinations of three of four colours forming curves not crossing in a cartwheel. Cartwheels are omitted, because they provide sparse rearrangements being unchanged by a rotation of 180^o. Each player chooses a colour and tries to build the longest line or loop before play runs out. Before and after each turn players fill forced moves resulting from hollows in the tiling. Tantrix records exist for the longest 'lines' and loops, including a computer solution to the "four longest lines" puzzle by Paul Martinsen & Jamie Sneddon, April 1998 totaling 146 -34 red, 40 green, 35 blue and 37 yellow links.

The 56 Tantrix pieces plus 8 forbidden ones can be deduced easily on a combinatoric basis from all possible three colour curves on the hexagon. Pieces can have short curves joining adjacent faces, long curves spanning a face, and diametric straights. There are $2C_3^4 = 2\frac{4.3.2}{1.2.3} = 2.4 = 8$ 'triple shorts'- 4 combinations of colours, each in 2 orientations. The same applies to the forbidden 'triple straight' cartwheels. The 'straights with long or short curves' each have 4 ways to pick the straight colour and 3 ways to eliminate the fourth colour, or $P_2^4 = 4.3 = 12$ each. They do not have orientations as a 180^o rotation has reverses colour orientation, due to their internal symmetry. Finally the "two longs and a short" pieces have two orientations as well, so have 4.3.2 = 24 pieces.

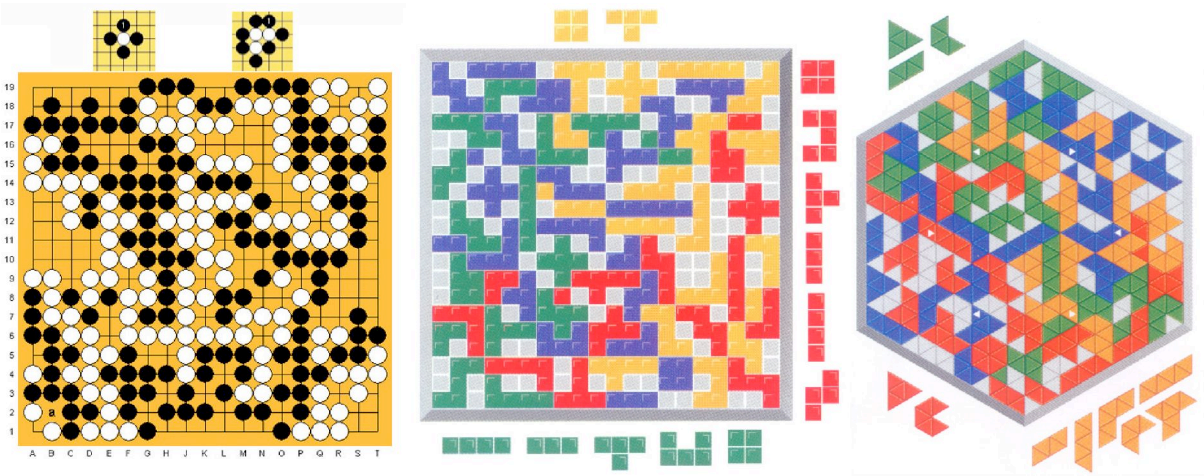


Figure 4.3a: Left: Go is based on capture by topological enclosure of regions following a discrete 4-cell von Neumann neighbourhood rule based on a player (black) holding the immediately adjacent squares (see insets).
Centre: Blokus uses all geometrical combinations of piece up to pentaminos to dominate space by building interpenetrating fractal trees, connectivity being maintained through adjacent corners.
Right: Blokus Trigon uses triangular pieces up to hexaminos and has different rules because triangular symmetry allows corner vertices to meet face vertices.

Examples 4.3: Go and Blokus, Dots and Lines

Go is based on capture by topological enclosure of regions following a 4-cell von Neumann neighbourhood rule based on a player (black) enclosing a region the immediately adjacent squares above and below and to either side (see insets). Topological connectedness thus becomes quantized and discrete. The winner is the player who has captured most of the board once the uncontested sites where a player dominates have been filled in. Critical is the idea that the game depends on topological reasoning although the moves are discrete on a discrete grid. The Go state space is huge. There are an estimated 4.63×10^{170} possible positions on a 19x19 board [22].

Blokus uses all geometrical combinations of four colours of piece up to pentaminos to dominate the board by building interpenetrating fractal trees, connectivity being maintained through adjacent corners. A given player playing in pieces of a given colour first builds from a corner of the board. Pieces of the same colour can only be placed corner to corner, so the corners of each piece constitute future sites of fractal growth. All the pieces played by a given player thus form a connected graph. Pieces of differing colours can meet on edges, so a player can fill spaces left by other colours. The graphs of two players can also cross one another and both be connected, an odd but obvious property of the discrete connectivity rule based on touching corners, which parallels the difference between discrete dynamics and continuous vector fields. The game begins with players building towards the centre and endeavouring to build fractal dendrites, which will permeate as many regions as possible in the face of blocking moves by the opponents, using the power of corners and the capacity to block opposing players vacant options. As the end-game unfolds, skill moves from strategies of fractal growth, in which the fractal-forming power of corners and the defensive blocking potential of edges is key, towards careful geometrical placing of the remaining pieces to play out. The game thus combines the fractal topology of discrete path connectedness with a geometrical tiling finale.

Both games illustrate the subtleties of the way discrete board games can give rise to implied topologies, despite appearing to be purely geometrical, or abstract strategic in nature and give a conceptual illustration of how quantization can affect classical properties of continua in a way which hints at similar properties of quantum transformations.

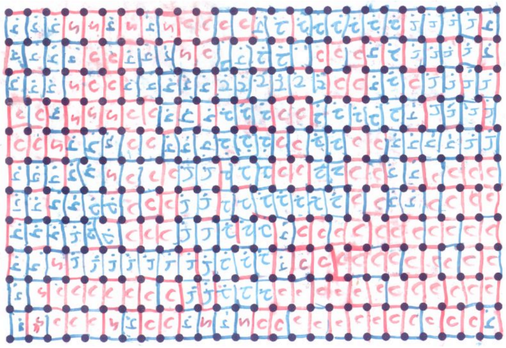


Fig 4.3b Dots and Lines final configuration

Dots and Lines, while an apparently simple filling-in game undergoes a complex phase transition from a ‘gas’ to a ‘solid’ similar to percolation[23]. Each player is allowed another turn each time they claim a square by completing the fourth side. At first players make defensive moves avoiding the opponent gaining squares, but this results in a crystallization of many edges to a point of self-organized criticality, when all moves will result in escalating cascades of claimed squares.

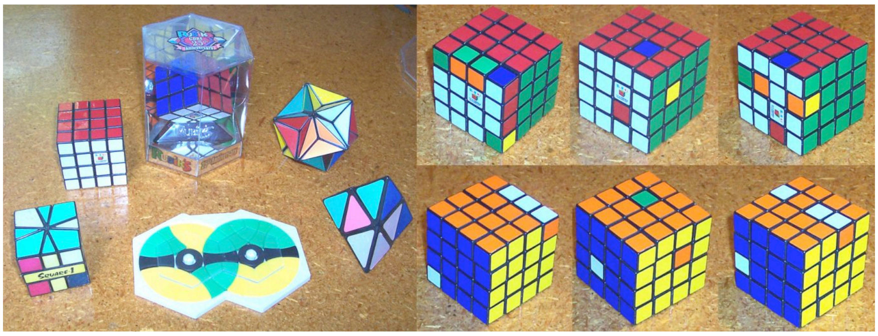


Figure 4.4a Left: Variants of the Rubik cube have a variety of geometrical and planar shapes, although all depend for their solution on the common algebraic method of examining the symmetries of a commutator of two rotations.
Right: The 4x4x4 Rubik revenge cube has three sets of symmetries generated by commutators.

Example 4.4: Rubik type Algebraic-Geometrical Puzzles

The Rubik cube is the most popular puzzle of al time having absorbed 1/8 of the world’s population. Rubik type puzzles, stemming from the initial 3x3x3 Rubik cube (centre left) now come in a wide variety of geometrical forms including cubes of various types, pyramids, stellated dodecahedron (Alexander star right left) cube-octahedron, truncated rhombic dodecahedron and planar configurations. All of them involve skill with mental rotation and keeping track of interacting rotation processes geometrically and each has a unique ingenious mechanical construction supporting its rotation set. However, all these puzzles are subject to a single basic algebraic strategy to complete the solution – examining the symmetries possessed by the commutators of two non-commuting rotations, which are the compound movements naturally closest to the identity *I*.

For the simpler puzzles there is only one commutator type $C = aba^{-1}b^{-1} \neq I$, which in the case of the original Rubik cube permutes 3 edges and exchanges two pairs of corners, at the same time rotating the corners. C^2 then permutes only edges and C^3 only corners. A succession of moves of the type $M = TC^nT^{-1}$, where T is a transformation moving the required edges, or corners, to the appropriate positions for C can then solve each of the puzzles straightforwardly. The 4x4x4 Rubik revenge cube has three possible types of commutator involving inner and outer rotations, some of which can generate odd permutations. These non-commuting operators give an everyday insight into the more mysterious non-commuting processes mediating quantum uncertainty of spin angular momentum in different directions, which form one basis for quantum uncertainty.

Although a puzzle like the revenge cube takes only 50 or so moves to reach the solution from an arbitrary state, the total number of states is huge: There are 8 corner pieces with 3 orientations each, 24 edge pieces with 2 orientations each, 24 centre pieces, giving a maximum of $8!\cdot 24!\cdot 24!\cdot 3^8\cdot 2^{24}$ positions. This limit is not reached because: (a) The total twist of the corners is fixed [3] (b) The edge orientation is dependent on its position [2^{24}] (c) There are indistinguishable face centres [$4!^6$] (d) The orientation of the puzzle does not matter [24]. This leaves $7!\cdot 24!\cdot 24!\cdot 3^6/4!^6$= 7,401,196,841,564,901,869,874,093,974,498,574,336,000,000,000 or $7.4\cdot 10^{45}$ positions, illustrating that the graph of states of realizable puzzles can be huge, comparable with the number of electrons in the universe or even the number of potential string theory candidates, thus forming an oracle for complex systems at the frontier of human knowledge.

Erno Rubik was initially interested simply in the mechanics of how to construct a cube in which the ‘subcubes’ would rotate, but having done so he discovered the implications: “"It was wonderful, to see how, after only a few turns, the colors became mixed, apparently in random fashion. It was tremendously satisfying to watch this color parade. Like after a nice walk when you have seen many lovely sights you decide to go home, after a while I decided it was time to go home, let us put the cubes back in order. And it was at that moment that I came face to face with the Big Challenge: What is the way home?"

Despite their many forms, most of these puzzles are regular in the sense that every position permits the same moves. An intriguing exception is the Square-1 puzzle, which admits a variety of irregular conformations, which have varying sets of moves into and out of these states, some of which permit odd permutations. Thus although the transformations form a group in which every element has an inverse, the group and the ensuing graph of puzzle states is highly irregular.

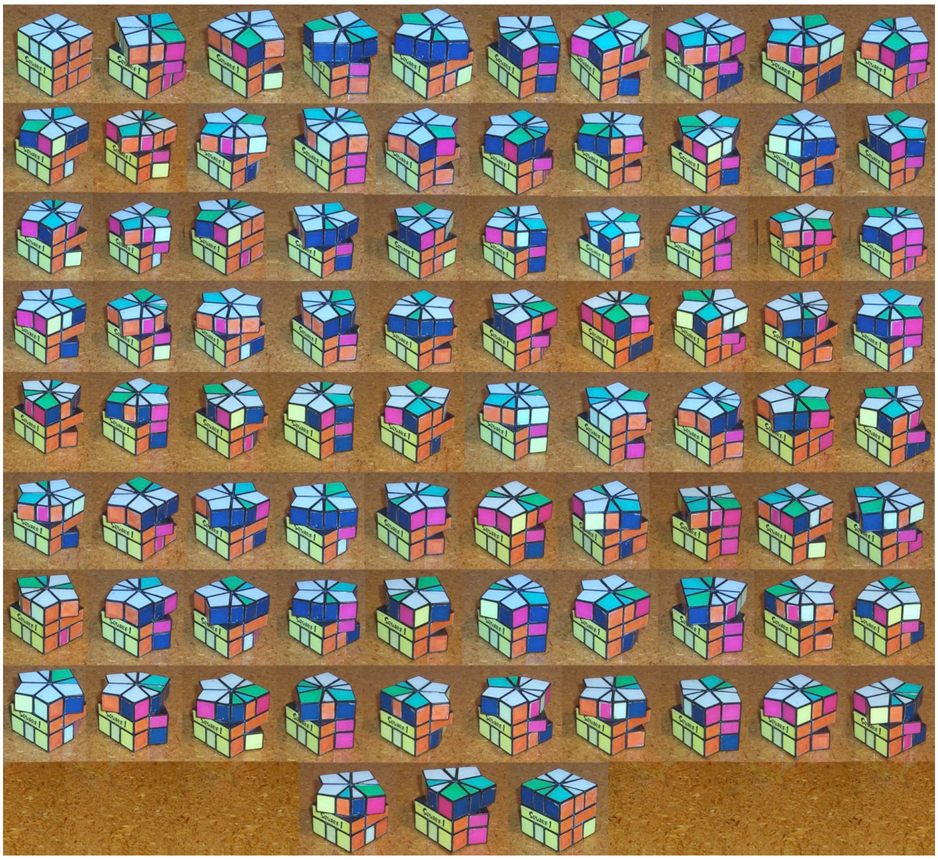


Figure 4.4b: The orbit of states of (t-) repeated 82 times on Square-1

The repeated operation {(t–)(82)}, (where ‘t’ rotates the top clockwise to the next flip position and ‘-’ is a flip of the right hand side of the puzzle) visits many such states before returning to the cube. The full periodicity back to the completed cube is 4 x 82 = 328 since the permutation of the corners (1-8) and edges (a-h) is (1728)(ag)(cd). The corresponding sequence {(tb–)(8)} returns to the cube permuted by (148)(263)(57)(afbecgdh), having an orbit length of 3x8x8=192.

5: State Space Graphs and Strategic Topologies

Virtually every puzzle, whether logical, conceptual, arithmetic, geometric, topological or strategic is navigated by a human subject in an abstract journey from beginning state to solution, through many possible cul-de-sacs in a journey which takes the form of a connected path along the nodes of a graph of states which constitutes a maze of intermediate positions. This is a process akin to a journey through the wilderness in which various conceptual attributes essential for solving the puzzle can point the way to the solution much as topographical signposts or at least sensibly reduce the huge space of possibilities to a feasible number of options.

Although every solvable puzzle is path connected, the form and size of the state graphs can vary extremely. A regular graph with a standard set of moves, such as the Rubik revenge cube, can have a huge state space. By contrast state spaces in which the transitions are complex, irregular may have a much smaller state space, despite being of non-trivial difficulty. We now examine several different types of puzzle to investigate the common topological thread involved in navigating a connected path from starting point to solution.

Example 5.1: Who Own the Zebra?

This logical puzzle sometimes incorrectly attributed to Einstein consists of a series of logical statements associated with five colours of house, five nationalities, five drinks, five pets and five brands of cigarette. The solution to the puzzle is most easily performed by making a table of the items, and then analyzing the logical statements, to specify successive entries of the table, branching to deal with contingencies as little as possible.

Given the statements listed below, we are asked: "Who owns the zebra?" and "Who drinks water?

1. There are five houses.
2. The Englishman lives in the red house.
3. The Spaniard owns the dog.
4. Coffee is drunk in the green house.
5. The Ukrainian drinks tea.
6. The green house is immediately to the right of the indigo house.
7. The Old Gold smoker owns snails.
8. Kools are smoked in the yellow house.
9. Milk is drunk in the middle house.
10. The Norwegian lives in the first house.
11. The man who smokes Chesterfields lives in the house next to the man with the fox.
12. Kools are smoked in the house next to the house where the horse is kept.
13. The Lucky Strike smoker drinks orange juice.
14. The Japanese smokes Parliaments.
15. The Norwegian lives next to the blue house.

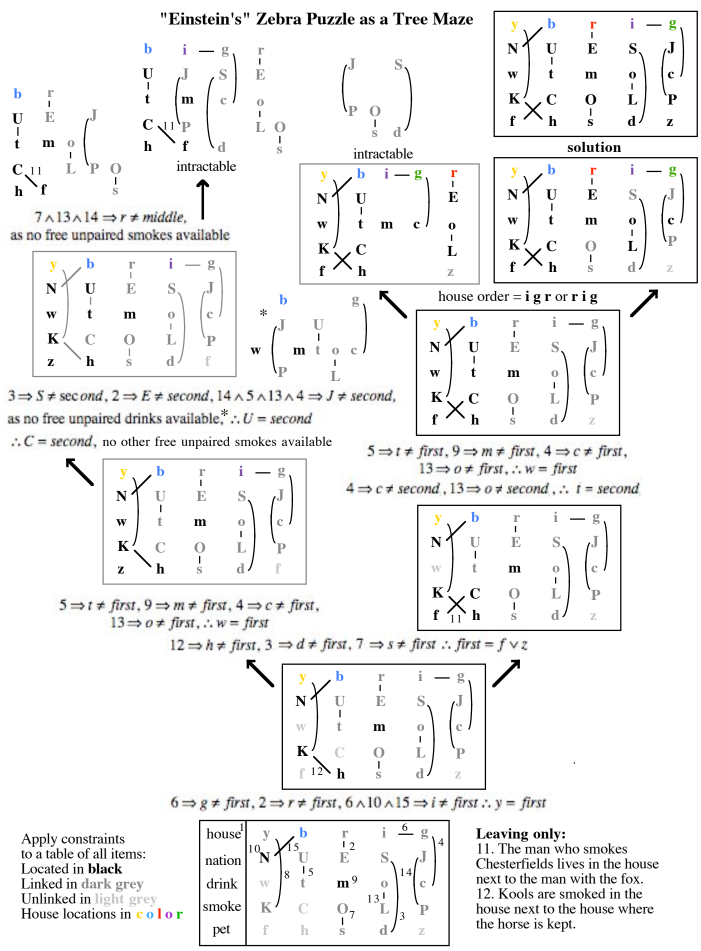
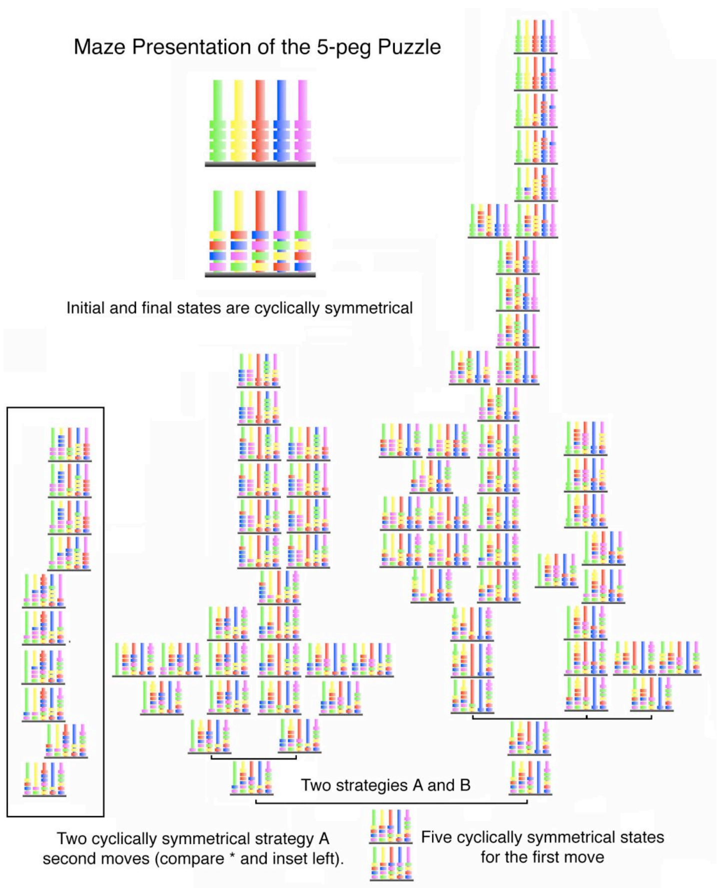


Figure 5.1: "Who Owns the Zebra?" portrayed as a strategic maze of puzzle states.

Figure 5.1 shows a decision-making tree maze for the puzzle, which is conveniently tabulated in the same way as the solution for ease of reading. Initially all but two of the statements are processed and incorporated into the table in terms of links between categories which determine the relative positions of the linked items. Although the puzzle is non-trivial its decision making state graph is a tree with only a few nodes once the logical statements, which can be processed simultaneously are grouped into one step.



Because there are many ways of prioritizing the statements and in which order to deal with the categories, a human subject will frequently navigate a version of the tree, adding one or two extra assumptions, only to find they have reached an intractable position, returning to the trunk of the tree, or a variant of it, to try again, releasing some or all of the assumptions which led to intractability. In doing so, they are navigating a logical space, abstractly akin to making a path-connected journey in a topographical landscape.

Figure 5.2: Maze Presentation of example 5.2

Example 5.2 The Five Peg Puzzle

Figure 5.2 shows the maze for a puzzle in which one or two top rings can be moved, but only on to a ring, or empty peg of the same colour as the lower one. Again there are only a limited number of states because many moves rapidly lead to intractability. The graph now has trivial loops but is unidirectional upward because the moves are not reversible. Again the subject is traversing a conceptual territory, which can be described as a path-connected region.

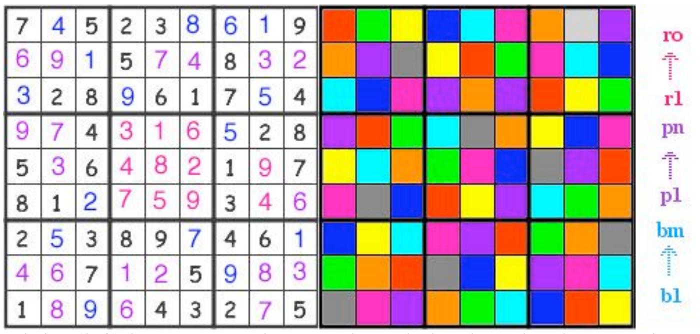
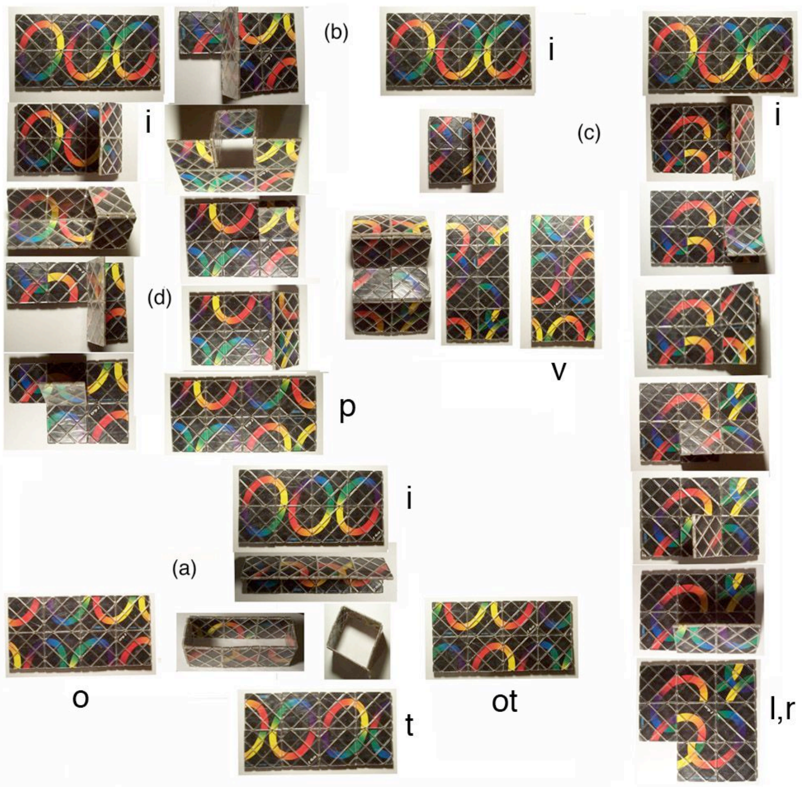


Figure 5.4 Elementary Sudoku (left) has no numeric operations, being based only on each row, column and sub-square having distinct entries, as the colour-coded version (centre) shows. Black: initial puzzle. Blue: clues from horizontal and vertical lines. Purple: clues using sub-squares. Red: final solution. This puzzle can be solved without contingencies and thus has a state space consisting of a meander maze - the unique path from start to solution (right). Similar colour codings are used to depict Cayley tables[24], which also have distinct entries in every row and column.

Example 5.4 Elementary Sudoku

Elementary Sudoku illustrates the ultimate simplicity of state space structure. Although it presents as a numeracy puzzle, it is simply a category-matching puzzle, as illustrated by colour coding, requiring only that every row, column and sub-square has nine distinct entries. Because all the numbers can be found simply by filling in numbers determined in sequence from the provided clues, the state graph is just a line, as in a meander maze figure 1.2, as illustrated right in figure 5.4. Advanced Sudoku however introduces fewer clues, requiring testing contingencies, and hence has a simply-connected tree maze as in example 5.1.

Figure 5.5a: Four sequences of geometric moves in the Rubik 3-rings puzzle.



Example 5.5 The Rubik Three Rings Puzzle

The Rubik 3-rings puzzle consists of a set of eight diagonally grooved plates held together by nylon strings woven over three successive plates in a circuit in overlapping succession, so that the plates can be folded along certain axes joining the plates, changing the way the strings link the plates and creating new puzzle geometries. The aim of the puzzle is to fold the plates into a new arrangement where the three unlinked rings have seemingly impossibly become linked. This is possible because the reverse sides of the plates have pieces of a second image of the three rings linked through one another as shown in the heart shape in the centre of figure 5.5c, associated with an L-shaped geometry differing from the rectangular starting position.

The puzzle presents an intriguing mix of geometrical and topological constraints, the weaving of the strings fixing the geometry of the hinged shapes, within the basic topology of a ring of plates in which some, but never all, of the plates are able to be hinged out of the loop, at least temporarily.

The weaving of the strings itself presents an interesting topological puzzle, which enables the eight rings in the rectangular configuration to be transformed in every possible way that retains their overall ring structure. There are 8 clockwise permutations of the plates, two directions of orientation around the ring and four orientations the square plates can adopt collectively. This gives 8 x 2 x 4 = 64 possible states of the rectangle, however we need to divide this figure by 2, since moving four steps round the ring rotates the whole rectangle through 180^{o,} if the top row is coded *abcd* and the bottom row is an inverted *efgh*. The way the strings are woven enables all of the 32 possible states to be reached, although this might seem impossible from the way they are woven.

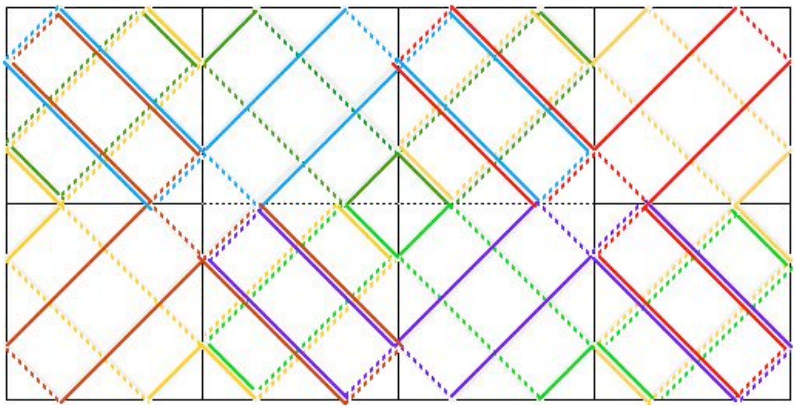
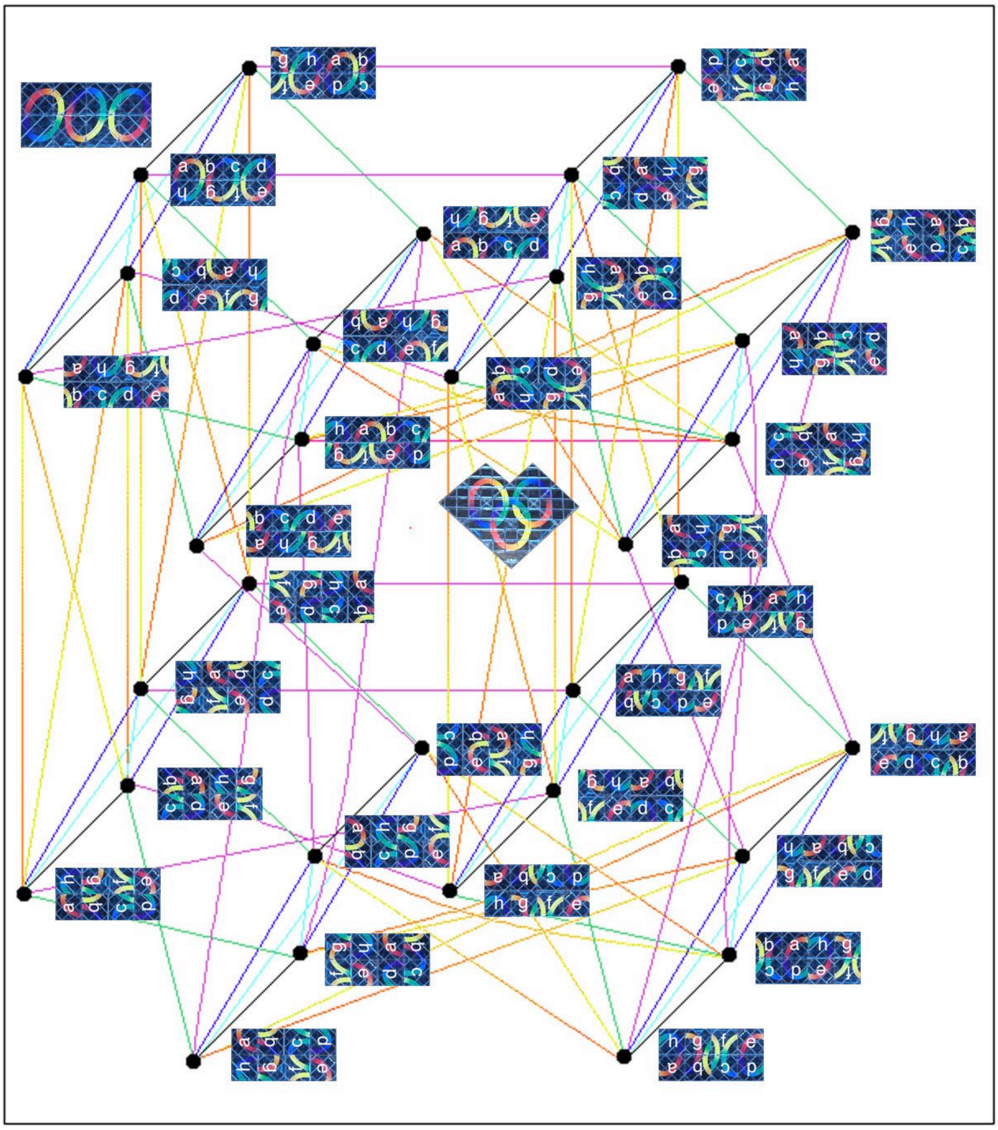


Figure 5.5b Colour-coded weaving of the 8 closed loop string pairs holding the 8 3-ring puzzle plates in a loop.

In figure 5.5b is shown the weaving of the 8 looped string pairs in a 4x2 arrangement, which remain unknotted throughout, although alternate plates have a double winding, with strings linking to 2 adjacent plates, spanning 3 in all. There are no vertical connections in the centre four plates, so the 8 plates form a ring.

To solve the puzzle requires negotiating a series of geometrical transformations, some of which lead to cul-de-sacs, however there are four sequences of transformations illustrated in figure 5.5a, which lead to a rearrangement of the rectangular arrangement. In (a) the ring is folded and becomes a literal ring of 8 plates, which can be unfolded to form four rectangular states involving 3 transformations from the identity. In (b) the plates can be folded together above and then unfolded in the vertical direction from below, effectively rotating the plates through 90^{o}. In (c) a sequence of moves takes the rectangle to a scrambled form of the L or heart-shape of the final solution, which can then be refolded from the other side of the L to gain a different transformation of the rectangle. There is a mirror image of this entire sequence, which forms an inverse transformation. Finally in (d) there is another move, which results in a new set of configurations, resulting in 7 transformations in all.

Figure 5.5c: The state space graph of the rectangular configurations presented as a non-commutative graph assembled on the 4-D hypercube of 32 states.



Because the geometry of these moves is complex, we have a non-trivial puzzle which has a state space graph which has only 32 nodes corresponding to the 32 possible transformations of the 8 plates above, so we see another example of the trade off between individual transition complexity and state graph size.

To analyze the state space graph, the seven possible transformations of the rectangular configuration, a Matlab simulation was made of each of the geometrical transformations and this was used to check the definition of each of the 7 transformations arising from each node. The result is shown in figure 5.5c.

Although this is a richly interconnected graph with a large number of loops, navigating from one position to another is still difficult because several of the operations fail to commute in diverse ways, causing operations performed out of order to arrive at unfamiliar destinations.

The seven transformations are colour-coded and the state of each node is illustrated and coded using the abcdefgh notation above. Each of the pairs of edges forming a parallelogram in the graph commute while the others do not. Each of the transformations are self-inverses, except for red-yellow shaded ones passing through the heart-shape intermediate, whose two forms are mutual inverses. There is a corresponding 32 node state graph for the heart-shapes, each of which is connected to two rectangles through inverse transformations, two of which emerge from each rectangle.

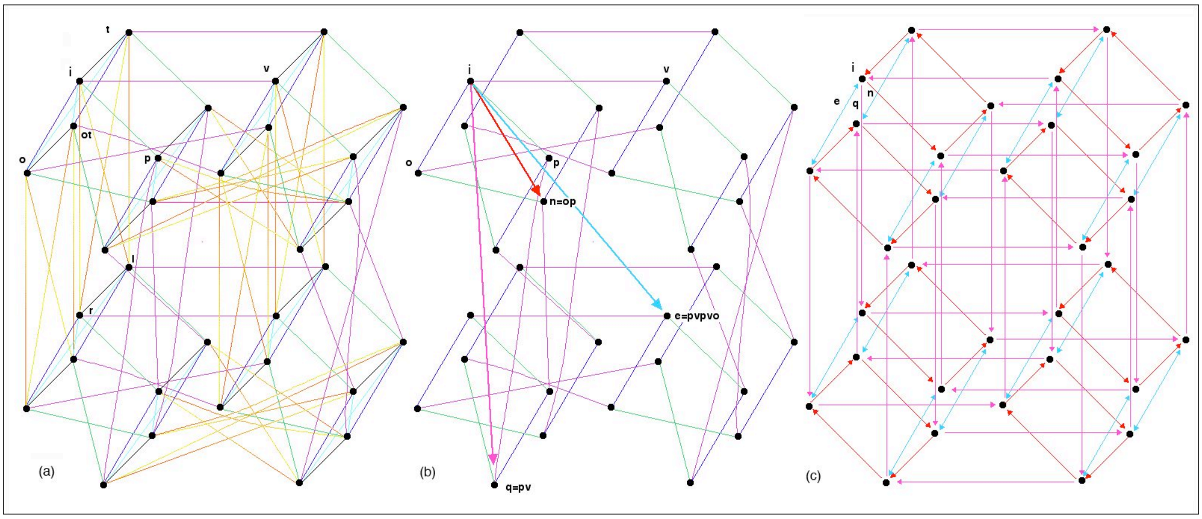


Figure 5.5d Stages in disentangling the transformation group. (a) Graph of the 7 geometric operations *t*, *o*, *p*, *v*, *r*, *l* and *op*. (b) Reduced graph with 3 generators o, p, v and defining new generators *n*, *q* and *e*. (c) Rearrangement of vertices using the new generators results in a non-commutative hypercubic Cayley graph.

To decode the actual 32-member group[25], first we eliminate redundant operations. Tracing the connections in figure 5.5d(a) we can see immediately that three of the key geometrical operations including the simplest (*t*) and the ones that pass through the heart solution (*l* and *r*) are composites of the others: *t*=*opop*, *l*=*povp*, *r*=*opvp*.

Removing these yields the graph in (b) and a group with a presentation[26]:

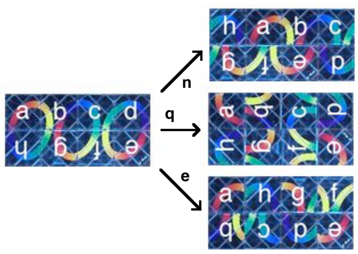
$$G = \left\{ o, p, v : o^2 = p^2 = v^2 = i,\ (op)^4 = (ov)^4 = (vp)^4 = i,\ opop = ovov,\ ov = opvp \right\} \qquad [5.5.1]$$

In so doing, we have eliminated the very geometrical transformations that enabled us to get to the heart shaped solution. We should note that a similar description could be mounted of all the transformations of the 32 hearts.

Examining the symmetries of the plates in figure 5.5c however, we can easily see that more natural operations are available which are composites of *o*, *p*, and *v* but represent fundamental symmetries of the rectangle.

We define three new transformations of the rectangle (RH composition):

1. *n*=*op* Moves all the plates cyclically right by one step, rotating plates 180° when they move around the end of the ring. Four such moves rotate the rectangle through 180° leaving the puzzle unchanged.
2. *q*=*pv* Rotates alternate plates 90° clockwise and anti-clockwise.
3. *e*=*pvpvo* Reverses the orientation of the ring of 8 plates leaving the top left and bottom right plates unchanged.



Given these generators, we can present the group *G*, with center[27] $\{i, nn = t, qq, nnqq\} \cong C_2 \times C_2$, as:

$$G = \left\{ n, e, q : n^4 = q^4 = e^2 = i,\ qe = eq,\ qn = nq^{-1},\ ne = en^{-1} \right\} \qquad [5.5.2]$$

We can then rearrange the vertices to reflect the symmetries and arrive at a non-commutative, hypercubic Cayley graph[28] for the transformation group as in figure 5.5d(c). If we use the notation $A \times_{-1} B$ for a semi-direct product[29] action by inverting elements: $ba = ab^{-1}$, then, from the relations, *G* can be characterized by: $G \cong D_4 \times_{-1} C_4 \cong (C_2 \times_{-1} C_4) \times_{-1} C_4$ since $eq = qe = qe^{-1}$ and $qn = nq^{-1}$, where D_4 is the dihedral group of transformations of the square and C_n is the cyclic group of order *n* (e.g. of integers modulo *n* under addition).

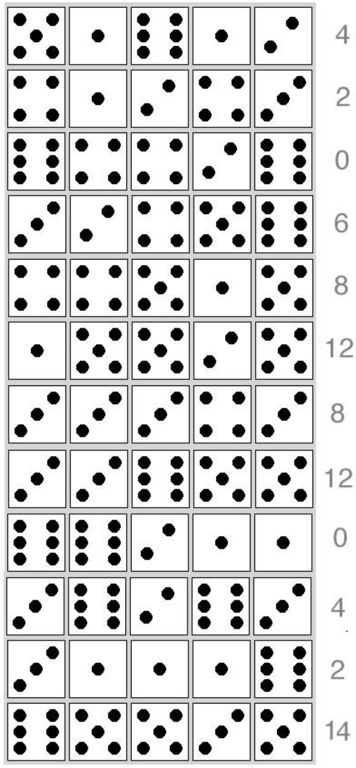


Figure 5.6 A sample of die throws from “Petals Around the Rose”

Example 5.6 Petals Around the Rose

“Petals Around the Rose”[30] is a puzzle that is famous for its account of Paul Allen and Bill Gates introduction to it in a crowd returning from a computing conference in 1977, in which Bill was the last active player in the group to discover the rule. The game has only two clues. One is that the answers are all even, which becomes obvious after a few throws, and the other is “Petals around the Rose”, which is significant. No one is supposed to reveal anything more than the answer to a throw – never the rule itself.

The problem to be solved, rather than one of deductive thinking as in the zebra puzzle is one of lateral thinking, faced with a seemingly irregular rule. The state space of the puzzle now consists of all the lateral shifts of thinking the subject might imagine, so it cannot be defined precisely in the way the previous examples were. There are a great variety of rules which could be applied, some involving adding or multiplying the values on the faces, others counting how many die of a given value appear, others dealing with the geometry of the faces, the way the dice fall or the order of them in sequence, but each of these conjectures form part of the topography of the state space which the subject explores till they see a contradiction, until eventually they discover the rule, which for convenience I will print upside down in light grey below, so you can read it only if you can’t deduce it from the instances in figure 5.6.

Critical to the irregularity is that the rule uses only partial information from the dice. This information is highlighted both by the high scores and the very low scores, which are over-represented in the list in the figure in the interests of quick analysis.

The "rose" is the center dot on any face that has one (i.e. 1, 3, and 5) and the "petals" are all dots around it -- the 1 face has no petals, the 3 face has two petals and the 5 face has four petals. The die faces without center dots (i.e. 2, 4, and 6) do not count.
Counting the total petal dots yields that round's answer.Alternately, one may sum the faces of odd dice, and then subtract the number of them.
Thus, in a roll of 1, 3, 4, 5, and 3, the odd faces sum to 12. There are four odd dice, so the solution is 12-4, or 8.

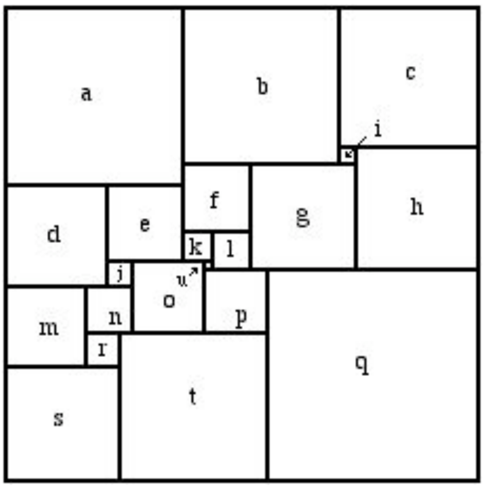


Figure 5.7a The unique simple perfect square of order 21 (the lowest possible order).

Example 5.7 Squaring the square and Magic Squares

Not all puzzles involve a state space. Some are better solved in one step, or a single defined process, e.g. by defining a system of equations. One such example is squaring the square[31 32], where we are asked to find the relative dimensions of the tiled unequal squares fitting into a single large square in figure 5.7.

This is an ideal candidate for using symbolic manipulation to take the boredom out of the algebra. We first investigate the geometry and compare a series of vertical and horizontal side lengths until we have generated enough equations for a unique solution, and set the smallest square to a suitable base number.

The Matlab symbolic toolbox provides an ideal solution platform:

```
syms a b c d e f g h i j k l m n o p q r s t u
S=solve('l=k+u,f=k+l,g=f+l,h=g+i,c=h+i,b+g=c+h,o=p+u,o=j+n,a=d+e,b+c=f+g+h,s=m+r,t=
r+s,q=p+t,a+b+c=d+e+f+g+h,a+b+c=s+t+q,a+b+c=m+n+o+p+q,a+d+m+s=a+e+j+n+t,a+d+m+s=c+h
+q,c+h+q=a+e+o+t,a+e+o=b+f+l+p,b+i=f+g,e+k=j+o+u,o+e=d+n,d+j=m+n');
C=struct2cell(S);
u=2;
for i=1:21
    fprintf('%c=%2.0f ',char(96+i),eval(C{i}));
end
a=42 b=37 c=33 d=24 e=18 f=16 g=25 h=29 i=4 j=6 k=7 l=9
m=19 n=11 o=17 p=15 q=50 r= 8 s=27 t=35 u=2
```

However such puzzles are neither common, nor as popular as those which require a conceptual hunt through a space of possibilities, and in this case the problem is unique, being the only simple perfect square of order 21 (the lowest possible order), discovered in 1978 by A. J. W. Duijvestijn[33].

8	1	6
3	5	7
4	9	2

Figure 5.7b Lo Shu the unique 3x3 magic square is associative and generated by the Siamese method..

To explore the problem of puzzle generation in numeric puzzles we can explore the problem of magic squares[34]. A magic square is a square array of numbers, consisting of the distinct positive integers 1, 2, ..., arranged such that the sum of the numbers in any horizontal, vertical, or *main* diagonal line is always the same number, known as the magic constant $\frac{n}{2}\left(n^2+1\right)$. The unique 3x3 square was known to the ancient Chinese as Lo Shu. This is also associative if pairs of numbers symmetrically opposite the centre sum to n^2+1. If *all* diagonals (including those obtained by wrapping around) of a magic square sum to the magic constant, the square is said to be a panmagic square also called a diabolic square.

It is an unsolved problem to determine the number of magic squares of an arbitrary order, but the number of distinct magic squares (excluding those obtained by rotation and reflection) of order 1-5 are 1, 0, 1, 880, 275305224, and an estimate of order 6 is 1.77×10^{19} using Monte Carlo simulation and methods from statistical mechanics. The number of distinct diabolic squares of order 1-5 are 1, 0, 0, 48, 3600.

Given the unbounded number of solutions one would expect there exists simple regular algorithms for generating magic squares and this is the case. The Siamese method consists of placing a 1 anywhere and placing 2, 3 etc. successively up the right hand diagonal (vector (1,1)) moving one down (break vector (0,-1) if we hit a filled square. Lo Shu in figure 5.7b can be seen to be generated in this way. The Siamese method will also generate diabolic magic squares of order 6*k*±1 with vector (2,-1) and break vector (1,1).

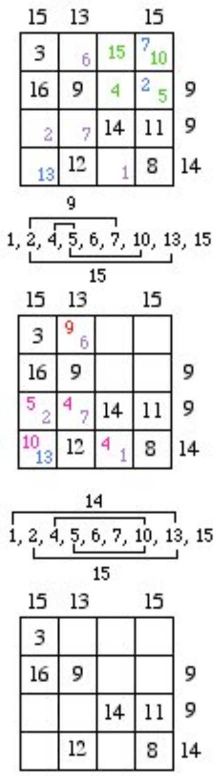


Figure 5.7c A sample 4x4 square puzzle made by removing magic square entries, has a simple tree maze with two branch points, corresponding to contingencies in the bottom left and top right corners.

Magic squares can be used to generate Sudoku-like puzzles with state space tree mazes of varying complexity. In figure 5.7c is shown a sample 4x4 diabolic magic square in which over half the entries have been omitted. The entries outside the square give the remainder when the existing entries are subtracted from the magic constant of 34. In the first stage the bottom- left entry is used to compare information from its row and column. This implies a corresponding set of contingencies in linked rows and columns leading to an impasse for one (the 9 in position (1,2). This information can now be used to perform the same analysis for the top-right entry leading to the solution. Once again the numeric puzzle leads to a path-connected graph, in this case a tree with two branch points, giving the puzzle an underlying topological basis.

Because the number of possible magic squares grows so rapidly, increasing the size of the square and reducing the number of given entries can rapidly lead to too many contingencies to make an interesting and ‘doable’ puzzle because of the load of multiplying contingences and the repetitious simple arithmetic involved.

Example 5.8 2-D and 3-D Tiling with Polyminoes

While some puzzles have one solution, which might be solved, like squaring the square, by a system of equations, an abstract proof, or a single algorithm, others have many possible solutions, often with their own internal irregularities, which require a brute force computational approach to find all the variations. One of the most persistent and intriguing types of puzzle to many people are geometrical tiling puzzles constructed out of systematic geometric variants, such as pentaminoes (all 12 configurations of 5 attached cubes in 2-D), the pieces of the soma cube (all 7 non-linear pieces of composed of 3 or 4 cubes in 3-D) and the Kwazy quilt made of all combinations of circles stellated with up to six regular apices.

Figure 5.8a Anti-clockwise from top: 6 variants of the soma cube, viewed front and back, 6 variants of the 'Lonpos Pyramid', one of only 2 possible 3x20 pentamino solutions, 'Kwazy Quilt', and compound happy cube and hypercube illustrate tiling puzzles with multiple solutions.

The soma cube was invented by Piet Hein[35] the scientist, artist, poet and inventor of games such as hex, during a lecture on quantum mechanics by Werner Heisenberg. There are 240 essentially distinct ways of doing so, as reputedly first enumerated one rainy afternoon in 1961 by John Conway and Mike Guy.
However, if we count the internal symmetries of individual pieces within themselves, i.e. $3\times 2^5 = 96$ and the $6\times 4\times 2 = 48$ symmetries of the whole cube we arrive at $96\times 48\times 240 = 1105920$. This might be compared with the maximum number of distinct, possibly non-tiling arrangements of the pieces in space $7!\times 24^7 / 96 = 2.4\times 10^{11}$. Because a subject assembles a cube using less tractable pieces first, it is relatively easy to find a solution, and to navigate in the maze of solution space using geometrical intuition using as many back step as necessary to retreat from cul-de-sacs near completion. A variety of other geometrical shapes can also be made with the soma pieces, having varying degrees of constraint and hence difficulty.

Likewise the 'Lonpos Pyramid' uses a subset of spherically-based 2-D polyminoes of sizes 3, 4 and 5 to build a pyramid, as well as rectangular solutions. Although the pieces are planar, the pyramidal solutions involve interlocking pieces aligned horizontally, vertically and obliquely. Since most are horizontal it is generally easier to solve from the apex of the pyramid, which places strong local constraints on the pieces to be used.

The 12 2-D pentaminoes, known from the 19th century, are capable of tiling several rectangles of area 60 units, as well as other shapes, such as using 9 to tile versions of the individual pieces expanded 3 times in size (45 units area). The number of rectangular solutions are: 6×10 (2339), 5×12 (1010), 4×15 (368), 3×20 (2). This might be compared with something like 3.2×10^{16} independent orderings and orientations of the 12 pieces. The rectangular puzzles each have similar difficulty, despite the varying number of solutions, because the narrower rectangles place more constraints on the feasible partial tilings.

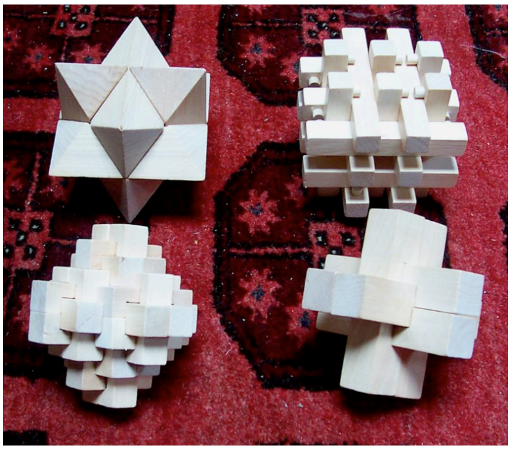


Figure 5.8b Four wooden interlocking puzzles.

The popularity of such puzzles with both adults and children, including their variants in wood puzzles (left) that generally have only one way of being assembled, illustrates a strong theme involving the geometry of mental rotation, the topology of navigating a path in abstract solution space, and a preference for dealing with mathematical problems which have a strong sensory basis, are capable of direct manipulation and promote lateral thinking, to open unperceived avenues and avoid tunnel vision, as well as deductive reasoning. These themes all support a linkage between puzzles and gatherer hunter skills, which have evolved over long epochs and stand diametrically opposed to the dominance abstract linguistic-based axiomatic manipulations have in proofs in classical theoretical mathematics. The gulf between these perspectives becomes ever more acute in an era when pocket calculators and laptop computers are making redundant many of the arithmetic skills of mental calculation we have come to assume go hand in hand with civilization.

Example 5.9 Peg Solitaire as a large State Space with Internal Symmetries

Table 5.9 Successive board positions in peg solitaire[36,37]

Holes	Moves	Positions	Winning	Terminal	Tot Positions	Dead Ends
1	0	1	1		1	0
2	1	1	1	0	4	0
3	2	2	2	0	12	0
4	3	8	8	0	60	0
5	4	39	38	0	296	0
6	5	171	164	0	1,338	0
7	6	719	635	1	5,648	32
8	7	2,757	2,089	0	21,842	0
9	8	9,751	6,174	0	77,559	0
10	9	31,312	16,020	0	249,690	0
11	10	89,927	35,749	1	717,788	280
12	11	229,614	68,326	1	1,834,379	31,920
13	12	517,854	112,788	0	4,138,302	0
14	13	1,022,224	162,319	5	8,171,208	386,416
15	14	1,753,737	204,992	10	14,020,166	1.82E+07
16	15	2,598,215	230,230	7	20,773,236	5.24E+07
17	16	3,312,423	230,230	27	26,482,824	5.69E+08
18	17	3,626,632	204,992	47	28,994,876	3.64E+10
19	18	3,413,313	162,319	121	27,286,330	3.80E+11
20	19	2,765,623	112,788	373	22,106,348	8.52E+12
21	20	1,930,324	68,326	925	15,425,572	1.96E+14
22	21	1,160,977	35,749	1,972	9,274,496	3.72E+15
23	22	600,372	16,020	3,346	4,792,664	5.31E+16
24	23	265,865	6,174	4,356	2,120,101	6.05E+17
25	24	100,565	2,089	4,256	800,152	4.41E+18
26	25	32,250	635	3,054	255,544	2.16E+19
27	26	8,688	164	1,715	68,236	8.25E+19
28	27	1,917	38	665	14,727	1.36E+20
29	28	348	8	182	2,529	2.11E+20
30	29	50	2	39	334	1.05E+20
31	30	7	1	6	32	1.63E+19
32	31	2	1	2	5	8.17E+16
		23475688	**1679072**	**21111**	**187636299**	**5.77117E+20**

Peg solitaire has a long and colourful history, being spuriously attributed both to native Americans and to a French aristocrat imprisoned in the Bastille, but can be specifically traced back to the court of Louis XIV in 1697, from when its repeated representation in art shows it had wide popularity. In the classical game, the board is filled with pegs except for the central position, and the aim is by jumping over and removing successive pieces, to end with a single peg remaining in the centre. The English board forms a cross comprising 33 holes, as shown in figure 5.9 and admits multiple solutions, but the European version with four extra pegs does not admit a classical solution, so we shall consider the English game, although there are also many puzzle variants.

A brute force attack on the possible number of positions in *n* moves gives the sequence in table 5.9. The total number of reachable board positions is the sum 23,475,688, while the total number of possible board positions is $2^{33}/8 \sim 10^{9}$ when symmetry is taken into account. So only about 2.2% of all possible board positions can be reached starting with the center vacant. 'Tot Positions' ignores the symmetries of board rotations and reflections which are factored out in 'Positions'. Counting successive board positions into a cumulative set of plays, there are 577,116,156,815,309,849,672 or 5.7×10^{20} different complete game sequences, of which 40,861,647,040,079,968 or 4×10^{16} are solutions. Thus although there are theoretically a huge number of solutions, the probability of finding one at random is about 1 in 10,000. Until a player finds a winning strategy, they tend to initially move in a haphazard way, hoping to arrive fortuitously at an end-game they can resolve more easily, and are thus unlikely to find a solution.

Since any jump exchanges 2 pegs and a hole with 2 holes and a peg and the start position exchanges holes and pegs as well, there is a symmetry between start and finish, which means that exchanging pegs and holes and playing backwards from the finish will provide a complementary strategy to the original. This can be seen from the symmetry in the winning positions in table 5.9. One appealing winning sequence first collapses the cross to one move off a smaller central diamond game before closing in with a grand circuit. The complement to this game counter-intuitively removes the centre diamond before the arms of the cross arriving back at the centre.

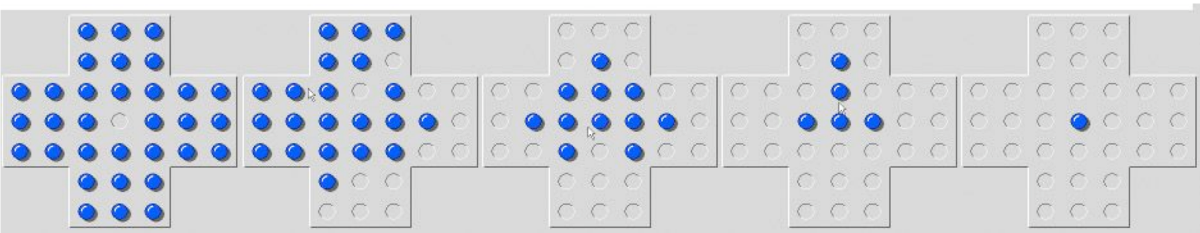


Figure 5.9 Five stages of a winning game of peg solitaire which first reduces the game to one move off a smaller diamond-shaped version of the game before making a grand tour leaving a single T which collapses to the solution. The reverse of this game with pegs exchanged for holes gives a second counter-intuitive solution in which the central diamond is first removed leaving the peripheral parts of the cross, finishing with a move to the centre. Other games win by an amorphous strategy.

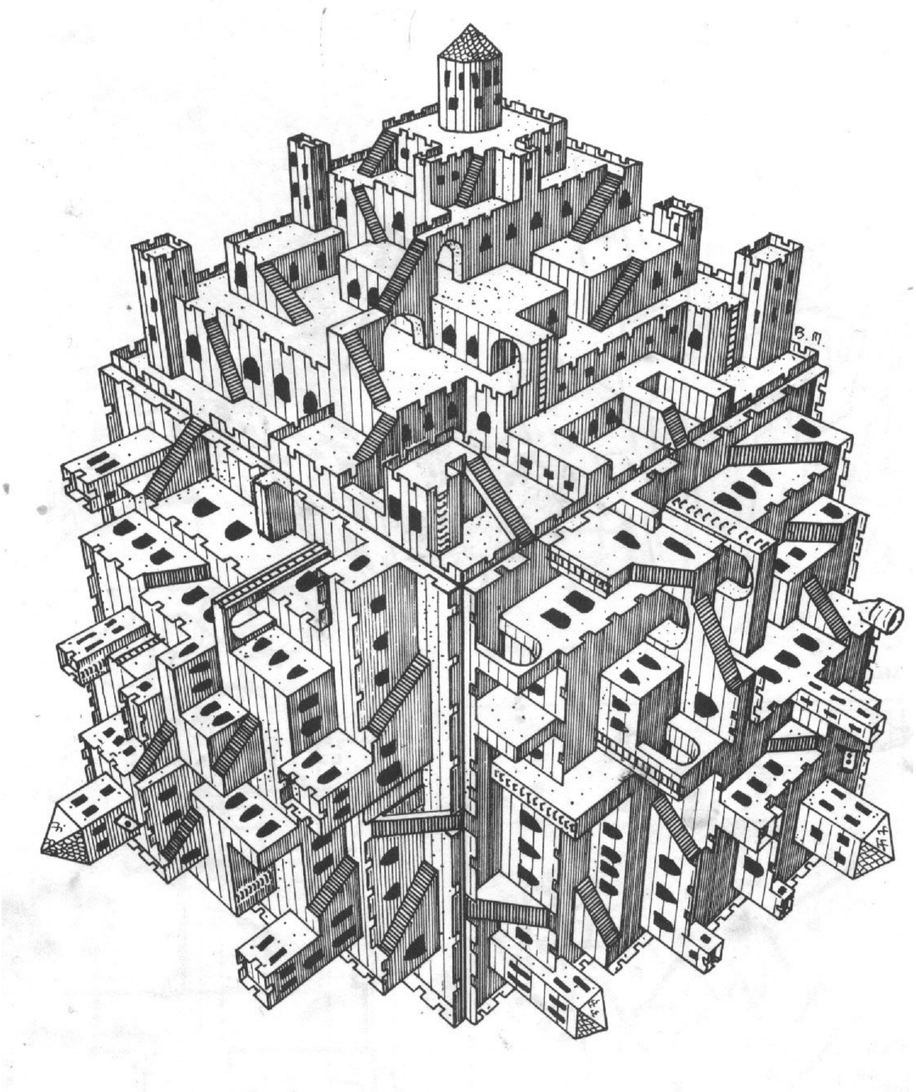


Figure 5.10 Cover maze from Supermazes[38]

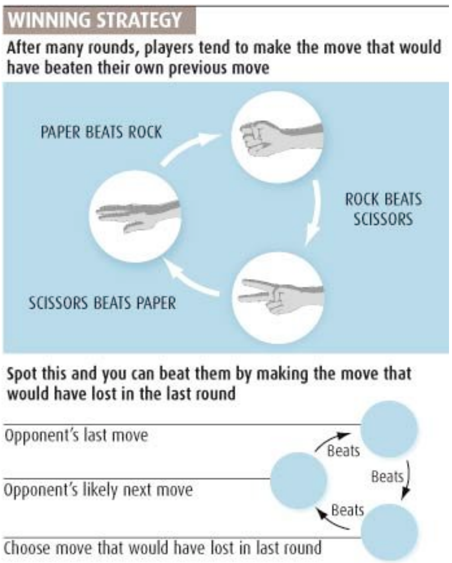
Example 5.10 Mazes as Topological Puzzles

Finally we return to mazes, which, in addition to underlying the solution space of every puzzle, constitute the most ancient and intrinsically topological form of puzzle known. The state space of the maze is precisely the set of positions negotiated in traversing it. Although, as in the example of figure 5.8 they may have a complex topology of overpasses and underpasses in the manner of knot theory, from the subjects point of view this is secondary to the path connecting the start and finish, so the structure of a maze is determined by its path-connected graph, which is trivially a line for a meander maze (figure 1.2), a tree for a simply-connected maze, which can then be traversed however laboriously by a systematic right hand rule following all cul-de-sacs to exhaustion, however in a maze with loops although there may be more than one path, the strategy needs to avoid becoming locked in cycles.

While we are told Theseus had to follow Ariadne's thread to return from slaying the Minotaur, this may have been merely to avoid becoming disoriented in the dark winding passage of the labyrinth, because the image of coins from Knossos from figure 1.2 suggests this, like Roman and floor mazes in many cathedrals was a simple meander maze requiring no choices, but just a long tortuous walk, in stark contrast to the duplicitous topological paths in the wilderness humanity has negotiated, since the dawn of history and the equally elaborate paths in state space we have discovered in analyzing the above puzzles.

Some of theses state spaces like the Rubik revenge with 7.4×10^{45} and even Solitaire with 5.7×10^{20} are huge, but Go with 4.63×10^{170} board positions and Chess with an estimated $10^{10^{50}}$ possible games[39] and between 10^{40} and 10^{120} board positions at the 40th move surely take the breath away and make one realize the Machiavellian theory of the evolution of intelligence, based on social strategic bluffing for sexual favours and personal power in a complex human society of many players has an invincible and convincing ring to it!

Example 5.11 Scissors-Paper-Stone Topological bifurcation as a basis for a complementary strategy space.

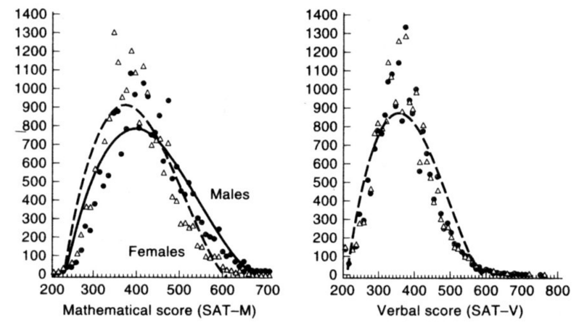


Scissors-paper-stone is a game consisting of an apparently irresolvable cyclic transitive relationship of dominance. There is thus no specific winning set of moves and winning play depends on a bifurcation between two complementary strategies of defense and attack. The defensive strategy is to randomize your moves as completely as possible so the opponent has no pattern they can fix on to take advantage. The complementary attacking strategy is to deduce the opponent's pattern and choose the move that will capture the move anticipated by the pattern. Various statistical deviations in human behavior can also be capitalized on. The choices are commonly skewed rock gaining 36% paper 30% and scissors 34% so a player can take advantage of the skew. Players also tend to pick moves that would have beaten their previous move, so choosing a move which your opponent would have just defeated is a paradoxically winning strategy[40].

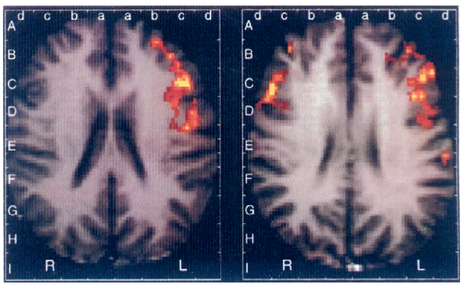
6: The Brain's Eye View of Mathematics

Despite the strides of such techniques as the electro-encephalogram and functional magnetic resonance imaging, research into how mathematics is processed in the brain is still in its infancy. Evidence from cultural and development studies and the effects of brain injury, are rapidly being complemented by research to elucidate the localization in the brain of various aspects of mathematical reasoning, however these have so far dealt mainly with basic level mathematical skills such as raw numeracy – e.g. comparing numbers and tasks such as mental rotation, which are already the fare of psychological experiment.

Figure 6.1 Sex differences in mathematical performance tests are not paralleled in verbal performance tests[41].



Views of the basis of mathematical reasoning in the brain vary widely. At one extreme is the notion that numeracy is a hard-wired genetically based trait[42] located in the left parietal lobe (related to finger counting) and even more basic than language. On a somewhat different tack, Stanislas Dehaene[43] the founder of the triple-code model discussed below, sees both hemispheres being involved in manipulating Arabic numerals and numerical quantities, but only the left hemisphere having access to the linguistic representation of numerals and to a verbal memory of arithmetic tables.



There is some evidence for a genetic basis in mathematical ability, in subtle gender differences in performance at mathematical tasks[44], which is not reflected in language acquisition (figure 6.1) despite the significantly different degree of language lateralization in male and female brains (figure 6.2).

Figure 6.2 Sexual differences in language processing[45].

These mathematics skill differences appear to be real and not just based on differences of educational opportunity. The most comprehensive study published in Science in 1995 found that in maths and science in the top ten percent, boys outnumbered girls three to one. In the top one percent there were seven boys to each girl. By contrast in language skills there were twice as many boys at the bottom and twice as many girls at the top. In writing skills girls were so much better, boys were considered 'at a rather profound disadvantage'[46].

Contrasting a biologically-based view of numeracy are studies which demonstrate cultural differences in the way the same numeracy problem is presented, such as those comparing Chinese and English speakers (figure 6.3). Whereas in both groups the inferior parietal cortex was activated by a task for numerical quantity comparison, such as a simple addition task, English speakers, largely employ a language process that relies on the left perisylvian cortices for mental calculation, while native Chinese speakers, instead, engage a visuo-premotor association network for the same task. Also raising doubts about the genetic basis of numeracy is the discovery of the Amazonian Pirahã [47] who live without any notions of numbers more specific than 'some' and cannot count. This is consistent with the fact that apart from some savant's and geniuses such as Ramanujan[48], most people have a digit span of only seven, and a mental calculation capacity vastly inferior to a simple pocket calculator.

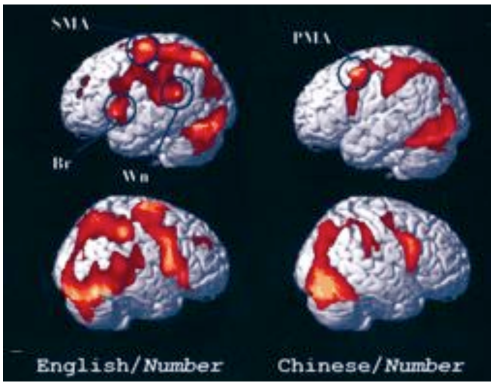


Figure 6.3 Language-based differences in mathematical processing[49].

While some people from Noam Chomsky's generative grammar[50] to Stephen Pinker's "Language Instinct"[51] contend that language is a genetically based evolutionary trait, other models of language[52] see the genetic basis as more generalized and that spoken languages have 'taken-over', as increasingly efficient systems more in the manner of a computer virus through their cultural evolution by colloquial use. This view has support in the much more rapid evolution of languages and the fact that, while we do not know how long ago people first began speaking, written language has only a short human history, consistent with our reading skills being an adaption of more generalized visual pattern recognition systems. Since numeracy and mathematics depend prominently on Arabic numerals, although having a basis in analog comparison and finger counting, the visual symbolic basis of mathematics is likewise likely to be a cultural adaption.

The brain consists of two hemispheres connected by a bunch of white matter called the Corpus callosum. Ever since split-brain experiments on monkeys there has been a fascination with the idea that the two hemispheres in humans may have different or complementary functions, stemming partly from the knowledge coming originally from war injuries and strokes that injury to the 'dominant' left hemisphere which is usually contra-laterally connected to the use of the right hand, is selectively devoted to language typified in Broca's area of the frontal cortex which facilitates fluent speech and Wernike's area of the temporal cortex, which mediates meaningful semantic constructions. Although this result came predominantly from men and brain scans on both sexes have subsequently showed that language acquisition in women is more bilateral than in men, the idea that the two hemispheres had complementary functions captured the imagination of neuroscientists.

There is some evidence generally for this idea with music and creative language use having a partially complementary modularization to structured language. This in turn led to the idea that the more structured aspects of mathematics, such as algebra, and the more amorphous entangled aspects such as topology might be processed in different ways in complementary hemispheres. While this idea is appealing, there are few actual experiments that have tested the idea, and brain scan studies have tended to concentrate on elementary mathematical skills, which psychologists and neuroscientists can test on a wide variety of subjects researching basic brain skills, such as mental arithmetic and mental rotation, rather than complex abstract procedures.

Theories about how mathematical reasoning is processed gravitate on common sense ideas linking specific sensory modalities, known linguistic capabilities and general principles of frontal cognitive processing to generate parallel processing models of brain-related modalities having a natural affinity with mathematical reasoning.

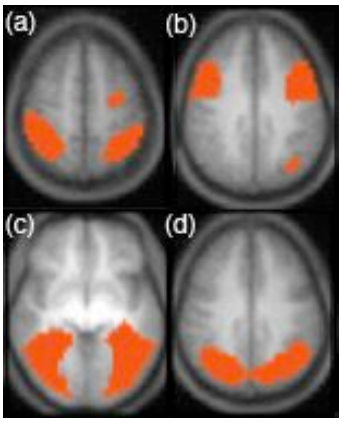


Figure 6.4 Triple-code model[53] of numerical process in has support from independent component analysis of fMRI scans of mental addition and subtraction revealing four components. (a) bilateral inferior parietal component may reflect abstract representations of numerical quantity (analog code) (b) left peri-sylvian network including Broca's and Wernike's areas and basal ganglia reflecting language functions. (c) ventral occipitotemporal regions belonging to the ventral visual pathway and (d) secondary visual areas consistent with a visual Arabic code.

For example the triple-code model[54] (figure 6.4) of numerical processing proposes that numbers are represented in three codes that serve different functions, have distinct functional neuro-architectures, and are related to performance on specific tasks. The analog magnitude code represents numerical quantities on a mental number line, includes semantic knowledge regarding proximity (e.g., 5 is close to 6) and relative size (e.g., 5 is smaller than 6), is used in magnitude comparison and approximation tasks, among others, and is predicted to engage the bilateral inferior parietal regions. The auditory verbal code (or word frame) manipulates sequences of number words, is used for retrieving well-learned, rote, arithmetic facts such as addition and multiplication tables, and is predicted to engage general-purpose language modules, associated with memory and sequence execution. The visual Arabic code (or number form) represents and spatially manipulates numbers in Arabic format, is used for multi-digit calculation and parity judgments, and is predicted to engage bilateral inferior ventral occipito-temporal regions belonging to the ventral visual pathway, with the left used for visual identification of words and digits, and the right used only for simple Arabic numbers.

In research focusing on the intra-parietal regions contrasting number comparison with other spatial tasks[55], number-specific activation was revealed in left IPS and right temporal regions, whereas when numbers were presented with other spatial stimuli the activation was bilateral[56].

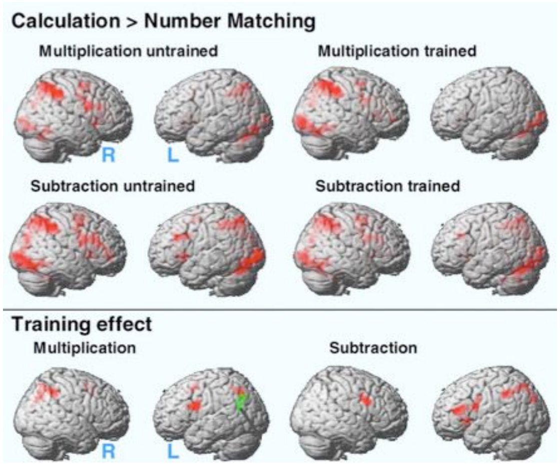


Figure 6.5 Unpracticed and learned tasks in multiplication and subtraction are contrasted[57].

Further support for the triple-code model comes from studies of learning complex arithmetic (multiplication)[58], where left hemispheric activations were dominant in the two contrasts between untrained and trained condition, suggesting that learning processes in arithmetic are predominantly supported by the left hemisphere. Activity in the left inferior frontal gyrus may accompany higher working memory demands in the untrained as compared to the trained condition. Contrasting trained versus untrained condition a significant focus of activation was found in the left angular gyrus. Following the triple-code model, the shift of activation within the parietal lobe from the intraparietal sulcus to the left angular gyrus suggests a modification from quantity-based processing to more automatic retrieval. A second study involving learning multiplication and subtraction supports similar conclusions (figure 6.5). This trend suggests that learned mathematical tasks of this kind become committed to linguistic memorization, once they are mastered.

In contrast with this, an experiment where subjects were asked to analyze a simple mathematical relationship[59], e.g. $x = A$, $B = A + 6$, $C = A + 8$ by either forming a number line picture, or constructing the left side of a solving equation e.g. $x + (x + 6) + (x + 8) = S$, in both cases visual processing areas were activated and there were no significant differences in processing in language areas. This suggests visual processing areas are involved in forming equations, at least unfamiliar newly presented ones.

An intriguing study, which has more implications for advanced mathematics, where real conjectures are examined and proved, or found false, examined brain areas activated when true and false equations were presented to the subject[60]. This study found greater activation to incorrect, compared to correct equations, in the left dorsolateral prefrontal left ventrolateral prefrontal cortex, overlapping with brain areas known to be involved in working memory and interference processing.

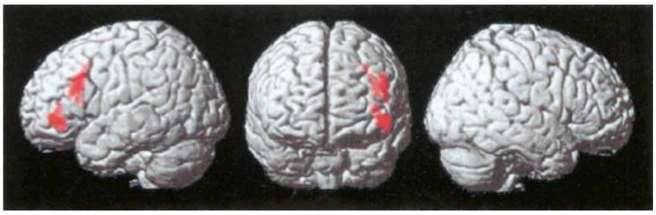


Figure 6.6 Prefrontal areas activated differentially when incorrect mathematical equations are presented.

Extending this into the geometrical area and specifically with gifted adolescents is a study of mental rotation (figure 6.8) involving images such as 3-D polyminoes. In contrast to many neuroimaging studies, which have demonstrated mental rotation to be mediated primarily by the right parietal lobes, when performing 3-dimensional mental rotations, mathematically gifted male adolescents engage a qualitatively different brain network than those of average math ability, one that involves bilateral activation of the parietal lobes and frontal cortex, along with heightened activation of the anterior cingulate.

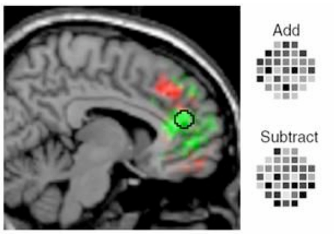


Figure 6.7 Differential activation of the medial prefrontal cortex can predict a person's intention to add or subtract two numbers[61].

It has also become possible to teach a computer to distinguish subjects' intention to add or subtract two numbers, using analysis of detailed differential activation of the medial prefrontal cortex, giving predictions which are 70% accurate (figure 6.7).

A brain imaging study of children learning algebra (simple linear equations)[62], shows that the same regions are active in children solving equations as are active in experienced adults solving equations, however practice has a more striking adaptive response in children. As with adults, practice in symbol manipulation produces a reduced activation in prefrontal cortex area. However, unlike adults, practice seems also to produce a decrease in a parietal area that is holding an image of the equation. This finding suggests that adolescents' brain responses are more plastic and change more with practice.

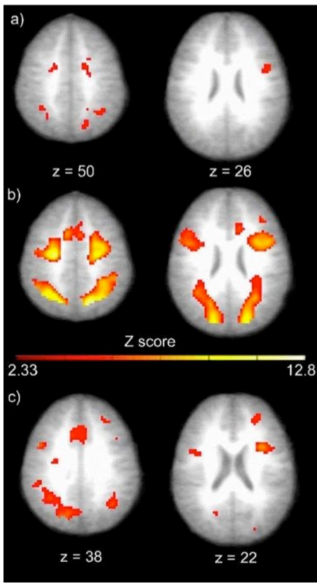


Figure 6.8 Mental rotation: Above average subjects, middle gifted subjects, below the difference in activation between the groups[63].

Other theorists have proposed differing models to the triple-code, in which there are modules for comprehension, calculation, and number production. The comprehension module translates word and Arabic numbers into abstract internal representations of numbers, calculations are performed on these representations, and then the abstract representations are converted to verbal or Arabic numbers using specific number production modules. Here amodal abstract internal representations of numbers are operated on, rather than numbers represented in specific codes (i.e., quantity, verbal, or Arabic).

The differences between these models are great. For example damage in the first would give rise to failures of one modality of processing or another, while in the second particular abstract operations would be impaired. Functional activation would be different in the two cases when stimuli involving mathematical processing are presented to the subject.

What do all these brain studies add up to and what bearing do they have on the sort of processes that go on in advanced mathematics? Although the subject trials rarely engage anything resembling the sort of advanced mathematics performed at the graduate level, they do suggest that a broad spectrum of brain areas are involved in mathematical reasoning, involving spatial transformations, visual representation of closeness and relative position on the number line, recognition of numbers and algebraic expressions, making strategic and semantic decisions and transforming many of these processes into coded linguistic transformations as they become familiar and memorized. They also suggest that much of the basis for the richness of mathematics as a palpable reality come from sensory and spatial processes in contrast to the emphasis placed on formal linguistic logic in advanced mathematics.

To ensure mathematics continues to be a real part of human culture and doesn't suffer the same fate as classical languages such as Latin in a world of pocket calculators and laptop computers which obviate the need for mathematical expertise in much of the population, mathematicians need to stay in touch with the perceivable richness of science and artistry and imaginative challenge many directly perceivable areas of mathematics do provide without consigning all such problems to the trash can of triviality in an era when new classical results at the research level can only be produced in esoteric spaces through formal processes that stretch far beyond the rich landscapes human imagination into the ivory towers of formalism.

7: The Fractal Topology of Cosmology

An acid test of abstract mathematics as a description of reality is how well it fits naturally with the emerging cosmological description of reality we are in the process of discovering. While physics had to face the demise of the classical paradigm forewarned in Kelvin's two small dark clouds of quantum theory and relativity, classical mathematics has not yet come to terms with these changes to its singular foundations.

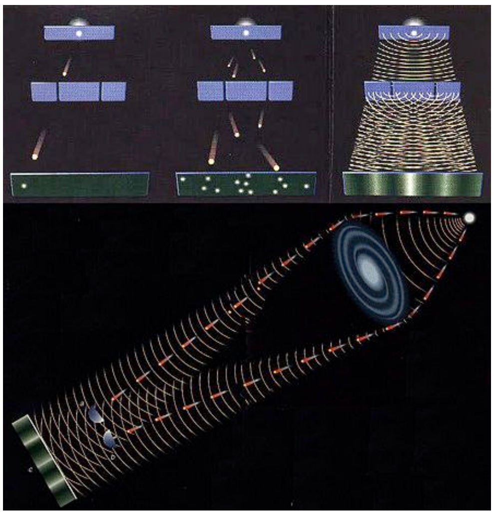


Figure 7.1 Top: Quantum interference invokes wave-particle complementarity. Bottom: Wheeler delayed choice experiment.

Quantum reality and cosmological relativity display troubling features which raise questions about the classical model of mathematics based on point-like singular elements in a space whose geometry is independent of its components. Rather than contrasting the discrete and continuous, quantum theory is indivisibly composed of complementary entities which posses both features through wave-particle complementarity, as illustrated in the interference experiment, figure 7.1, in which quanta released as localized 'particles' from individual atoms traverse a double slit as waves, only to be reabsorbed by individual atoms on a photographic plate in the interference fringes. Such complementarity arises from a feedback process between dynamical energy and wave geometry, as expressed in Einstein's law:

$$E = h\nu \qquad [7.1]$$

However the space-time properties of these quanta are counterintuitive, as can be seen from the Wheeler delayed-choice experiment, where changing the absorbing detector system, from interference detection to individual particle detection can change the apparent path taken by the quanta, long before they arrived.

Worse still, in contrast to quantum theory, which is usually couched within space-time, general relativity applies a second feedback between energy and geometry, in the form of curvature of space-time, so that the geometry and topology of space is also a function of the dynamics. This makes integrating quantum theory and relativity a conceptual nightmare, because, in the event virtual black holes can be created by quantum uncertainty, space-time is locally a seething foam of wormholes, resulting in contradictory descriptions.

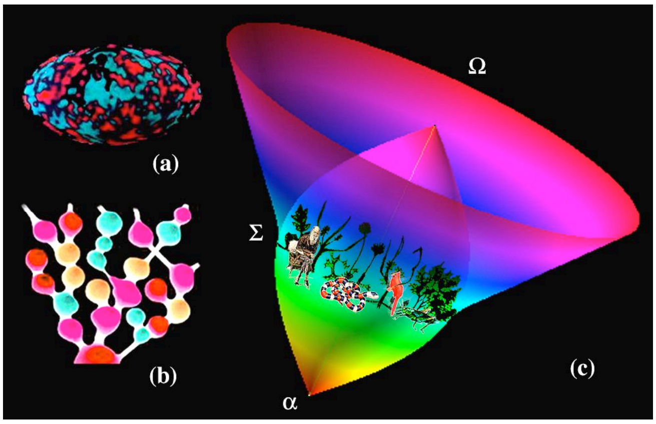


Figure 7:2 The red-shifted cosmic fireball (a) has fluctuations consistent with being inflated quanta. Fractal inflation (b) provides a topological model of how the large-scale structure of the universe might expand forever. Whether or not it does is also a topological question between a closed and open manifold structure.

An oracle for the fit of classical mathematics with reality is the elusive TOE, or theory of everything, which has remained just around the corner since Einstein made inroads into both quantum theory and relativity. In every respect, the search for a unified cosmological theory fundamentally brings topology into the picture and lays siege to classical notions such as point singularities.

Inflation, as a key candidate theory of cosmic emergence, links events at the quantum and cosmological levels. Symmetry-breaking between the forces of nature at the quantum level is coupled to a switching from a phase of cosmic inflation in which an 'anti-gravity' causing an exponential decline in the curvature of the universe switches to attractive gravity, the kinetic energy thus equaling the gravitational potential energy, enabling the universe to be born out of almost nothing.

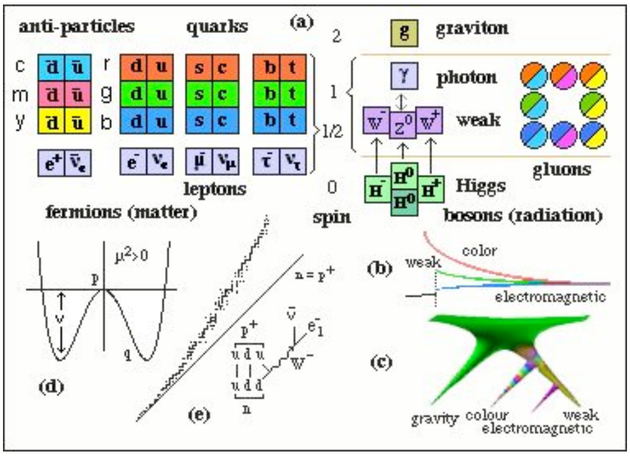
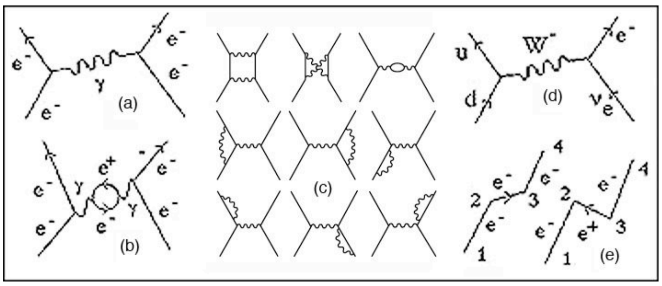


Figure 7.3 The standard model of physics involves a symmetry-breaking between electromagnetism and the weak and colour nuclear forces. A deeper symmetry-breaking is believed to unite gravity with the others.

At the quantum level, theories uniting gravity and the other forces are based on a variety of forms of symmetry-breaking, in which the differences between the two nuclear forces, electromagnetism and gravity arise from symmetry-breaking transformations of a super-force.

In the standard model of particle physics, the divergence, first of the weak force from electromagnetism, and then the color force of the quarks and strong nuclear force are mediated by forms of symmetry-breaking in which the bosons carrying the weak force take up a scalar Higgs' particle and thus gain non-zero rest mass, at the same time quenching the inflationary anti-gravity effect of the Higgs' field. The latent energy released by this process gives rise to the hot shower of particles in the big-bang's aftermath. A similar but slightly different symmetry-breaking applies to the colour force that binds quarks, involving massless bound gluons.

Figure 7.4 Feynman diagrams (a) 2nd order and (b) sample 4th order terms in the infinite series determining the scattering interaction of two electrons. (c) The full set of 4th order terms. (d) The weak *W* particles act as heavy charged photons indicating symmetry-breaking. (e) Time-reversed electron scattering is positron-electron creation annihilation, showing virtual particles are time reversible[64].



Quantum field theories are fractal theories, because they define the force, say the electromagnetic scattering between two electrons, in terms of a power series of terms, mediated by virtual photons, summing every possible virtual particle interaction permitted by uncertainty, each of which corresponds to an increasingly elaborate Feynman space-time interaction diagram (figure 7.4 a-c). The series is convergent in the case of electromagnetism because the terms diminish by a factor

$$\frac{e^2}{hc4\pi\varepsilon_0} \sim \frac{1}{137} \qquad [7.2]$$

the so-called fine-structure constant. A major quest of all theories is such convergence, to avoid infinite energies or probabilities.

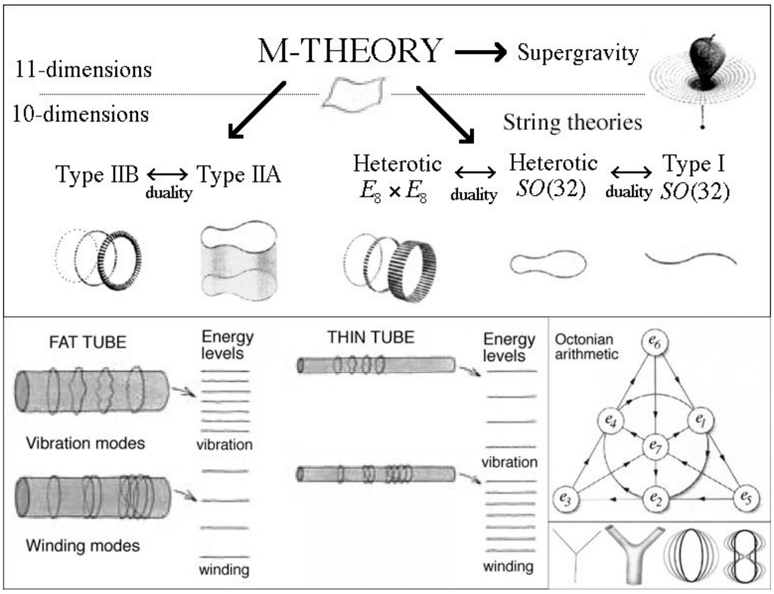


Figure 7.5 Top: M-theory can unite several 10-D string theories and 11-D supergravity through dualities[65]. The holographic principle allows an n-D theory to be represented on an (n-1)-D surface. Lower left: dualities between theories can exchange vibration and topological winding modes of strings on the compactified dimensions[66]. The algebra of the groups may invoke the octonians, lower right. String excitations, bottom right, avoid point singularities, but result in topological connections when strings meet.

Attempts to unite gravity with the other forces have proved more difficult, with a series of theories striving to hold the centre ground, from supergravity, through superstring[67] to higher dimensional (mem)brane M-theories[68]. All these theories have topological features attempting to get at the root of the singularities associated with the classical notion of point singularity and its infinite energy. They are broadly based on supersymmetry[69] – the idea that every force carrying boson of integer spin is matched by a matter-forming fermion of half-integer spin to ensure their independent contributions balance to give rise to a convergent theory. All string and brane theories are founded on removing the infinite energy of a point singularity by invoking the quantum vibrations of a topological loop or string, or membrane for small distance scales, resulting in a series of excited quantum states. Connecting several of these theories are principles of duality in which two theories with differing convergence properties can be seen to be dual, so that a non-convergent description in one corresponds to a convergent description in the other. This can result in dual descriptions of reality in which supposed fundamental particles, like quarks and neutrinos exchange roles with supposed composites of exotic particles like the magnetic monopole singularities of symmetry-breaking. These theories also share a basis in invoking a higher dimensional space, usually of 10 to 12 dimensions to make the theories convergent. This in turn raises the notorious compactification problem of how some of these dimensions can be topologically 'rolled up' into closed loops forming internal spaces representing the 10-12 internal symmetries of the twisted form of the forces of nature we experience as well as the four dimension of space-time. These theories involve topological orbifolds[70] – orbit generalizations of manifolds factored by a finite group of isometries, Calabi-Yau manifolds[71], topological bifurcations, and potentially up to 10^{500} candidate string theories[72] hypothetically representing multiverses with differing properties, only a vanishing few of which would support life and sentient observers, thus invoking the Anthropic principle[73], rather than cosmologically unique laws of symmetry and symmetry-breaking.

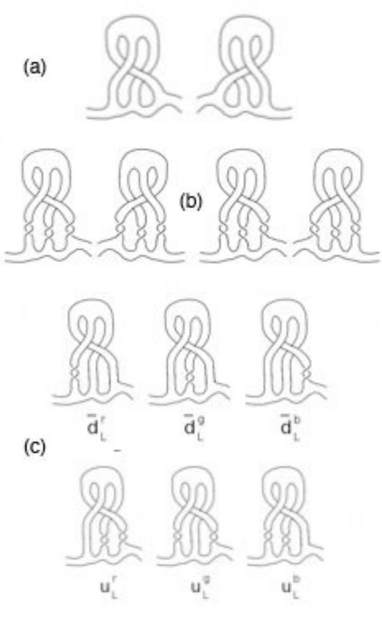


Figure 7.6 How the lightest family of particles in the standard model appear as braids[74, 75, 76]. Each complete twist corresponds to a third unit of electric charge depending on the direction of the twist. (a) Electron neutrino and anti-neutrino correspond to mirror-image braidings. (b) Four states corresponding to the electron and positron with charge depending on the orientations of the twists. (c) Three colours of up quark and anti-down quark.

An alternative to string and brane theory is loop quantum gravity[77] and topological quantum gravity based on braided preons (figure 7.6). Here again we have a topological basis, in which the fundamental particles are braids in space-time, consisting of more fundamental units called preons, three of which make up each quark and each lepton. The orientation of the twists in these braids determine a fractional electric charge of $\pm\frac{1}{3}$

which can sum in differing ways to the charge on the electron and positron or up and down quarks. The theory predicts many features of the standard model including the relationship between quarks and leptons, the charges of the two flavours of quark – up and down and the fact that each of these come in three colours corresponding to the combinations of one and two twist braids on the triplet and can model particle interactions through concatenation and splitting of braids.

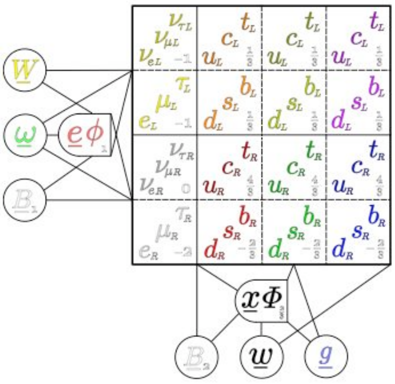


Figure 5: A periodic table of $E8$.

More recently Garrett Lisi's "Exceptionally simple theory of Everything" [78, 79] attempts to integrate all the forces including gravity and interactions of both fermions and bosons in terms of the root vector system generating E8, with its subalgebras such as G2 and F4 representing sub-interactions, such as the colour force. The connection he uses again represents the curvature and action on a four-dimensional topological manifold

We thus find that all the candidate theories of reality have an intrinsic topological as well as an algebraic basis and all lead to situations in which the classical view of mathematical spaces is replaced by quantized versions, which fundamentally alter the founding assumptions. One can then ask whether the difficulty at arriving at a theory of everything results from the obtuseness of physicists, or the inadequacy of abstract mathematics as a cultural language of ideals to come to terms with the actual nature of the universe we find ourselves within.

8: References

In the interests of the maze-like nature of mathematics, these references have an emphasis on interlinked internet resources, particularly those from Mathworld and Wikipedia, which themselves provide direct access to the magical maze in the noosphere, which mathematics has become.

1 http://en.wikipedia.org/wiki/Popperian_cosmology
2 Andersen Johannes 1927 Maori String Figures, Steele Roberts Ltd., Wellington ISBN 1-877228-38-9
3 Walker, Barbara 1971 A Second Treasury of Knitting Patterns, Pitman ISBN 0-273-36073-6 p139
4 http://en.wikipedia.org/wiki/Travelling_salesman_problem
5 Dienes Zoltan, Holt Michael 1972 ZOO, Longman Group ISBN 0 582 18449
6 Holt Michael 1973 Inner Ring Maths II, Ernest Benn ISBN 0 510 07872 9
7 http://mathworld.wolfram.com/CollatzProblem.html
8 http://en.wikipedia.org/wiki/Collatz_conjecture
9 http://www.ericr.nl/wondrous/pathrecs.html
10 Stewart Ian1997 The Magical Maze: Seeing the world through the mathematical eye, Weidenfield & Nicholson, London
11 De Chardin Teilhard 1955 The Phenomenon of Man William Collins Sons & Coy. Ltd., London.
12 http://en.wikipedia.org/wiki/Lipschitz_continuity
13 http://en.wikipedia.org/wiki/Weierstrass_function
14 K. Falconer 1984 The Geometry of Fractal Sets, Oxford.
15 Barnsley, Michael 1988 Fractals Everywhere, Academic Press, New York.
16 http://mathworld.wolfram.com/WeierstrassFunction.html
17 http://en.wikipedia.org/wiki/Lebesgue_measure
18 http://en.wikipedia.org/wiki/Vitali_set
19 http://en.wikipedia.org/wiki/Hyperreal_number
20 http://en.wikipedia.org/wiki/Long_line_(topology)
21 http://en.wikipedia.org/wiki/Poincaré_conjecture
22 http://mathworld.wolfram.com/Go.html
23 Schroeder M. 1993 Fractals, Chaos and Power Laws ISBN 0-7167-2136-8.
24 http://mathworld.wolfram.com/FiniteGroupT.html
25 http://mathworld.wolfram.com/FiniteGroup.html
26 http://en.wikipedia.org/wiki/Presentation_of_a_group
27 http://en.wikipedia.org/wiki/Center_of_a_group
28 http://en.wikipedia.org/wiki/Cayley_graph
29 http://en.wikipedia.org/wiki/Semidirect_product
30 http://en.wikipedia.org/wiki/Petals_Around_the_Rose
31 http://en.wikipedia.org/wiki/Squaring_the_square
32 http://mathworld.wolfram.com/PerfectSquareDissection
33 Bouwkamp, C. J. and Duijvestijn, A. J. W. *Catalogue of Simple Perfect Squared Squares of Orders 21 Through 25*, Eindhoven Univ. Technology, Dept. Math, Report 92-WSK-03, Nov. 1992.
34 http://mathworld.wolfram.com/MagicSquare.html
35 http://en.wikipedia.org/wiki/Piet_Hein_(Denmark)
36 http://en.wikipedia.org/wiki/Peg_solitaire

[37] http://www.durangobill.com/Peg33.html
[38] Myers Bernard 1979 Supermazes No1, Fredrick Muller Ltd., London ISBN 0-584-10296-8
[39] http://mathworld.wolfram.com/Chess.html
[40] New Scientist 25 December 2007, 66-67
[41] Benbow Camilla 1988 *Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects and possible causes*, Behavioral and Brain Sciences 11 169-232.
[42] Butterworth, Brian 1999 What Counts: How every brain is hardwired for math, Free Press NY ISBN 0-684-85417-1
[43] Dehaene, Stanislas 1997 Number Sense: How the mind creates mathematics, Oxford University Press NY.
[44] Kimura, Doreen 1992 *Sex Differences in the Brain*, Scientific American, Sept 81.
[45] Shaywitz, B. and S. et. al. 1995 *Sex differences in the functional organization of the brain for language*. Nature 373, 607-9.
[46] Blum Deborah 1997 Sex on the Brain, Penguin, N.Y.
[47] von Bredow, Rafaela 2006 *Living without Numbers or Time*, Nature news May 3
[48] http://en.wikipedia.org/wiki/Srinivasa_Ramanujan
[49] Tang, Yiyuan et. al. 2006 *Arithmetic processing in the brain shaped by cultures*, Proc. Nat. Acad. Sci. 103 10775-10780
[50] http://en.wikipedia.org/wiki/Generative_grammar
[51] http://en.wikipedia.org/wiki/The_Language_Instinct
[52] Christiansen Morten, Kirby Simon (ed.) 2003 Language Evolution, Oxford University Press
[53] Schmithorst VJ. Brown RD. 2004 *Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group Independent Component Analysis of the mental addition and subtraction of fractions*, Neuroimage *22(3) 1414-20*.
[54] Dehaene S et. al. 1999 *Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence* Science 284 970-974.
[55] Cohen Kadosh R et. al. 2005 *Are numbers special? The comparison systems of the human brain investigated by fMRI* Neuropsychologia 43 1238–1248.
[56] Eger E et. al. 2003 *A Supramodal Number Representation in Human Intraparietal Cortex* Neuron, 37, 719–725.
[57] Ischebeck A 2006 *How specifically do we learn? Imaging the learning of multiplication and subtraction* NeuroImage 30 1365–1375.
[58] Delazer M et. al. 2003 *Learning complex arithmetic—an fMRI study* Cognitive Brain Research 18 76-88.
[59] Terao A et. al. *An fMRI study of the Interplay of Symbolic and Visuo-spatial Systems in Mathematical Reasoning* http://act-r.psy.cmu.edu/papers/679/paper507.pdf
[60] Menon V et. al. 2002 *Prefrontal Cortex Involvement in Processing Incorrect Arithmetic Equations: Evidence From Event-Related fMRI* Human Brain Mapping 16:119–130.
[61] Haynes, John Dylan 2007 *Current Biology* DOI: 10.1016/j.cub.2006.11.072.
[62] Qin Y et. al. 2004 *The change of the brain activation patterns as children learn algebra equation solving* Proc. Nat. Acad. Sci. 101 5686 –5691.
[63] Boyle M. et. al. 2005 *Mathematically gifted male adolescents activate a unique brain network during mental rotation* Cognitive Brain Research 25 583–587
[64] King Chris 2006 *Quantum Cosmology and the Hard Problem of the Conscious Brain* in "The Emerging Physics of Consciousness" Springer (Ed.) Jack Tuszynski 407-456. http://www.math.auckland.ac.nz/~king/Preprints/pdf/tuz6.pdf
[65] Hawking Stephen 2001 Universe in a Nutshell Bantam Books, NY
[66] Duff, Michael 2003 *The theory formerly known as strings* (2nd ed.) in The Edge of Physics, Scientific American
[67] http://en.wikipedia.org/wiki/Superstring_theory
[68] http://en.wikipedia.org/wiki/M-theory
[69] http://en.wikipedia.org/wiki/Supersymmetry
[70] http://en.wikipedia.org/wiki/Orbifold
[71] http://en.wikipedia.org/wiki/Calabi-Yau
[72] Brumfiel Geoff 2006 *Outrageous Fortune* Nature 439 5 Jan p10.
[73] http://en.wikipedia.org/wiki/Anthropic_principle
[74] Castelvecchi D 2006 *You are made of space-time* New Scientist 12 Aug.
[75] Bilson-Thompson Sundance 2006 A topological model of composite preons http://www.arxiv.org/pdf/hep-ph/0503213
[76] Bilson-Thompson S, Markopoulou F, Smolin L 2006 Quantum Gravity and the Standard Model http://www.arxiv.org/pdf/hep-th/0603022
[77] http://en.wikipedia.org/wiki/Loop_quantum_gravity
[78] http://arxiv.org/pdf/0711.0770
[79] http://en.wikipedia.org/wiki/An_Exceptionally_Simple_Theory_of_Everything