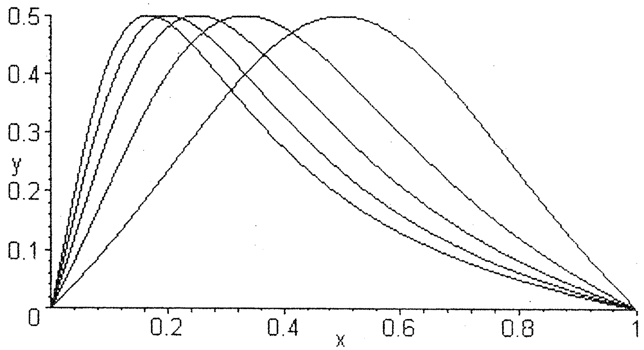
Function

A School Mathematics Journal

Volume 28 Part 3 June 2004



School of Mathematical Sciences – Monash University

Reg. by Aust. Post Publ. No. PP338685/0015

Function is a refereed mathematics journal produced by the School of Mathematical Sciences at Monash University. It was founded in 1977 by Prof G B Preston, and is addressed principally to students in the upper years of secondary schools, but also more generally to anyone who is interested in mathematics.

Function deals with mathematics in all its aspects: pure mathematics, statistics, mathematics in computing, applications of mathematics to the natural and social sciences, history of mathematics, mathematical games, careers in mathematics, and mathematics in society. The items that appear in each issue of *Function* include articles on a broad range of mathematical topics, news items on recent mathematical advances, book reviews, problems, letters, anecdotes and cartoons.

* * * * *

Articles, correspondence, problems (with or without solutions) and other material for publication are invited. Address them to:

The Editors, *Function*
School of Mathematical Sciences
PO BOX 28M
Monash University VIC 3800, AUSTRALIA
Fax: +61 3 9905 4403
e-mail: michael.deakin@sci.monash.edu.au

Function is published five times a year, appearing in February, April, June, August, and October. Price for five issues (including postage and GST): $33* ; single issues $7. Payments should be sent to: The Business Manager, Function, School of Mathematical Sciences, PO Box 28M, Monash University VIC 3800, AUSTRALIA; cheques and money orders should be made payable to Monash University.

---

* $17 for *bona fide* secondary or tertiary students.

THE FRONT COVER

This issue's front cover shows a famous set of graphs. Five of the set are drawn, although it is actually infinite. The relevant formula is

$$y = \frac{nx(1-x)}{n^2x^2 + (1-x)^2}$$

and the graphs drawn are those for $n = 1, 2, 3, 4, 5$.

It is readily apparent from the cover illustration and also from the formula given above that all the graphs pass through the points (0, 0) and (1, 0). Each graph also has a unique maximum point and we can find this by differentiating. The details of this calculation are left to the reader, but the result is that the maximum occurs at the point $\left(\frac{1}{n+1}, \frac{1}{2}\right)$.

If we take *any particular value* of x, and ask what happens as n gets larger and larger, then for very large n,

$$y \approx \frac{1-x}{nx}$$

and so *this* value of y tends to zero as n tends to infinity. Thus if we designate the *particular* value of x by X (say), we have in symbols

$$\lim_{n\to\infty} \frac{nX(1-X)}{n^2X^2 + (1-X)^2} = 0.$$

However, the interesting (and paradoxical) point is that the curves themselves, taken as a whole, never approach the curve $y = 0$, as there is always a point somewhere on each curve for which $y = ½$.

[Something like this formed the basis for the April Fools' Day letter of our 2002 issue, although *that* case is rather more complicated.]

This example is due to Georg Cantor, who is better remembered for his research into the question of transfinite numbers. [See *Function*, *Volume* **2**, *Parts* **1** and **2**.] Before Cantor developed this interest (for which he is best remembered today) he looked into the theory of trigonometric series (like those in our issue for April 2002) and set up much of the analytic machinery for dealing with such series. It was in this context that he advanced the example illustrated on this issue's front cover.

Cantor was a pupil of Weierstrass, who is responsible for much of the careful work needed to put the theory of functions onto a sound footing. [Some of this story was told in our special issue, released at the end of 1996 to commemorate 100 regular numbers of *Function*.]

A biography of Cantor may be found at the website

http://www-history.mcs.stand.ac.uk/history.Mathematicians/Cantor.html

He was a major figure in Mathematics toward the end of the 19th century and into the early 20th. Sadly his latter years were clouded by mental illness, but this in no way diminishes the significance of his contributions to Mathematics. Our cover illustrates one of these: one of the lesser-known of them, but an important and interesting one nonetheless.

∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞

NEWS ITEMS

Another Prime Number Result Proved?

In *Function* last June and again last October, we reported the results of recent investigations into the "twin prime conjecture". This is the statement (not yet proved) that there are infinitely many pairs of primes like 3 and 5, 4 and 7, 11 and 13, and so on whose difference is 2. The conjecture also turned up in the History Column for April 2001.

In an attempt to investigate the problem, Dan Goldston and Cem Yildrin developed a new approach that seemed to show a lot of promise. At first they believed that by using it they had been able to show a strong partial result. However in this they were mistaken, for their proof of the technical result was found to contain an error. However, as was remarked in our October story, the new approach was nonetheless perceived as very promising.

This judgement would seem to be borne out by another related development. Prime pairs are examples of 2 primes in arithmetic progression. Now consider 3, 7, 11. This is a sequence of length 3 with a common difference of 4. A rather meatier example is the 10-term arithmetic progression of primes with a common difference of 210: 199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089. This is the sort of area of Mathematics where people strive for records. Currently the record is shared by two sequences both 22 terms long. The first, found in 1993, has terms of 11,410,337,850,553 + 4,609,098,694,200k. This was equaled in 2003 by 376,859,931,192,959 + 18,549,279,769,020k, where in each case $k = 0, 1, \ldots, 21$.

Back in 1923, G H Hardy (whose story was told in the June 1995 issue of *Function*) and his long-time collaborator, Littlewood, made a very general conjecture known as the "k-tuple conjecture" which implies that there exist prime arithmetic progressions of any length k, although the full statement goes beyond this.

However it now seems to be established that the sequence of prime numbers *does* contain arithmetic progressions of length k for all k. This result has been announced by Ben Green and Terence Tao (the expatriate Australian who was the subject of a news story in our issue for last February). The work has not yet been formally published, but in accordance with much modern practice it has been posted on the Internet for scrutiny by other mathematicians, and so far seems to have met with approval. Their work makes use of several sophisticated techniques, but among them is the approach of Goldston and Yildrin.

If this work does meet with full approval, it will be published in the refereed literature and will become a theorem. Meanwhile it is posted on the Internet at:

http://arxiv.org/abs/math.NT/0404188.

But beware! Only an expert in advanced Number Theory will be able to follow it!

∞∞∞∞∞∞∞∞∞∞∞∞∞∞∞

The Second Abel Prize

Our cover story last October dealt with the award of the first ever *Abel Prize* to Jean-Pierre Serre. The Abel prize honours the memory of the Norwegian mathematician Niels Henrik Abel, and is now awarded annually by the Norwegian Academy of Arts and Letters.

The 2004 prize has now been announced and this time it is a joint award. The recipients are Sir Michael Atiyah of the University of Edinburgh, and Isadore M Singer of the Massachusetts Institute of Technology. Atiyah and Singer collaborated back in the 1960s to prove the "Atiyah-Singer index theorem", that has been described by the American Mathematical Society's website as "bringing together topology, geometry and analysis", and as "building new bridges between mathematics and theoretical physics".

For a readable account of the impact of their work on Mathematics and Physics, see the website

http://www.abelprisen.no/nedlastning/2004/english_2004.pdf

Both winners have been the recipients of numerous other awards: in particular, Atiyah has already received a Fields medal (prior to the institution of the Abel prize, seen as the mathematical equivalent of a Nobel Prize). For detailed biographies of the two prizewinners, look up:

http://www-history.mcs.st-and.ac.uk/history/Mathematicians/Atiyah.html

for Atiyah, and

http://en.wikipedia.org/wiki/Isadore_Singer

for Singer.

"There is no branch of mathematics, however abstract, which may not some day be applied to phenomena of the real world."

Nikolai Lobachevsky

BOOK REVIEW

Shrine to University: A Geometry Journey along St Kilda Rd and Swanston Street (Second Edition), by Jill Vincent.
(Mathematical Association of Victoria, 2004)
Price $35; Members $25

The first edition of this delightful and informative book appeared in 1999, in advance of World Mathematical Year 2000. The year 2000 saw the Mathematics 2000 Festival hosted by the University of Melbourne. The first edition was a contribution to this endeavour. Now the MAV have published a second edition, that expands and updates the first. In that first edition, there were 14 chapters. This one has 15, and takes account of changes in the Melbourne streetscape in the last five years.

Sadly one of the original chapters is no more. The "Ansett A" has gone from the corner of Swanston and Franklin Streets. However, we have acquired in that time a new landmark: Federation Square. This has a chapter and there is also another new chapter, devoted to the Town Hall.

The author has been wonderfully diligent in tracking down original architectural drawings and suchlike material and has clearly devoted a lot of time and energy (not to mention legwork) to her project.

In this new edition, there are 15 landmarks discussed; all are on public view and all reflect significant mathematical concepts and structures. In order, they are:

1. The Shrine of Remembrance
2. The Victorian College of the Arts Logo
3. The Nautilus Fountain
4. The Floral Clock
5. The Victorian Arts Centre Spire
6. The Victorian Arts Centre Logo
7. The Footbridge over the Yarra
8. Federation Square
9. St Paul's Cathedral
10. The Melbourne Town Hall

11. The "Architectural Fragment" Sculpture
12. The Melbourne Central Cone
13. The Geodesic Dome at Melbourne Central
14. Storey Hall at RMIT
15. The Tessellating Pentagon Pavement

The book takes the form of a mathematical stroll through the heart of Melbourne, beginning at the Shrine of Remembrance and heading north.

The Shrine was erected as a memorial especially to the fallen of World War I, and is based on classical models. The lower part echoes the Parthenon in Greece and descriptions of the Mausoleum at Harlicanassus, one of the Seven Wonders of the Ancient World; above is placed a truncated pyramid, with classical proportions. Right at the top is an aperture that admits sunlight in such a way that the light falls exactly onto the "Stone of Remembrance" at 11 am on November 11 each year. This is the time and date of the Armistice that brought the hostilities of World War I to an end. To do this required careful surveying and a sound knowledge of astronomical principles. A further complication arose with the introduction of Daylight Saving. This is now compensated for by means of an ingenious arrangement of mirrors.

On the front of the Victorian College of the Arts is a logo based on the pentagon (symbolic of our five senses) and superimposed circular motifs. These overlap to surround a central five-pointed star. The overall pattern resembles a Celtic braided figure and sits astride the initials VCA. The book notes similar designs in Mediaeval and Islamic art. In particular, one of the windows of Exeter Cathedral (UK) incorporates a similar combination of pentagonal and circular features.

The structure of the Nautilus shell was briefly discussed in *Function's* October 1992 issue, where it was related to the "Golden Section". The third chapter of this book discusses a sculpture that follows the lines of a Nautilus shell and is to be found outside the National Gallery of Victoria, at the North end of the moat.

The Floral Clock on the other side of the road superposes the twelve hours of the day on a doubly octagonal design. Although the actual plants are changed from time to time, the underlying geometry remains the same throughout.

The spire on the Victorian Arts Centre is another of the mathematical sights. It also appeared in an earlier issue of *Function*, as our cover story in April 1980 (before it was actually built!). Here there is an extremely thorough study of the spire in all its aspects, with detailed architectural drawings to go with it. An accompanying table quotes some of the dimensions to the nearest thousandth of a millimetre! The lower part takes the form of a hyperbolic paraboloid: $z = y^2 - x^2$.

The Arts Centre also sports a logo that represents in two dimensions, what the spire models in three. It is realized by the superposition of part of the hyperbola $y = 1/|x|$ onto an equilateral triangle.

Proceeding further north, we come to the Yarra River. With the inauguration of the Southbank development, a new footbridge was built to span the river. It is of a bow truss construction and the walkway is suspended from a parabolic arch.

Federation Square on the North bank of the Yarra, and the East side of Swanston Street is newly built since the first edition of this book was published. Its assertively "modern" appearance incorporates several features of mathematical interest. One of these is the "pinwheel tiling" that supplies the underlying geometry of many of the surfaces. This is based on the right-angled triangle whose shorter sides are 1 and 2. There are several examples of these proportions illustrated in the book.

Part of the controversy surrounding the plans for Federation Square arose from the supposed clash between the style of its architecture and the more traditional lines of the adjacent St Paul's Cathedral. This too has its mathematical interest, with its variety of Gothic arches, trefoil and quatrefoil windows and suchlike.

A brief chapter on the Melbourne Town Hall has also been added for this edition. The semi-circular windows of the façade incorporate the feature known as the "arbelos". This shape has also previously made a brief appearance in *Function*: in our History Column for June 1998.

Further north still on the corner of La Trobe St, we see, outside the Public Library, what looks like the corner of a building sinking slowly into the pavement. It is a sculpture whose official title is "Architectural Fragment". It is essentially a triangular pyramid with embellishments. Its largest face has the dimensions of a 3:4:5 triangle.

Opposite the "fragment" is Melbourne Central, with its prominent conical spire designed to cover and shelter the old shot tower. Again we see the juxtaposition of the old and the new in the architecture. The clearly "modern" dome encloses the heritage-listed shot tower.

Chapter 13 is an interesting and ambitious one, dedicated to the geodesic dome that is another feature of Melbourne Central. The construction of such domes is discussed in considerable detail: the way they are constructed from triangles, but on an underlying pattern of pentagons and hexagons. They relate to the classical solids of Greek antiquity, notably the icosahedron and the dodecahedron.

RMIT's Storey Hall is another blend of the old and the new. Prominent on its façade is a pattern of rhombuses. There are "fat" ones and "thin" ones that together tile the plane in a manner first discovered by the British mathematician Sir Roger Penrose, and today known as "Penrose Tiling". This also is the material for a long and informative chapter.

Related material forms the basis of the final chapter, on the pentagonal tiling of a pavement at the University of Melbourne. Between the building that now houses the department of Mathematics and Statistics and the building to its immediate West ("Old Geology") is an area paved with pentagonal sandstone tiles. The construction follows a pattern known as the "Margaret Rice tiling". This uses eight irregular pentagons that fit together to form a tessellating unit that allows a regular tiling by a process of translation, and thus makes for a periodic pattern (unlike the Penrose tiling which is not periodic).

The author of this delightful work is a Melbourne Mathematics teacher. She has supplemented her accounts of these landmarks with a set of activities, graded according to Year Level, and then the answers given to the problems set in conjunction with these activities.

The work ends with a double bibliography: books on page 87, and websites on page 88.

Did you know that Melbourne had so much Mathematics on public display? Read this book, visit the sites and understand the underlying Mathematics to enhance your enjoyment of our city' treasures!

FROM THE ARCHIVES

The Game of Slither

[The following discussion comes from *Function*, *Volume* 1, *Number* 2, April 1977. In those early days of *Function*, there was much discussion of various games. This extract formed part of that discussion. It has been slightly edited for this issue.]

In the June 1972 issue of *Scientific American*, Martin Gardner described the following game called SLITHER. This is a game for two players, played on a 5×6 lattice or 'grid', as shown in Figure 1 below. The rules are simple. Opponents take turns marking a horizontal or vertical segment of unit length. (For example, the move shown in Figure 2 is permitted, but that shown in Figure 3 is not.)

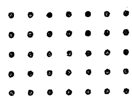


Figure 1

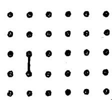


Figure 2

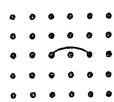


Figure 3

The segments must form a continuous path, but at each move the player must add to either end of the preceding path. The player forced to close the path is the *loser*. (Figure 4 shows a situation in which the next play must be a losing one.)

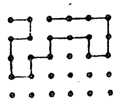


Figure 4

At the time Gardner wrote his article no winning strategy for either player was known[1]. In a tabulation of several hundred games the wins were about equally divided between the first and the second player so there was no indication whether one player or the other held the advantage. Soon after publication of his article Gardner received a flood of correspondence containing strategies of steadily mounting generality until finally Ronald C Read, a graph theorist at the University of Waterloo in Ontario, Canada, reduced the standard game to a monumental triviality.

Try to find a winning strategy for one of the players in the game of SLITHER just described. You might also try to formulate a general strategy for the game played on a grid of $m \times n$ dots for any natural numbers m and n greater than 1.

[The article went on to consider further possible generalisations; for example to grids where the dots did not necessarily form a rectangular pattern. In a later issue (October 1977), a full analysis of the $m \times n$ case was given. The conclusion was that if mn is even, then the first player has a forced win, whereas if mn is odd then the second player has a winning strategy. Details are given on pp 85-87 as to how to accomplish these wins. The discussion of grids other than the rectangular was left in an inconclusive state. An addendum published a contribution including a partial analysis from Graham Farr, then a student at Melbourne High School, today an Associate Professor in Monash University's School of Computer Science and Software Engineering.

An account of Ronald Read's definitive discussion of the original game is also included (on p 85).]

---

[1] Any game where chance plays no part, where two players move alternately, where the state of the game is on view to both players at all times, and where no draws are possible is the subject of the *winning strategy theorem*: either the first or the second player to move can force a win by the right choice of moves, *whatever the opponent does*. In this case, it is clear that these conditions are all fulfilled. The only one that might give us pause for thought is the question of whether draws are possible. But as long as there is a "free end" to the chain of interconnected dots, then a move is possible. Eventually we run out of dots and so such a "free end" must become joined to some point of the chain. At this juncture, the game is necessarily lost, not drawn! Of course, knowing that one player or the other has a winning strategy does not tell us either which player actually has it, nor how it is to be achieved. *Eds*

HISTORY OF MATHEMATICS

The Mathematics of Darwin and Mendel's Insights

Michael A B Deakin, Monash University

In the late 1850s, two separate lines of enquiry were proceeding simultaneously, but almost independently. In England, Charles Darwin was slowly working his way to his theory of natural selection. He came to the view that what fanciers of dogs, pigeons and other domestic animals achieved by selective breeding was also taking place in the wild as the result of competition between individuals for space in a crowded world. This led him to his version of the theory of evolution, for which he is now celebrated.

Meanwhile, in what is now the Czech Republic, Gregor Mendel was conducting experiments on garden peas, and sorting out the laws governing the inheritance of characteristics such as height, flower colour and the like. If, for example, a strain of pea that bred true for red flowers, was crossed with another that bred true for white flowers, then the offspring were all pink (which, perhaps, is what we might expect); however, if these pink flowers were then crossed with *one another*, the offspring could be red, white or pink (which is perhaps not what we might expect). When he looked at the proportions of these different colours, Mendel found that about ¼ were red, another ¼ were white and the remaining ½ pink.

Our current understanding of this result represents the red type as having two copies of a "gene" R, so that the red-flowering pea is represented as RR. The white type has two copies of another gene r, so that the white-flowering pea is represented as rr. The cross produces the combination of an R–type gene from one parent with an r–type from the other, resulting in the production of Rr progeny, which were identifiable as having the pink flowers. When these, in their turn, interbred, each passed on the R–type gene to half the progeny and the r type to the other half, and hence the result Mendel observed.

However, when other characteristics were investigated, things were rather different. If, for example, *seed colour* was studied, then the seeds were always either green or yellow, with yellow seeds making up about ¾ of the plants arising from the second cross. Here Mendel's results are now seen as resulting from two versions of a gene: *Y* (for "yellow") or *y*. After two generations of cross-breeding, there were still the three types *YY*, *yy* and *Yy* in the proportions ¼ : ¼ : ½ , but the *YY* and *Yy* types could not be told apart by simple inspection.

In this case, the *Y* was said to be *dominant* over the *y*, which was termed *recessive*. The only way to distinguish the *YY* plants from the *Yy* was to conduct breeding experiments. Nowadays we say that these two types have different *genotypes* but share the same *phenotype*.

There was no real attempt to draw these two strands (the Darwinian and the Mendelian) of research together for many years. Indeed, while Darwin's work became famous (even notorious), Mendel's was forgotten. It was not till 1900, when three geneticists (Correns, Tschermak and de Vries) independently rediscovered it and drew attention to it, that its significance was appreciated.

After that, however, things started to move quite rapidly. The first question to arise was: Why do not dominant genes drive out recessive ones? This question was answered (again independently) in 1908 by two researchers, the English mathematician G H Hardy (whose story was told in this column in June 1995) and Wilhelm Weinberg, a German physician.

The answer went like this. Suppose that a gene exists in two possible forms, *A* and *a*. Then there are three possible genotypes: *AA*, *Aa* and *aa*. But look not at these but rather at the proportions of the two forms of the gene: *A* and *a*. Suppose the *A* form exists in a proportion p of the total; then the other, *a*, will be present in a proportion q, where of course $p + q = 1$. Suppose now that all the genes are mixed up and combined randomly. Then the combinations will be *AA* with a frequency p^2, *Aa* with a frequency $2pq$ and *aa* with a frequency q^2. The values of p and q are unaltered and so these proportions continue thereafter. Notice that the question of dominance is entirely beside the point.

This result is now called the *Hardy-Weinberg Law*. It may be seen as the first of a series of results in the mathematical analysis of Mendel's laws of inheritance. However, it makes no allowance for Darwin's insight on natural selection. This extra complication took some time to be studied. In the early 1920s, R A Fisher and J B S Haldane sought to unify the two lines of enquiry. [Fisher made brief appearances in my columns for June and August 2000, and Haldane in February of that same year.]

The way in which this unity was achieved was to assign to each of the genotypes a "fitness". We now often use this term to refer to sporting prowess and the like, but this is not what is meant here. The "fitness", or *fitness coefficient*, of an individual is its aptness for its environment, as measured by the relative success of its offspring. The "fittest" individuals are those that contribute most to the next generation.

So suppose that the fitness of *AA* is u, that of *Aa* is v (often set equal to 1) and that of *aa* is w. Then the Hardy-Weinberg proportions need to be modified. There will be ratios p^2u, $2pqv$, q^2w of the three genotypes *AA*, *Aa*, *aa* (respectively) in the next generation. In order to make these three ratios add up to 1, we divide each by their total: $p^2u + 2pqv + q^2w$, which, because it is a mean of the three values u, v and w, is called the *mean fitness*. It is usually represented by the symbol $\overline{w}$.

It is now possible to write equations that show how the value of p (and hence of q) varies from one generation to the next under selection. One very common set of fitness values is the case $u = v = 1$, $w = 0$. This is the situation of a recessive lethal genetic disorder. In such a case, the form *aa* is slowly eliminated from the population and when it is finally eliminated, we have $\overline{w} = 1$. Prior to this, $\overline{w} = p^2 + 2pq = 1 - q^2 < 1$. So the effect of natural selection is to increase the mean fitness toward its maximum value.

This same principle applies to other situations as well. One well-documented case is that of a disease called *sickle-cell anaemia*. Here the gene responsible comes in two forms *S* and *s*. *SS* individuals are normal, and so to outside appearances are those with the *Ss* genotype. However *ss* individuals are condemned to die of a blood disorder called sickle-cell anaemia. It would thus appear that the case is the same as that just discussed. And so it is in some situations. Sickle-cell anaemia is a fatal disorder among black Americans.

However, in their ancestral homeland, back in Africa, things are more complicated. The phenotypes for *SS* and *Ss* are in fact not quite the same. Under the microscope, the red blood cells of the *Ss* individuals show a slight abnormality that in fact turns out to be a blessing; such individuals have an increased resistance to malaria, which is endemic in sub-saharan Africa. We thus have

$$u = 1 - \alpha \text{ (say)},\ v = 1,\ w = 0 \text{ so that } \overline{w} = p^2(1-\alpha) + 2pq .$$

In this case also, once selection is complete, $\overline{w}$ is again maximised. The final value is $\left(\dfrac{1-\alpha}{1+\alpha}\right)^2$ and it corresponds to the value $q = \dfrac{\alpha}{1+\alpha}$ which typically is small but not zero. Nature allows a number of deaths to occur (from sickle-cell anaemia) in order to prevent a larger number of people dying from malaria.

Now in fact very few traits are determined by the simple action of a single gene with just two possible forms. The best example of the next complication is that of the ABO blood groups among humans. Here the relevant gene comes in three forms that I will call *A*, *B* and *O* (although this is not the standard symbolism). There are six possible genotypes: *AA*, *AB*, *AO*, *BB*, *BO* and *OO*. These correspond to four distinct phenotypes: Type A (*AA* and *AO*), Type AB (*AB*), Type B (*BB* and *BO*) and Type O (*OO*).

[This was a most important piece of medical research as its elucidation allowed the practice of safe blood transfusion without the complication of adverse reactions.]

But this led to the proposal of a further mathematical problem. Suppose a gene were to exist in n different forms. There would them be $\dfrac{n(n-1)}{2}$ different genotypes and to each of these a fitness coefficient would be assigned and the mean of all these coefficients would be the value of $\overline{w}$. The question then arises: Is it still the case that $\overline{w}$ is maximised when the final equilibrium is achieved and selection is complete?

This question was investigated in the late 1950s by three separate teams of researchers. The answer is 'yes'. P A G Scheur and S P H Mandel proved this and published their proof in the journal *Heredity* in 1959; H P Mulholland and C A B Smith published their (different) proof in *American Mathematical Monthly* that same year. In 1960, a third proof (different again) appeared in the *Quarterly Journal of Mathematics*. The authors were three Australians: F V Atkinson, G A Watterson (a former editor of *Function*) and P A P Moran, a professor at the ANU.

A few years later, a considerably simpler proof was given. It also appeared in the *Quarterly Journal of Mathematics*, and its author was then a student, J F C Kingman (but now Sir John Kingman, FRS and a highly distinguished professor of Statistics at Cambridge).

All this endeavour tended to reinforce the belief that the effect of selection was to bring about "the survival of the fittest" in the quantitative sense that the mean (average) fitness always increased until a maximum was achieved.

However, this was a belief soon to be challenged. Before we look into this however, notice one point. In every case, we are looking at the frequency of the type of gene, not (except after the event) at the frequency of either the genotype of the individual nor of its phenotype. This extremely fruitful viewpoint has been standard since Hardy and Weinberg, and has since been popularised by books like *The Selfish Gene*. These look at the underlying mechanism of evolution and speak in terms that make what was initially a mathematical convenience into a governing principle. (This view is not, however, immune to challenge.)

By about 1960, it had long been known that the genes occur on intracellular structures called *chromosomes*, and a distinction was made between where on the chromosome a gene was to be found and what form it took at that point. The word "gene" tended to drop out in favour of two different terms: *locus*, which referred to the place in the set of chromosomes ("genome") where the gene was to be found, and *allele*, which referred to the form the gene took.

Think of the case of two loci, at each of which one or other of two alleles may be found. (This is really the very simplest case taking into account the complications of real Biology.) Then the various genetic combinations that can occur are AB, Ab, aB and ab. That is to say at the first locus, either the allele A or the allele a may be found while at the second the possibilities are B or b. Thus we have these four possible combinations.

Initially it had been thought that this situation could be analysed in terms of the frequencies of A and a on the one hand and separately of B and b on the other. So a body of literature arose in which the overall frequency of A (say) was p_1, with that of a being q_1 $(=1-p_1)$ and the overall frequency of B (say) being p_2, with that of b being q_2 $(=1-p_2)$. It was thought that the situation could be pictured by taking a square in which on one axis, values of p_1 were plotted between 0 and 1, while on the other values of p_2 were plotted, again between 0 and 1. It was assumed that to each point in this square a value of $\overline{w}$ could be assigned and that the maximum of all these values would represent the situation at equilibrium, once selection was complete. Such a diagram was called an "adaptive topography".

Two things happened to disturb this comfortable picture. The first was the derivation of exact equations governing the case under discussion. Several authors worked on this, but probably the most influential analysis was the joint work of the US geneticist R C Lewontin and a Japanese colleague K-I Kojima. They published their analysis in the journal *Evolution* in 1960.

They considered the four allele-combinations AB, Ab, aB, ab and looked on these as if they were just four versions of a gene as in the earlier analyses by Kingman and those who preceded him. Now if this were all there was to the matter then the earlier analysis would apply to this case also. But there was a further complication.

They took the four types just listed and assigned frequencies to them, in order: x_1, x_2, x_3, x_4, where $x_1+x_2+x_3+x_4=1$. They also assigned fitness coefficients according to the following table:

AB & AB: w_{11}	AB & Ab: w_{12}	AB & aB: w_{13}	AB & ab: w_{14}
Ab & AB: w_{21}	Ab & Ab: w_{22}	Ab & aB: w_{23}	Ab & ab: w_{24}
aB & AB: w_{31}	aB & Ab: w_{32}	aB & aB: w_{33}	aB & ab: w_{34}
ab & AB: w_{41}	ab & Ab: w_{42}	ab & aB: w_{43}	ab & ab: w_{44}

They were able to simplify this table somewhat. Biological considerations led to the conclusions that (1) for all i and j, $w_{ji}=w_{ij}$, and (2) $w_{14}=w_{23}=w$ (say). Indeed it would be possible to go further and set $w=1$, but they did not do this. The next thing to do was to set up fitness coefficients for each of the four types individually, and this they achieved by setting $w_i=w_{i1}x_1+w_{i2}x_2+w_{i3}x_3+w_{i4}x_4$ for each of the four possible values of i. The overall mean fitness is then found to be

$$\overline{w}=w_1x_1+w_2x_2+w_3x_3+w_4x_4\ .$$

Lewontin and Kojima then gave equations for the progress of the selection process. Here, however, I will only quote the equilibrium case, achieved when selection is complete. We have in this instance:

$$\begin{aligned} x_1(w_1 - \overline{w}) &= RwD \\ x_2(w_2 - \overline{w}) &= -RwD \\ x_3(w_3 - \overline{w}) &= -RwD \\ x_4(w_4 - \overline{w}) &= RwD, \end{aligned} \qquad (1)$$

where R is a number between 0 and 0.5, and D is a shorthand for $x_1x_4 - x_2x_3$. It is the presence of the right-hand sides that distinguishes this case from the other one discussed earlier.

They arise because of a complication resulting from the biological process by which the sex-cells (*gametes*, i.e. sperm and ova) are formed. The genes on the different chromosomes do not stay in their original configurations, but recombine into different patterns. R is in fact the probability that such recombination occurs between the two loci involved. D is a measure of the effect of that recombination.

These equations allowed Moran to look again at the theory that had been advanced for the case involving two loci. In 1964 he published in *The Annals of Human Genetics* a paper called "On the nonexistence of adaptive topographies", that threw out much of what had been accepted up till then. As he wrote: "The purpose of the present paper is to show that the above theory [the study of adaptive topographies] and all the quoted work based on it is wrong because when there exists general selection the genotypes at one locus do not associate at random with the genotypes at the other locus."

Specifically he demonstrated that:

- the mean fitness $\overline{w}$ is not a function of the two variables p_1 and p_2 introduced above, but rather of three variables (any three of x_1, x_2, x_3, x_4);
- the situation satisfied by Equations (1) did not necessarily maximise $\overline{w}$, and that
- there were even situations where selection could actually *decrease* $\overline{w}$.

I claim for myself the credit for finding a fourth paradox arising from Equations (1):

- The fittest of the combinations *AB*, *Ab*, *aB*, *ab* can never be the most frequent.

The proof is perfectly straightforward. Suppose for definiteness that *AB* is the fittest combination (it makes no difference which one we choose). Then w_1 is the largest of the coefficients w_i. Then in particular, $w_1 > w_4$. But now combine the first and the fourth of Equations (1) to find

$$x_1(w_1 - \overline{w}) = x_4(w_4 - \overline{w}),$$

from which it follows that $x_1 < x_4$, and thus there are more *ab*s than there are *AB*s, even though these latter are more fit!

The work of Moran in particular implied that if selection was maximising something, then that something was not $\overline{w}$. There began a search for what that something could possibly be. The answer, when it came, was rather surprising.

Suppose that there are n alleles at a single locus. This is the case already decided by the work of Kingman and his predecessors. If an individual has an allele of type i and another of type j, then the genotype can be represented as ij, and such an individual would be assigned a fitness w_{ij}. If the frequency of the type i allele is q_i and that of type j is q_j, then the mean fitness $\overline{w}$ is the sum of all the products $w_{ij}q_iq_j$ taken over all the values of i and j. The change in the value of $\overline{w}$ from one generation to the next is caused by changes in the values of the frequencies q_i (and q_j). Write Δq_i for the change in q_i, etc.

The overall change in $\overline{w}$ is the sum of two terms, each itself a sum of other terms. The first is twice the sum of all the products $w_{ij}q_i\Delta q_j$, and the second is the sum of all the products $w_{ij}\Delta q_i\Delta q_j$. In some cases, this second term is small compared to the first, and if it is valid to ignore it then the mean fitness will increase. But a better way to express this is to say that the *partial increase in mean fitness* is always positive.

This result had been advanced by Fisher back in the 1920s, and named by him as "the fundamental theorem of natural selection". Later, however, it was thought to have had only limited validity, or else to have the status merely of an approximation. However, in 1972, G R Price, writing in *The Annals of Human Genetics*, suggested that Fisher had been misunderstood, and that he had meant that the *partial* increase in mean fitness was always positive, not that the mean fitness itself always increased. He wrote: "The mystery and the controversy [over the 'fundamental theorem'] result from incomprehensibility rather than error."

Later work by W J Ewens (then at Monash University) led to his being able to construct a function that does indeed increase under the influence of natural selection. This he published in the journal *Theoretical Population Biology* in 1989. Ewens agreed with Price that Fisher had been misunderstood, and was also able to show that his function always increased no matter what biological complications were included and no matter how complicated the underlying equations might become.

What is lost, though, is the immediacy of biological interpretation. As Ewens wrote: "An interpretation of this theorem is put forward here which implies that it is correct as a mathematical statement, but of less biological value than claimed by Fisher".

The old, but incorrect, understanding had been that $\overline{w}$ itself increased from one generation to the next. This makes for a simplicity of interpretation, that is, however, deceptive. Reading popular works on Genetics, such as *The Selfish Gene*, one wonders quite how much of the true mathematical theory has been absorbed!

Thoughts along these lines have found trenchant expression in the writing of the mathematician Ian Stewart (in his book *Life's Other Secret*, Chapter 12). Here is a sample.

"... it is not unusual to be told that people are having children 'in order to pass on their genes to future generations'. ... but I know that when I was deciding to have children, I didn't pay much attention to my genes at all. I blame this kind of nonsense on a widespread misunderstanding of the *selfish gene viewpoint*, which maintains that the only reason we exist is so that our genes can reproduce. .... However, it is equally possible to promote the *slavish gene theory*, in which genes worry enormously about the survival of their organisms. (If the genes don't produce a viable organism, they die out, right?)"

COMPUTERS AND COMPUTING

"Computation is Exclusive"

The title of our column in this issue is taken from the webpage

http://www.mathpages.com/home/kmath106.htm

which begins by discussing the notion of a "computable number". This is not perhaps quite what one might expect. While it does include numbers like $\sqrt{4}$ whose *exact* value may be computed, it also includes others like $\sqrt{2}$ for which this is not the case. However it is possible to write an algorithm in finite terms that will compute $\sqrt{2}$ to any desired degree of accuracy. (You might like to try this as an exercise!) Although the values achieved by such a program will never *exactly* equal $\sqrt{2}$, the various values that *are* achieved converge to this value and to no other.

It is like the case of the sequence 1, ½, ¼, … , whose terms approach, but never attain, the value 0. If we continue the sequence long enough, we can get arbitrarily close to 0, and no number other than 0 has this relation to the sequence.

Similarly to compute $\sqrt{2}$ or π, say, we need to construct a sequence of approximations that converge, in this same sense, to $\sqrt{2}$ or π. We can produce ever better and better approximations to the true values of these numbers without ever actually reaching them.

However, because of the convergence property, we can positively *rule out* any incorrect value – just as long as we have enough time and computing power! This is what is meant by "exclusive" in the title of this column.

From time to time, people come up with the notion that the accepted theory of (say) π is wrong. Recently we had a letter from a reader who espoused such a view. His value differed from the accepted decimal expansion in the third decimal place. However, the area of a unit circle, i.e. π, can be expressed as being smaller than that of a regular polygon whose sides are all tangents to it. In the case of our correspondent, it may be shown that his value is greater than that of an enclosing regular 64-gon, whose area in its turn is greater than π.

SOLUTION TO 'SLITHER'

First, let's look at the standard 5×6 game. Two of the sides have 6 dots along them. This means that each has 5 intervals between adjacent dots. Player *A*, moving first, takes the central interval in either one. Thereafter, *A* mirrors every move the opponent *B* makes. Clearly, *A* cannot be the first to close the path, as such a move would be a mirror image of a move by *B*, which would already have closed the path, and so lost the game. This is Professor Read's "monumental triviality".

Notice that this same strategy applies to all rectangular grids in which the number of dots is even. There will be either two or four sides with an even number of dots along them, and any one of these may be used to initiate the same strategy.

Thus, if the number of dots is even, the first player has an extremely simple winning strategy. It is not, however, the only possible one. There are many more. The illustrations below apply to a 4×4 grid, but the principles invoked are general. In this case there are 16 dots and 24 possible first moves. From these 24, pick out a set of exactly 8 with the following two properties:

1. No two share a common end-point,
2. Each is the beginning or end of exactly one possible connection.

Such moves will be called 'favoured moves'. It is always possible to choose a set of favoured moves, and in fact it is possible to do this in many ways. For the 4×4 case, Figures 1, 2 and 3 below show some of the possibilities. The third illustrates a pattern whose analogues can always be set up.

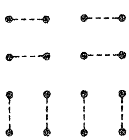


Figure 1

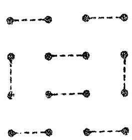


Figure 2

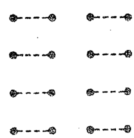


Figure 3

A generalised winning strategy for A, playing first, is to make any one of the favoured moves. This will force B, the opponent, to make a non-favoured move. A will now be able to make another favoured move, and so it goes. B will never have the opportunity to make a favoured move. Because the favoured moves are never connected to one another, no favoured move can close the path. Eventually B will be forced to make a closing move (return to a previously visited point) and so lose the game.

Notice that there are mn dots and that each dot belongs to exactly one of a set of pairs, the favoured moves. It follows that to set up the winning strategy requires mn to be even, as there must in general be $mn/2$ favoured moves.

So what about the case mn odd? In this case B has the winning strategy, but it is not as simple to apply because the initiative still lies with A, who moves first. However, *whatever A does*, B has a forced win. This is illustrated in Figures 4 and 5, which use the case of a 5×5 grid for the demonstration.

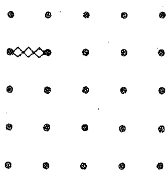


Figure 4

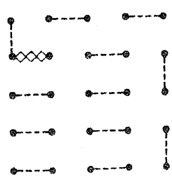


Figure 5

To exploit a winning strategy, B must first wait to see what A does. An example is displayed in Figure 5. But now one end or other of the interval actually chosen by A must lie an even number of intervals away from each of the four corner points. In the case shown in Figure 4, it is the right-hand that lies 2 intervals from the top left-hand corner (or 4 from the bottom left-hand one, etc).

B must avoid this point and consider the remaining 24 points of the grid. These 24 can be paired off into 12 pairs that form a set of favoured moves. One possibility is illustrated in Figure 5. One point in the set of 24 will be the other end of the interval chosen by *A*. From *this* end, *B* makes a favoured move. Note that there will always be such a move. The game now proceeds as in the earlier case, except that it is now *B* who forces the win.

You might care to explore specific cases of simple grids. The 2×2 case is completely trivial, and the 2×3 case not much harder; however, the 3×3 case may be completely analysed without too much headache, and the next few cases are easily accessible.

A CORRECTION AND AN ADDENDUM

In the course of the Computer column in our previous issue, we inadvertently misspelt the name of Professor W M Kahan several times. Our apologies to everybody!

Our discussion of Avni Pllana's "Tie-knot problem" (Problem 27.4.3) showed two possible interpretations of the problem, leading to two somewhat different solutions. For more on tie-knots, see *The Australian Mathematics Teacher, Volume* **60**, *Part* **1** (March 2004), p 32.

"Although to penetrate into the intimate mysteries of nature and thence to learn the true causes of phenomena is not allowed to us, nevertheless it can happen that a certain fictive hypothesis may suffice for explaining many phenomena."

Leonhard Euler, 1748

PROBLEMS AND SOLUTIONS

First the solutions to the problems set in *Volume* **27**, *Part* **5** (October 2003).

SOLUTION TO PROBLEM 27.5.1 (Submitted by Willie Yong (Singapore), Jim Boyd (USA) and Richard Palmaccio (USA), jointly).

The problem read:

Evaluate $4\sin 20^\circ + \tan 20^\circ$.

Solutions were received from Keith Anker, Šefket Arslangić (Bosnia), John Barton (2 solutions), Julius Guest, Joseph Kupka, Carlos Victor (Brazil) and the proposers. Here is Barton's first.

Inserting tabulated values suggests that the required value is $\sqrt{3}$.

Then:

$$\sqrt{3} = \tan 60^\circ = \tan(3\times 20^\circ) = \frac{\sin(3\times 20^\circ)}{\cos(3\times 20^\circ)} = \frac{3\sin 20^\circ - 4\sin^3 20^\circ}{4\cos^3 20^\circ - 3\cos 20^\circ}$$

$$= \tan 20^\circ \times \frac{3-4\sin 20^\circ}{1-4\sin 20^\circ} = \tan 20^\circ \left\{1 + \frac{2}{1-4\sin 20^\circ}\right\}$$

$$= \tan 20^\circ + \frac{2\sin 20^\circ}{\cos 20^\circ - 2\sin 20^\circ \sin 40^\circ}$$

$$= \tan 20^\circ + \frac{2\sin 20^\circ}{\cos 20^\circ - (\cos 20^\circ - \cos 60^\circ)}$$

$$= \tan 20^\circ + 4\sin 20^\circ, \qquad \text{since } \cos 60^\circ = 1/2.$$

This sequence of identities is reversible, which establishes the result.

Barton's second proof was shorter, but less transparent.

SOLUTION TO PROBLEM 27.5.2 (Submitted by Šefket Arslangić (Bosnia))

The problem read

Prove that

$$\frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{2n} < \frac{1}{\sqrt{2}}$$

for all positive integers n.

Solutions were received from Keith Anker, John Barton, Julius Guest, Joseph Kupka, Carlos Victor (Brazil) and the proposer, most of whom proved a somewhat stronger result. Here is Kupka's.

Let $S_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$.

Then $S_{n+1} = \frac{1}{n+2} + \frac{1}{n+3} + \cdots + \frac{1}{2n} + \frac{1}{2n+1} + \frac{1}{2n+2}$.

So $S_{n+1} - S_n = \frac{1}{2n+1} + \frac{1}{2n+2} - \frac{1}{n+1} = \frac{1}{2n+1} - \frac{1}{2n+2} > 0.$

Thus $\{S_n\}$ is an increasing sequence. Furthermore

$$\begin{aligned} S_n &= S_1 + (S_2 - S_1) + (S_3 - S_2) + \ldots + (S_n - S_{n-1}) \\ &= \frac{1}{2} + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) + \cdots + \left(\frac{1}{2n-1} - \frac{1}{2n}\right) \\ &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots + \frac{1}{2n-1} - \frac{1}{2n} \end{aligned}$$

This last series is the partial sum of an infinite series whose limiting value is known to be ln 2. Because $\{S_n\}$ is an increasing sequence, the limiting value is greater than any individual partial sum.

Thus $S_n < \ln 2 \ (\approx 0.693) < \frac{1}{\sqrt{2}} \ (\approx 0.707)$.

SOLUTION TO PROBLEM 27.5.3 (Submitted by Julius Guest)

The problem read:

Let

$$S_n = \frac{1^2}{2\times3\times4\times5} + \frac{2^2}{3\times4\times5\times6} + \cdots \frac{n^2}{(n+1)(n+2)(n+3)(n+4)}$$

Find an explicit formula for S_n and determine $\lim_{n\to\infty} S_n$.

Solutions were received from Keith Anker, Šefket Arslangić (Bosnia) (2 solutions), John Barton, Joseph Kupka, Carlos Victor (Brazil) and the proposer. All were rather similar and so what follows is a composite.

It may be proved that

$$\frac{n^2}{(n+1)(n+2)(n+3)(n+4)} = \frac{1}{6}\frac{1}{n+1} - \frac{2}{1}\frac{1}{n+2} + \frac{9}{2}\frac{1}{n+3} - \frac{8}{3}\frac{1}{n+4}$$

[This may readily be proved once we know it; to establish it in the first place requires partial fractions:]

Then

$$S_n = \frac{1}{6}\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n+1}\right) - 2\left(\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{n+2}\right)$$
$$+\frac{9}{2}\left(\frac{1}{4}+\frac{1}{5}+\cdots+\frac{1}{n+3}\right) - \frac{8}{3}\left(\frac{1}{5}+\frac{1}{6}+\cdots+\frac{1}{n+4}\right)$$

But $\frac{1}{6} - 2 + \frac{9}{2} - \frac{8}{3} = 0$, so many of these terms cancel out. After some simplification, we are left with the answer the first question:

$$S_n = \frac{5}{36} - \frac{1}{6}\frac{1}{n+2} + \frac{11}{6}\frac{1}{n+3} - \frac{8}{3}\frac{1}{n+4}$$

As $n\to\infty$, $S_n \to \frac{5}{6}$ which answers the second question.

SOLUTION TO PROBLEM 27.5.4 (Submitted by Keith Anker)

The problem read:

> Lines l_1 and l_2 are perpendicular to one another and lie in the plane of a triangle ABC. Using only measurements in the directions of l_1 and l_2, determine the area of ABC.

Solutions were received from Šefket Arslangić (Bosnia), John Barton, Carlos Victor (Brazil) and the proposer.

Details varied, but all came down to the recognition that l_1 and l_2 could be taken as the axes of a rectangular co-ordinate system. Suppose then that $A=(x_1,y_1)$, $B=(x_2,y_2)$ and $C=(x_3,y_3)$ in these co-ordinates. The result is now a standard one. The area is given by the value of the determinant

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

Barton cites Sommerville's *Analytical Conics* and Osgood & Graustein's *Plane and Solid Analytic Geometry* as examples of texts where the result may be found.

We close with four new problems.

PROBLEM 28.3.1 (submitted by Julius Guest)

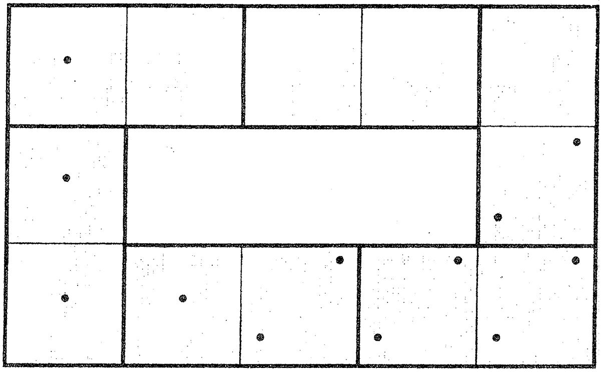
> Prove that all the points of inflection of the curve $y=\sin x$ lie on the curve $y^2=\dfrac{4x^2}{x^2+4}$.

PROBLEM 28.3.2 (submitted by Keith Anker)

> Let A be the region contained between the x-axis and the parabola $y=1-4x^2$. Determine the largest rectangle that can be inscribed within A.

PROBLEM 28.3.3 (submitted by Paul Grossman)

Let us define a *Domino set of rank n* as a set of tiles, the rectangular faces of which are separated into two squares, each marked with dots representing numbers from zero to n, such that no two tiles contain the same pair of numbers and all combinations of pairs are represented.



The figure shows a set of rank 2 laid out in a closed chain. The contacting squares on adjoining tiles have matching numbers and each tile was placed in the clockwise direction at the end of the previous tile, either in the same direction or at right angles. Now:

1. Prove that a closed chain with the above conditions can be established with a set of rank 6 (the standard domino set) but not with sets of rank 3, 4 or 5.

2. Show what ranks will allow a continuous chain to be formed with matching numbers on adjoining squares and tiles placed at the end of the previous tile.

PROBLEM 28.3.4 (submitted by Šefket Arslangić, Bosnia)

Let $x^2+y^2+z^2+2xyz=1$, where $x,\ y,\ z\ \geq 0$.

Prove that $x^2+y^2+z^2 \geq \dfrac{3}{4}$.

PH: +61 3 9905 4432; email: barbara.hardie@sci.monash.edu.au

* * * * *

Published by the School of Mathematical Sciences, Monash University