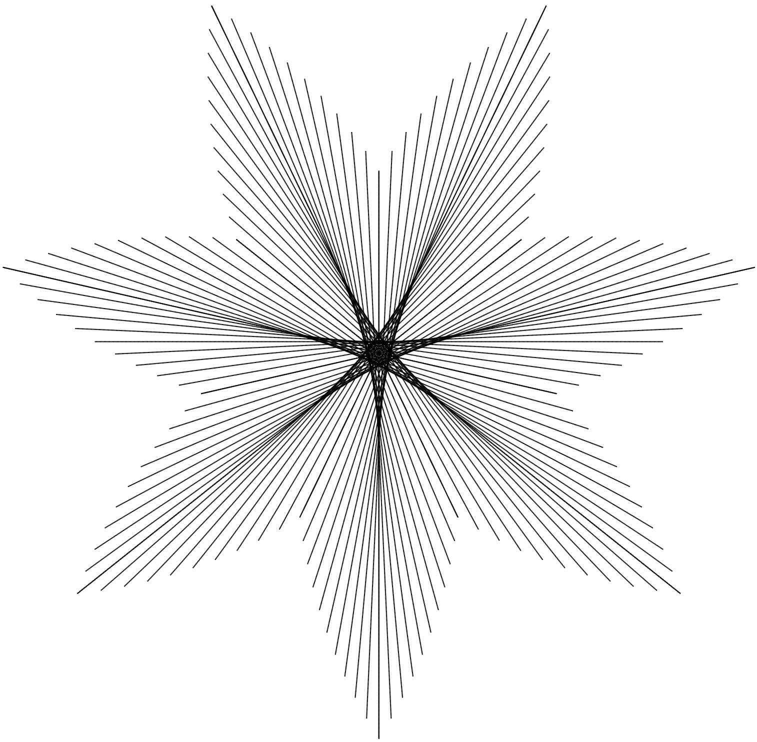
*

ISSN 1350-8539

M500 223



The M500 Society and Officers

The M500 Society is a mathematical society for students, staff and friends of the Open University. By publishing M500 and 'MOUTHS', and by organizing residential weekends, the Society aims to promote a better understanding of mathematics, its applications and its teaching. Web address: www.m500.org.uk.

The magazine M500 is published by the M500 Society six times a year. It provides a forum for its readers' mathematical interests. Neither the editors nor the Open University necessarily agree with the contents.

MOUTHS is 'Mathematics Open University Telephone Help Scheme', a directory of M500 members who are willing to provide mathematical assistance to other members.

The September Weekend is a residential Friday to Sunday event held each September for revision and exam preparation. Details available from March onwards. Send s.a.e. to Jeremy Humphries, below.

The Winter Weekend is a residential Friday to Sunday event held each January for mathematical recreation. For details, send a stamped, addressed envelope to Diana Maxwell, below.

Advice to authors. We welcome contributions to M500 on virtually anything related to mathematics and at any level from trivia to serious research. Please send material for publication to Tony Forbes, above. We prefer an informal style and we usually edit articles for clarity and mathematical presentation. If you use a computer, please also send the file to tony@m500.org.uk.

Contra Cantor

Sebastian Hayes

Passing in review the various paradoxes, linguistic and mathematical, that bothered logicians around the beginning of the last century, Russell and Whitehead in their *Principia Mathematica*—I shall henceforth just say Russell—found that 'they all result from a certain kind of vicious circle' that consists in 'supposing that a collection of objects may contain members which can only be defined by means of the collection as a whole' (R & W, 37). As an example of what they had in mind they cited the statement **'All propositions are either true or false'**. Russell comments:

> It would seem that such a statement could not be legitimate unless 'all propositions' referred to some already definite collection, which it cannot do if new propositions are created by statements about 'all propositions'. (R & W, 37).

More mathematical examples are the Set of All Sets—is it a member of itself?—or Burali-Forti's[1] paradox of the Ordinal Number of All Ordinals.

Russell suggests stopping such statements being made, or at any rate being accepted as meaningful by logicians—'Whatever involves *all* of a collection must not be one of the collection' (R & W, 37). Poincaré coined the term 'impredicative' for statements that define an object in terms of a collection to which the object being defined belongs. He considered that impredicative definitions should be banned from mathematics.

But what happens if we *do* want to talk about 'all' the members of such a collection? This, Russell assures us, need not pose any insuperable difficulties. A statement about 'all' of a certain collection is of 'higher type' than a statement about specific members of the collection and in consequence must be excluded from the range of application of the statement. The Set of All Sets is 'of higher type' than any Set you like to mention which will be one of its members, and so we do not get the ridiculous situation of the Set of All Sets being at one and the same time a member, and yet not a member, of itself.

At first sight Russell's solution sounds both sensible and effective. However, it soon became a major embarrassment to him, for not only did strict application of the theory of types make a lot of proofs very cumbersome it actually invalidated a lot of them. As Weyl and others pointed out, analysis turned out to be littered with impredicative formulae. This stimulated the

[1] Apparently, contrary to what I have always believed up to now, Burali-Forti is a singleton set from the Set of All People, being not two persons but one.

Intuitionists to reformulate the whole subject but Russell had no intention of taking such a heroic course. He states airily in the Introduction to the 1927 re-edition of *Principia Mathematica* that 'though it might be possible to sacrifice infinite well-ordered series to logical rigour, the theory of real numbers ... can hardly be the object of reasonable doubt' (R & W, xlv, 1927). But why not? Russell's reply sounds suspiciously like an eighteenth century clergyman's assertion that 'the eternal existence of a Creator God cannot seriously be questioned'.

Subsequent mathematical discussion of these issues has clouded rather than clarified the basic principles at stake: in particular far too much attention has been given to the validity or otherwise of the so-called Axiom of Choice. As it seems to me, the problem is not 'impredicative statements' as such—this is something of a red herring—but a failure to distinguish between 'definite' sets and 'indefinitely extendable' sets. By definition the former are fully constituted once and for all, and thus listable, whereas the latter are not. Confusing the two is the real 'category mistake' at the root of all the kerfuffle.

In conversation we normally deal with two, and only two, types of sets or collections, those that are what I call definite and those that are continually being extended. The persons living in the UK at the present moment constitute a definite set which can be (and actually is) listed—at any rate within the bounds of bureaucratic error. The set of all human beings, past, present and future, is not a definite set but a continually expanding one, and one that will presumably continue to expand as long as the species exists.

Self-referential statements of the type 'Whatever I say is untrue' only cause trouble because there is a certain ambivalence about the type of collection we are dealing with. One schoolboy philosopher exclaims to another during the lunch-break, "You know, there's not a single thing I'm sure about!" His companion rejoins, "Ah! but there is one thing at least you're sure about, and that is that you aren't sure about anything!"

Sceptic's first statement only referred to the fully definite set of all beliefs he had *actually considered up to that moment*, and a standpoint of all-round scepticism was not one of them. It would be quite perverse to consider his first statement as referring to the collection of all possible beliefs the human species might conceivably entertain. The belief 'I don't believe in anything' was not, at the beginning of the discussion, a member of the Set of All Beliefs Sceptic Had Considered (a definite set) but after the end of the conversation it was. His first statement was *time and context dependent*: it was not an intemporal assertion.

At a future date Sceptic might say, "I'm not sure about anything—except the statement I made to you yesterday that I wasn't sure about anything I'd considered up to then." The Set of Beliefs Sceptic Was Sure About starts off empty, then contains one member, perhaps goes on to containing two members, and so on.

All this hardly seems worth dwelling on. So why the fuss? Because, when it comes to *mathematics*, the situation is very, very different. Mathematical assertions are not generally considered to be time and context dependent, they are in some sense held to be 'eternally true', true even before human beings or the universe we live in existed.

So far it has not been necessary to introduce the fatal word 'infinite' but it cannot be withheld any longer. Can any so-called 'infinite' set ever be a fully constituted totality, a 'definite set'? I do not see that it can. The only sensible way of treating 'infinite' sets is to view them as open-ended partly definite sets which can be extended as far as we wish. This is entirely in line with the way we proceed in normal speech and conversation—which, one strongly suspects, is the main reason mathematicians disapprove of such an approach.

What we must above all not do is treat an open-ended indefinite set as a *fully constituted one*. But in mathematics, ever since the advent of Cantor, this is exactly what is done in mathematics. This is the essential 'category mistake', not Russell's 'self-referential misapprehension'. Some mathematicians, notably Cantor himself, were frank enough to put their hands on the table and declare that they really did believe in the existence of the transfinite. Even Russell, though at the time a positivist, introduced into his *Principia* the controversial Axiom 'That infinite classes exist' (R & W, *120.03). Most modern mathematicians are, however, content to evade the issue: as Davis and Hersh point out, the modern mathematician is two things at once, a Platonist in the study, but a Formalist when confronting the outside world.

Cantor's proofs

Cantor's proofs are of two main types, one acceptable (to me), one not. Let us first take his proof that the rational numbers between 0 and 1 form a null set. This depends on two prior results, his ingenious diagonalization of $\mathbb{Q}$, the rational numbers, and the well-known limit (as usually formulated)

$$\lim_{n\to\infty}\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots+\frac{1}{2^n}\right) \;=\; 1.$$

Since, for any positive rational number you like to name, say $1/N$, I can always find a smaller one, namely $1/(N+1)$, it looks at first sight as if it were impossible to list the rational numbers, first, second, third &*c*.; i.e. put them in one–one correspondence with $\mathbb{N}$, the natural numbers. But Cantor showed how this could be done. For example, those between 0 and 1 can be listed as follows:

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \ldots.$$

This is not an ordering by increasing or decreasing size but that does not matter, nor does it matter (too much) that there will be some redundancy—2/4 will appear though we already have 1/2. The point is that given any specified fraction between 0 and 1, it will eventually crop up and can be attributed an ordinal from the natural numbers, hundred and seventy-seventh, ten-thousandth, or what have you. We do not need to know what this ordinal is, but we do know that we can provide it if challenged to do so if given enough time. There is nothing objectionable in this procedure since we do not have to envisage the rational numbers between 0 and 1 as a definitively constituted totality existing in some Platonic Never Never Land—though this is undoubtedly how Cantor himself viewed them. The sequence $S = 1/2, 1/4, 1/8, \ldots, 1/2^{n-1}$ is a geometric sequence with constant ratio 1/2. The terms are respectively $t_1 = 1/2$, $t_2 = 1/2^2$, $\ldots$, $t_n = 1/2^n$. If we take partial sums S_1, S_2, $\ldots$, S_n we have

$$S_1 = \frac{1}{2} = \left(1 - \frac{1}{2}\right), \quad S_2 = \frac{1}{2} + \frac{1}{4} = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{4}\right) = 1 - \frac{1}{4}$$

and $S_n = \left(1 - \dfrac{1}{2^n}\right) < 1$ for all $n \in \mathbb{N}$.

The series S_p of successive partial sums is clearly, in my terms, not a fully constituted totality but an indefinitely extendable one. Many slapdash authors, who ought to know better, talk about 1 being 'the sum to infinity' of the series: in fact, as is generally the case with series, the limit 1 is unattainable and there is no definitive sum, only a perpetually changing one as n increases, which is why we speak of 'partial' sums, though the word is misleading.

Cantor now invites us to construct a sequence of open intervals where each interval $\{I_n\}$ has centre r_n. Each interval starts at the point $r_n - k/2^{n+1}$ and ends at the point $r_n + k/2^{n+1}$ so it has length twice $k/2^{n+1}$ or $k/2^n$. Since we have got a way of listing the rational numbers we drop them one after the other into these intervals. And the total length of n intervals is

$$\frac{k}{2} + \frac{k}{2^2} + \frac{k}{2^3} + \cdots + \frac{k}{2^n} \;=\; k\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots\right) \;=\; k\left(1 - \frac{1}{2^{n-1}}\right) \;<\; k$$

since $1 - 1/2^{n-1} < 1$. Provided we can decrease k as much as we wish, we can squeeze 'all' the rationals between 0 and 1 into an arbitrarily small compass. So a line segment a foot (or a millimetre) long is nonetheless capable of containing an 'infinite' quantity of numbers, many more than there are stars in the sky. Cantor has thus, to his own satisfaction at least, shown that the rationals between 0 and 1 are what he calls a 'null set': they take up so little space it's as if they weren't there at all.

One might baulk a little at this over-literal way of considering numbers as points on a line (which they are not) and, of course, in the real world there would be a definite limit to the size of k—it could not be made smaller than that of an elementary particle, for instance. However, one might be prepared to let this pass as temporary exercise of mathematical licence. The main thing is that there is no need to view this procedure as having been carried out for 'all' the rationals $0 < q < 1$ but only for as many as someone likes to mention. It is usually stated in maths books that this result (the infinite compressibility of $\mathbb{Q}$) is 'counter-intuitive': it would be more accurate to describe it as being totally unrealistic. This is not a problem if we make sure to continually bear in mind that a mathematical model or construction is not itself part of the physical world to which it can sometimes be applied successfully.

Other bizarre results such as the length of the Koch curve fall into the same category. Starting with an equilateral triangle, then building one on the middle third of each side and continuing in this way *ad infinitum*, it appears that the perimeter of the curious jagged figure can be made to exceed any stipulated length provided you go on long enough even though the whole creature can be inserted in a disc of, say, radius one metre. Of course, in any actual situation there would once again be a limiting size beyond which it would not be possible to go: there is certainly no need to conclude that we have here a case of 'infinity in the palm of your hand' (Blake), though some people seem to think so.

If now we pass to Cantor's 'proof' that the real numbers are not denumerable, we have a very different kettle of fish. A collection is considered denumerable if it can be put in one-one correspondence with $\mathbb{N}$, the natural numbers—broadly speaking can be listed. We have seen that this is possible for the rationals between 0 and 1. Cantor now invites us to consider an enumerated list of all the real numbers (rationals + irrationals) between 0 and

1. These reals are exhibited in the form of non-terminating decimals—any other base would be just as good. To avoid ambiguity a fraction like 1/5 has been listed as 0.19999999… instead of 0.2—absurd though it would be to do any such thing. So there they all are:

$$\begin{aligned} s_1 &= 0.a_{11}a_{12}a_{13}\ldots \\ s_2 &= 0.a_{21}a_{22}a_{23}\ldots \\ s_3 &= 0.a_{31}a_{32}a_{33}\ldots \\ &\ldots, \end{aligned}$$

where every a is a natural number between 0 and 9.

Cantor now produces out of a hat a 'number' that has not appeared in the list, call it b. We concoct b by 'doing the opposite' as it were. If a_{11} is 1, make b_1 (the first digit of b) $= 2$, if $a_{11} \neq 1$, make $b_1 = 1$. Likewise for a_{22}, a_{33}, giving b_2, b_3 and so on. This defines $b = 0.b_1b_2b_3\ldots$. But this 'number' has not appeared in the list since it differs from s_1 in the first place, from s_2 in the second place and so on. Therefore, the real numbers between 0 and 1 are not denumerable, and since these are only a small part of $\mathbb{R}$ as a whole, $\mathbb{R}$, the Set of all Reals is not denumerable—a paradoxical result since $\mathbb{N}$ is already an 'infinite' set, so $\mathbb{R}$ must be of a higher type of infinity than $\mathbb{N}$, Q.E.D.

Now this proof by contradiction wholly depends on the original assumption that *all* the reals between 0 and 1 have been listed—not one has been left out. Since Cantor shows one that *has* been left out, the assumption must have been wrong in the first place. However, if we do not view the reals between 0 and 1 as a wholly constituted totality, listable and enumerable, but as an open-ended extendable set, the argument collapses like a burst balloon. All that could ever be on view at a single moment in time is an array of decimals—or digits in some other base r—taken to n places. A competing generator handled by Cantor in person cannot produce any real number for given r and n which is not on show since all possibilities are covered. All Cantor can do is print out an arbitrary 'diagonal' rational number between 0 and 1 to m places with $m > n$.

Since the base used is immaterial let us use base 2 and print out numbers between 0 and 1 using only the symbols 0 and 1. In the first print out we only go so far as one digit, then we print out all numbers with two digits after the point and so on. We have

$$0.0, 0.00, 0.000; 0.1, 0.01, 0.001; 0.10, 0.010;$$

$$0.11, 0.011; 0.100; 0.101; 0.110; 0.111.$$

To keep ahead Cantor has to counter with a number containing at least one more digit after the point, but, whatever number he chooses, this number will appear in the next print out. Thus the struggle is ding-dong and inconclusive. It should be stressed that the ability to view $\mathbb{R}$ as a whole does not depend on our limited range of vision or the size of the memory of the computer or any other technicality: the reals are simply not exhibitable in their full extent because strictly speaking they do not have a 'full extent'. Even God would not be able to view 'all' the real numbers at one fell swoop because there is no 'all' to view.

Very similar is Cantor's 'proof' that, for all non-empty sets A, the cardinality of A is less than the cardinality of the power set of A. (The power set, remember, consists of the sets that can be constructed from the members of A; e.g. if $A = \{1, 2, 3\}$, then $P(A)$ consists of A itself $= \{1, 2, 3\}$, also the sets $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$, the singleton sets $\{1\}$, $\{2\}$ and $\{3\}$ and $\emptyset$, the empty set.) Obviously, for ordinary 'finite' sets the theorem holds, but, since things become so disconcerting when we pass to consider transfinite sets, Cantor wonders whether it remains valid.

In typical fashion Cantor proceeds to assume that an exhaustive mapping from A to $P(A)$ has been carried out. Since for any a we can, *faute de mieux*, pair it off with the set of which it is the sole member, namely $\{a\}$ and this is far from exhausting all the possibilities, Cantor concludes that the cardinality of $P(A)$ cannot be less than the cardinality of A. We are now invited to consider the set B given by $B : \{a \in A, a \neq f(a)\}$; i.e. the set containing all those elements which are not members of the sets they have been paired off with in the mapping. It would seem that B is non-empty. But if so, B, being a bona fide subset of $P(A)$ must have a pre-image under this mapping, a_b say, i.e. there is an a_b in A such that $f(a_b) = B$. But a_b itself must either belong to B, or not belong to B. We find that if it does it doesn't and if it doesn't it does. Thus contradiction. Therefore there can be no such mapping $f : A \to P(A)$ and so card $A <$ card $P(A)$. Q.E.D.

This argument is worthless because Cantor has envisaged a mapping that cannot ever be carried out in full, even in theory: he is treating an ongoing, indefinitely extendable mapping as a completed act.

Sets with oscillating membership

If we regard the proposed function f, not as already existent, but as in the process of being defined, we get a different picture. Suppose we have carried out a bijection from A to $P(A)$ to n places—which is all we can ever hope to do—and we have a non-empty set B satisfying Cantor's condition, namely that individual members of B have not been paired off with themselves viewed as sets. But B does not as yet have a pre-image in A; so, noting this, we pick some element in A not yet used, a_b say, and form (a_b, B).

Now, prior to its being assigned an image under the function f, the element a_b did not have an image; however, now that it has acquired one we realize that it has automatically become a member of B (which it was not before) and so is disqualified from being the pre-image of B. We thus remove a_b from (a_b, B) and look for another pre-image. The same situation develops and one might justifiably conclude that, since we are perpetually going to have to change B's pre-image as soon as we assign one, then any function of the desired type A to $P(A)$ is going to be of a very provisory nature and so, we might decide, for this reason, to conclude that the cardinality of $P(A)$ must be 'greater' than that of A. This is not quite what Cantor says though. This oscillating procedure whereby one or more element changes sets perpetually is entirely normal outside mathematics—in fact it is really only in the unreal world of mathematics that sets ever do get constituted definitively once and for all. Individuals are always changing their set membership as their age changes, as their beliefs mature, as the frontiers of countries are redrawn and so forth. Even species evolve and change into radically different ones, so we are told, and nothing stays exactly the same for very long.

A typical example of 'oscillating membership' is provided by Russell's Village Barber Paradox though Russell did not realize this. Russell invites us to consider a Village Barber who claims he shaves everyone in the village who does not shave himself and only such persons. The big question is: Does he shave himself? If he does shave himself, he shouldn't be doing so—since, as a barber, he shouldn't be shaving self-shavers. On the other hand, if he doesn't shave himself, that is exactly what he ought to be doing.

The contradiction only arises because Russell, like practically all modern mathematicians, insists on viewing sets as being constituted once and for all in the usual Platonic manner. Let us see what would actually happen in real life. It is first of all necessary to define what we mean by being a self-shaver: how many days do you have to shave yourself consecutively to qualify? Ten? Four? One? It doesn't really matter as long as everyone agrees on a fixed length of time, otherwise the question is completely meaningless. Secondly, it is important to realize that the barber has not always been the Village Barber: there was a time when he was a boy or perhaps inhabited a different village. On some day d he took up his functions as Village Barber in the village in question. Suppose our man has been shaving himself for the last four days prior to taking on the job, so, if four days is the length of time needed to qualify as a self-shaver, he classes himself on day d as a self-shaver. He does not get a shave that day since he belongs to the self-shaving set and the Village Barber does not shave such individuals.

The next day he reviews the situation and decides he is no longer in the self-shaving category—he didn't get a shave the previous day—so he shaves himself on day 2. On day 3 he carries on shaving himself—since he has not yet got a run of four successive self-shaving days behind him. This goes on until day 6 when he doesn't shave himself. The barber spends his entire adult active life oscillating between the self-shaving and the non-self-shaving sets. There is nothing especially strange about this: most people except strict teetotallers and alcoholics oscillate between being members of the set of drinkers and non-drinkers—depending of course on how much and how often you have to drink to be classed as a 'drinker'.

This example was originally chosen by Russell to show that the self-referential issue has nothing necessarily to do with infinity. Nor does it, but it does depend on the question of whether sets or collections are treated as time and context dependent or not.

Conclusion

One understands, of course, why mathematics as the exact science *par excellence* does not want to be bothered with such messy creatures as sets with varying membership but it is worth stressing how different the abstract systems of mathematics are from conditions in the real world. Perhaps, in the future a kind of mathematics will arise which will be time and context dependent while still remaining more precise than ordinary speech. Mathematics does indeed model time dependent processes, notably via differential equations, but only from the outside; time itself, in the sense of change, is never allowed to be present within the boundaries of the mathematical system itself. Once true, always true—or so it would seem in mathematics.

Mathematics has managed to do something which sounds equally difficult, namely to model randomness (up to a point) and there is an interesting chapter discussing this complex issue in a recent book, *How Mathematicians Think*, by William Byers (Chapter 7). However, randomness is still, like time, studied from the outside although it is getting steadily closer and closer to the fixed, ideal world of mathematics via Heisenberg, chaos theory, Gödel's Incompleteness and so forth. Maybe the twin shadows of time and chance will in the end darken out the pure light of eternity after all.

Quaternionic space

Dennis Morris

The 4×4 matrix form of the quaternions is

$$\begin{bmatrix} a & b & c & d \\ -b & a & d & -c \\ -c & -d & a & b \\ -d & c & -b & a \end{bmatrix}.$$

The real part of this commutes with the non-real part. We thus have

$$\exp\left(\begin{bmatrix} a & b & c & d \\ -b & a & d & -c \\ -c & -d & a & b \\ -d & c & -b & a \end{bmatrix}\right)$$

$$= \exp\left(\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{bmatrix}\right) \exp\left(\begin{bmatrix} 0 & b & c & d \\ -b & 0 & d & -c \\ -c & -d & 0 & b \\ -d & c & -b & 0 \end{bmatrix}\right).$$

The successive powers of the 'imaginary' matrix are surprising; we have

$$I^2 = \begin{bmatrix} -b^2-c^2-d^2 & 0 & 0 & 0 \\ 0 & -b^2-c^2-d^2 & 0 & 0 \\ 0 & 0 & -b^2-c^2-d^2 & 0 \\ 0 & 0 & 0 & -b^2-c^2-d^2 \end{bmatrix},$$

$$I^3 = (-b^2-c^2-d^2)\begin{bmatrix} 0 & b & c & d \\ -b & 0 & d & -c \\ -c & -d & 0 & b \\ -d & c & -b & 0 \end{bmatrix},$$

$$I^4 = \begin{bmatrix} (-b^2-c^2-d^2)^2 & 0 & 0 & 0 \\ 0 & (-b^2-c^2-d^2)^2 & 0 & 0 \\ 0 & 0 & (-b^2-c^2-d^2)^2 & 0 \\ 0 & 0 & 0 & (-b^2-c^2-d^2)^2 \end{bmatrix},$$

leading to

$$\exp\left(\begin{bmatrix} a & b & c & d \\ -b & a & d & -c \\ -c & -d & a & b \\ -d & c & -b & a \end{bmatrix}\right)$$

$$= \begin{bmatrix} e^a & 0 & 0 & 0 \\ 0 & e^a & 0 & 0 \\ 0 & 0 & e^a & 0 \\ 0 & 0 & 0 & e^a \end{bmatrix} \begin{bmatrix} \cos\theta & \frac{b}{\theta}\sin\theta & \frac{c}{\theta}\sin\theta & \frac{d}{\theta}\sin\theta \\ -\frac{b}{\theta}\sin\theta & \cos\theta & \frac{d}{\theta}\sin\theta & -\frac{c}{\theta}\sin\theta \\ -\frac{c}{\theta}\sin\theta & -\frac{d}{\theta}\sin\theta & \cos\theta & \frac{b}{\theta}\sin\theta \\ -\frac{d}{\theta}\sin\theta & \frac{c}{\theta}\sin\theta & -\frac{b}{\theta}\sin\theta & \cos\theta \end{bmatrix},$$

where $\theta = \sqrt{b^2 + c^2 + d^2}$.

Normalizing this expression shows that the particular functions are projections onto the axes of the space. The rotation matrix is the exponential of a matrix with zero trace and thus has determinant unity.

Setting the polar form equal to a Cartesian form matrix and taking the determinants of both sides leads to the distance function:

$$r^4 = \left(a^2 + b^2 + c^2 + d^2\right)^2, \quad r = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

Since the functions in the rotation matrix are projections onto the axes, we are justified in calling them the trigonometric functions of quaternionic space. However, they do not have the simple differentiation cycles that we normally associate with trigonometric functions. We have

$$\frac{\partial(\cos\theta)}{\partial b} = -\frac{b}{\theta}\sin\theta,$$

and similarly for the other variables, but

$$\frac{\partial\left(\frac{b}{\theta}\sin\theta\right)}{\partial b} = \frac{b^2\theta\cos\theta + (c^2 + d^2)\sin\theta}{\theta^3}$$

and similarly for the other variables. Setting any two of the variables to zero restores the usual differential cycles.

The quaternions are a bona fide division algebra (though not an algebraic field). Since we have a distance function and a rotation matrix containing four (trigonometric) functions that are projections onto the axes, we have a geometric space. However, the failure of the trig functions to have a four-fold differentiation cycle shows that the axes of this space are not fixed together 'properly'. It seems that what we have is a kind of 'Siamese triplet'; we have three copies of 2-dimensional euclidean space joined together along the real axis. Such, it seems, is quaternionic space—many writers of science fiction make the most of it.

The sequence 1, 1, 3, 7, 19, 51, . . .

Patrick Walker

Pascal's triangle is well known.

1									
1	1								
1	2	1							
1	3	3	1						
1	4	6	4	1					
1	5	10	10	5	1				
1	6	15	20	15	6	1			
1	7	21	35	35	21	7	1		
1	8	28	56	70	56	28	8	1	
1	9	36	84	126	126	84	36	9	1

The middle binomial coefficients are

$$1, 2, 6, 20, 70, 252, 924, 3432, 12870, \ldots .$$

We can divide them by the natural numbers to give the sequence

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, \ldots .$$

The result of this division is always an integer since

$$\frac{1}{n+1}\binom{2n}{n} \;=\; \frac{1}{2n+1}\binom{2n+1}{n}.$$

These are the *Catalan numbers.*

A similar triangle is obtained by forming each number in a row as the sum of the one above plus the two on either side of it.

1										
1	1	1								
1	2	3	2	1						
1	3	6	7	6	3	1				
1	4	10	16	19	16	10	4	1		
1	5	15	30	45	51	45	30	15	5	1

The sequence of middle numbers is

$$1, 1, 3, 7, 19, 51, 141, 393, 1107, 3139, 8953, 25653, \ldots.$$

The two 'triangles' appear at first sight to have quite different patterns. But both contain a sloping line of triangular numbers,

$$1, 3, 6, 10, 15, 21, 28, \ldots.$$

The sum of any row of Pascal's triangle is a power of two. The sum of any row in the second triangle is a power of three. In Pascal's triangle, adding and subtracting in turn each element of a row gives a total of zero. In the second triangle, adding and subtracting in turn each element of a row gives a total of 1. And so on.

These are intriguing comparisons, and it might be thought that the sequence of middle numbers in the second triangle would possibly bear some relation to the Catalan numbers. One can prove by induction that the nth term is the coefficient of x^n in the expansion of $(1 + x + x^2)^n$, but I cannot find a closed formula. Can anyone help?

Problem 223.1 – Mud

Norman Graham

Mud flies off the hindmost point of a wheel rolling at a uniform speed. Will it hit the wheel again if it leaves the wheel in the same direction as the hindmost point (a) at the same speed, (b) slower, (c) faster? Of course, the vehicle is being driven in a perfect vacuum.

Problem 223.2 – Gun

Norman Graham

If a gun has a maximum range of r on a level plain, what is it from the top of a cliff of height h? Find a construction for the angle of elevation.

Problem 223.3 – Factorization

Tony Forbes

For which integer values of d does

$$x^4 - x - d$$

factorize?

How to solve quartics

Tony Forbes

Going through past issues of M500 I notice that regularly and often we have occasion to find the roots of fourth-degree polynomials. We always just quote the solution, referring to the literature or to computer software for the method of getting there. I find this unsatisfactory. And it is regrettable that quartics are no longer actively solved in schools. So I thought it would be a good idea if M500 were to fill a possible gap in your mathematical education by presenting a short exposition. Surely it can't be that difficult.

The problem is simple to state, 'Solve

$$x^4 + a\,x^3 + b\,x^2 + c\,x + d \;=\; 0 \tag{1}$$

in terms of the coefficients a, b, c and d.'

But that's far too difficult!

We can simplify the problem by getting rid of the x^3 term with an affine transformation. Make the substitution $x \to x - a/4$ and the second term of $(x - a/4)^4$ is $-ax^3$, which exactly cancels the ax^3 of the original expression (1). Assuming you want to see it written out in full, the equation becomes

$$x^4 + \left(b - \frac{3a^2}{8}\right) x^2 + \left(c + \frac{a^3}{8} - \frac{ab}{2}\right) x + d - \frac{3a^4}{256} + \frac{a^2 b}{16} - \frac{ac}{4} \;=\; 0$$

and, as you can see, the x^3 term really has vanished. However, we will continue to use the original coefficients, as in equation (1), but henceforth we shall assume that $a = 0$. I suppose it's worth pointing out that if $a = 0$ anyway, there is no need to perform this initial step.

We introduce a parameter u the value of which we shall decide at a time when it is convenient for us to do so. Remembering that $a = 0$, we rewrite (1) as

$$(x^2 + u)^2 \;=\; (2u - b)\,x^2 - c\,x + u^2 - d. \tag{2}$$

Now for the clever part. We want to make the right-hand side of (2) a square, just like the left-hand side. That means the discriminant Δ, say, of the quadratic

$$(2u - b)\,x^2 - c\,x + u^2 - d \tag{3}$$

must be zero. But

$$\Delta \;=\; c^2 - 4(2u - b)(u^2 - d) \;=\; -8u^3 + 4bu^2 + 8du + c^2 - 4bd. \tag{4}$$

Now is a good time to decide the value of that parameter. We choose u to be anything that makes $\Delta = 0$. This involves solving the cubic on the right of (4); but I have already explained how to do that—see 'How to solve cubics' in M500 **202**. Given that u satisfies $\Delta = 0$, we know the quadratic (3) must factorize into a square. Hence (2) becomes

$$(x^2+u)^2 \;=\; (2u-b)\left(x - \frac{c}{2(2u-b)}\right)^2. \tag{5}$$

You can verify that on multiplying (5) out you should recover the original equation (1) (with $a = 0$) when you substitute zero for $8u^3 - 4bu^2 - 8du + 4bd - c^2$.

Taking square roots of both sides of (5), we obtain

$$x^2+u \;=\; \pm\sqrt{(2u-b)}\left(x - \frac{c}{2(2u-b)}\right), \tag{6}$$

a pair of quadratics, which can be solved using a neat method sent to me by **Martin Hansen**; see below.

The problem is solved. However, one small detail is bothering me. As a quartic, (1) is supposed to have four solutions. But the cubic on the right of (4) has three roots, and together with the plus-or-minus choice of sign in (6) and the fact that a quadratic has two roots, there are twelve solutions altogether. We would be interested if someone can explain how to dispose of eight unwanted solutions.

How to solve quadratics

Martin Hansen

To solve $ax^2 + bx + c = 0$, multiply through by $4a$,

$$[4a^2x^2 + 4abx] + 4ac \;=\; 0,$$

complete the square within the square brackets,

$$[(2ax+b)^2 - b^2] + 4ac \;=\; 0,$$
$$(2ax+b)^2 \;=\; b^2 - 4ac,$$

square-root both sides,

$$2ax + b \;=\; \pm\sqrt{b^2-4ac},$$

and the well-known formula is there for the taking.

Letters to the Editor

Problem 220.4 – Biseptic

Tony,

Relying on intuition and putting $x = \pi$ into my old and trusted pocket calculator, working at ten sig. figs, I get $x^{14} + 508x^4 = 9171655.016$. Wow!

Solution time = 27 seconds which, for recent M500 problems, must be record (for me at least).

Jim James

Cylinders

Re: M500 **218** page 13, 'Mutually touching cylinders.'

It occurred to me that six cylinders could be arranged thus. Take three of the cylinders (I used pencils) and bind them with an elastic band, near to the ends. Hold the bunch loosely by this binding and force the ends apart, twisting clockwise as you do. You now have a crude tripod. Repeat the process with the three other cylinders, this time twisting anticlockwise. Nest the second tripod into the first and bind the ends together in three pairs (one from each 'tripod' forming a pair). I tried this but I couldn't decide if all six were in mutual contact (possibly because the elastic bands got in the way!).

However, I recall seeing a solution with seven cylinders in a book many years ago (it could have been from the *Scientific American* series 'Mathematical Puzzles and Diversions' by Martin Gardner, but at a quick glance it didn't turn up in my collection of this and similar publications). The arrangement is in the diagram. The pencils meet at infinity if necessary, since the cylinders are allowed to be of infinite length. Looking at this diagram it does seem that the arrangement is valid. If so then given a set of cylinders of unit radius, what is the minimum length required? How would you prove that all seven are in mutual contact?

Ken Greatrix

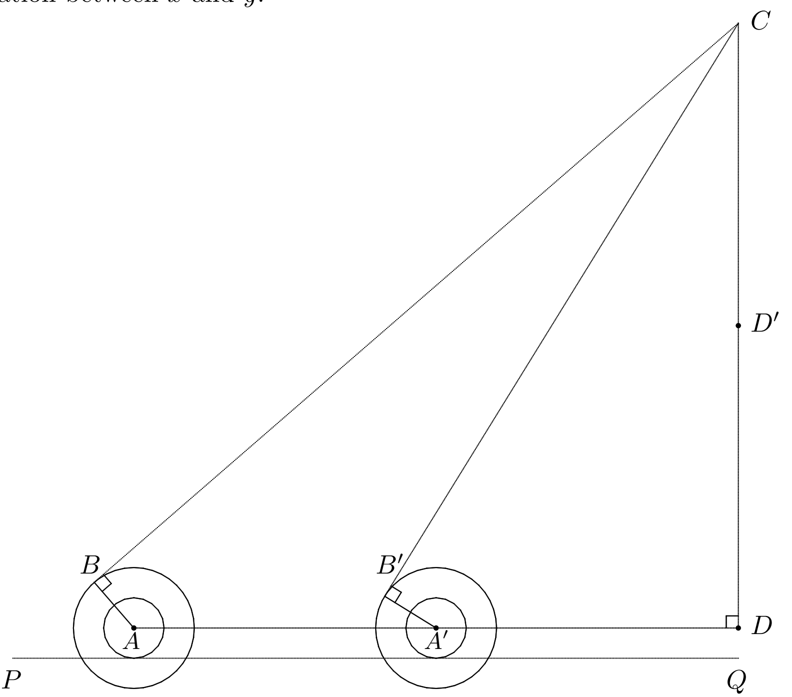
TF writes. To amuse myself I had a go at making the thing out of pencils and rubber bands. The resulting photograph I think illustrates the situation far better than any diagram I can reproduce. Moreover I was surprised to see how stable this configuration actually is. And I can even convince myself that each cylinder touches the other six—especially if you count points at infinity.


The ultimate test
The ultimate test
logic!
SUDOKU
CHALLENGER
ultimate test of logi
The ultimate tes

Solution 210.5 – A monkey and a pole

> A monkey climbs a pole along DC to a height y. It is tethered by a line to a drum of radius R which rolls on its axle of radius r along a flat bed PQ. During this rolling, the drum moves horizontally by a distance x. You may assume that the pole isn't high enough for the drum to collide with it, that the drum rolls back to its start position when the monkey climbs down again and that the line is always tangential to the drum. What is the relationship between x and y?

ADF — On reading this after an interval of two years it has become evident that we have been a little sloppy with the statement of the problem. One interpretation is implied by the solution offered below, in which the monkey is tethered by a fixed piece of rope, which goes initially from D to A to B to C and afterwards from D' to C via D, A' and B', but I have no idea of the mechanism that keeps the rope tangent to the drum. Perhaps we can treat the problem as an exercise in geometry. Given the parameters r, R, h, x, y, z, and that the polygonal lines $DABC$ and $D'A'B'C$ are equal, find a relation between x and y.



Steve Moon

In the diagram, $x = AA'$, the distance moved by the drum, $y = DD'$, the distance climbed by the monkey, $z = AD$ and $h = CD$.

The length of the rope at the start is $|CB| + R + z$ and at the end $|CB'| + R + z - x + y$. Hence $y = x + |CB| - |CB'|$. By Pythagoras, $|A'C|^2 = |CB'|^2 + R^2$ and $|A'C|^2 = h^2 + (z-x)^2$. Therefore $|CB'|^2 = h^2 + (z-x)^2 - R^2$. Also $|AC|^2 = |BC|^2 + R^2 = h^2 + z^2$; therefore $|CB|^2 = h^2 + z^2 - R^2$. Hence

$$y \;=\; x + |CB| - |CB'| \;=\; x + \sqrt{h^2 + z^2 - R^2} - \sqrt{h^2 + (z-x)^2 - R^2}.$$

ADF again — Well, I am unhappy because I notice that the parameter r has disappeared! And judging by the simplistic answer I am convinced that the intended problem has not been solved. As the real solution is likely to be of interest, can someone please enlighten us as to the true nature of the real problem.

Problem 223.4 – The arbelos

Norman Graham

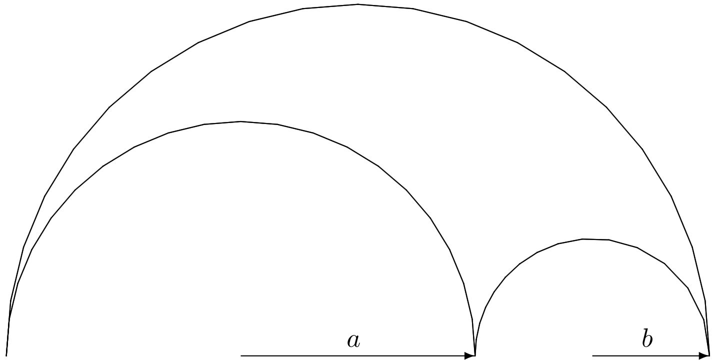
Also known as the 'shoemaker's knife' of Archimedes, this shape, the *arbelos*, is bounded by three touching circles of radii a, b and $a + b$, as in the upper diagram on the next page.

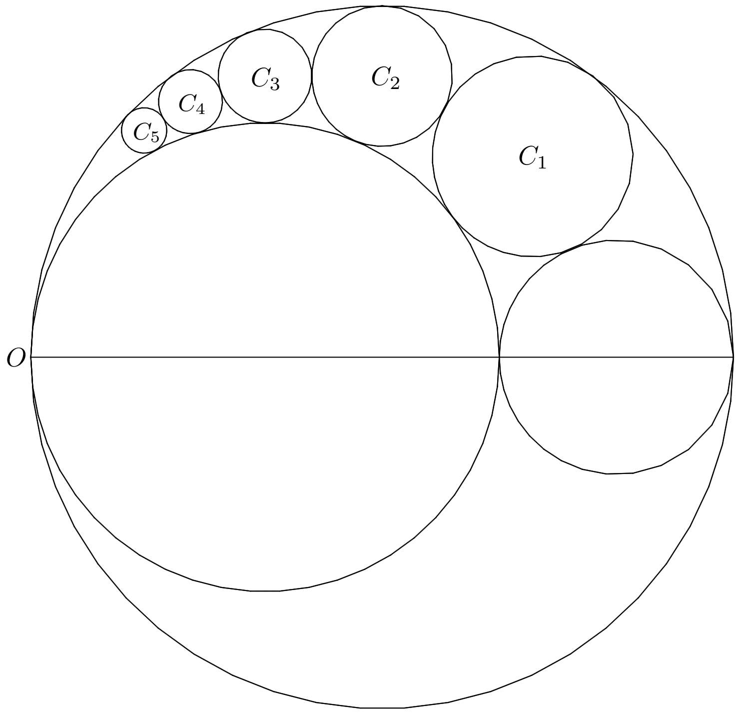
A 'train' of circles C_n, $n = 1, 2, \ldots$, is constructed inside the arbelos as shown in the lower diagram. Show that their centres lie on an ellipse, and find the diameter d_n of each circle in terms of a, b and n. Also show that for each C_n, the centre is at a distance nd_n from the centre-line.

Hint. There is a beautiful proof using inversion geometry. With O as the centre of inversion, any point X is inverted to X' on OX, where $|OX| \cdot |OX'| = k^2$, for some constant k. The two tangent circles are inverted to two parallel straight lines, so that the train becomes a set of equal-radii circles touching those lines.

Reference. Martin Gardner, *Fractal Music, Hypercards and More: Mathematical Recreations from* Scientific American *magazine*, Chapter 10.

[To get a better idea of the inversion process, you can look at the front cover of M500 **203**, where the train of shrinking circles and their tangent lines do really become a column of constant-radius circles sandwiched between two vertical lines. — **TF**]





What did the string theorist say when his wife caught him in bed with his mistress?

"It's all right, dear, I can explain everything!"

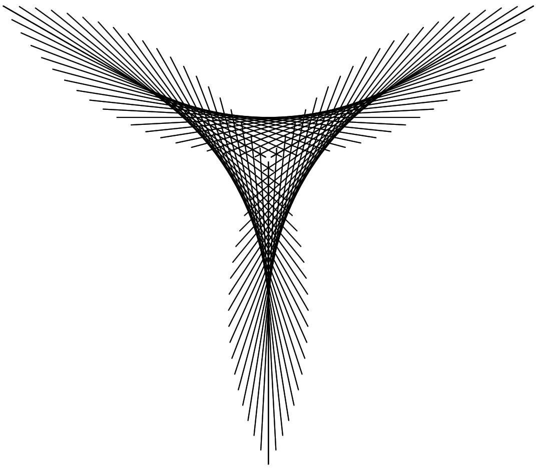
Problem 223.5 – Reversing a needle

Tony Forbes

A needle of length 2 is lying in the plane pointing straight up and with its centre at the origin. It is required to reverse its direction, so that it points straight down, again with its centre at the origin. The needle must remain in the plane throughout the reversing manoeuvre; so picking it up, turning it around and putting it back would be cheating.

Question. *What is the smallest possible area on the plane swept out by the needle during the process?*

Think of the needle as being very dirty, so that it leaves marks on the all parts of the plane it comes into contact with. The marked parts then identify the area swept by the needle. If you rotate it through 180 degrees about its centre, the area is π. But perhaps one can do better by imagining that the needle is a car doing an n-point turn along arcs of radius r for some odd integer n and some positive number r. Here, for instance, is what happens when $n = 3$ and $r = 1.77$.



Thanks to **Emil Vaughan** for showing me something like this problem. As a warm-up, see if you can find the exact value of r for which the hole in the centre of the diagram just ceases to exist. Do it for any odd n.

If you disregard the very simplest cases, there is in all of mathematics not a single infinite series whose sum has been rigorously determined. In other words, the most important parts of mathematics stand without a foundation.—**Niels Abel**

Contents — M500 223 – August 2008