*

ISSN 1350-8539

M500 217

○	♠	♠	♠	♡	♡	♠	♡	♡
♠	♡	♡	♣	♠	♣	♣	♢	♣
♠	♡	♡	♣	♣	♢	♣	♣	♠
♠	♢	♡	♢	♢	♠	♢	♣	♡
♢	♠	♣	♡	♡	♣	♣	♣	♠
♢	♢	♡	♠	♢	♢	♣	♠	♡
♠	♡	♢	♢	♡	♣	♢	♠	♢
♢	♡	♢	♣	♡	♠	♠	♢	♢
♢	♣	♠	♣	♠	♣	♡	♣	♡

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 on a PC diskette or via e-mail.

The Twins Paradox and related issues

Sebastian Hayes

MOST CURRENT books take the line that the Twins Paradox is no paradox at all and is simply a consequence of Special Relativity, a theory that is now perfectly respectable and amply supported by experiment. Is this really so?

Suppose twin A leaves the Earth and sets off in a space-ship travelling at three-fifths of the speed of light, $\frac{3}{5}c$, while twin B stays at home. It is to be understood—a point not sufficiently stressed in textbooks—that the Earth and the destination star are absolutely at rest relative to one another and thus share the same 'inertial reference frame'. (In practice this would never occur, of course.) Moreover, we assume that twin A, prior to his departure, was earthbound and so, initially, shared the same reference frame with his twin and the hypothetical 'stationary' star.

The star is, say, six light years away. Note that this is a distance, not a time interval, and corresponds to (roughly) $6 \cdot (9.46 \cdot 10^{15})$ metres since light in a vacuum travels at $3 \cdot 10^8$ metres/second or $9.46 \cdot 10^{15}$ metres/year.

If twin A were actually travelling at the speed of light he would reach the star six 'years' later. These are Earth years, which would be meaningless to the traveller, of course: he will have his own way of measuring the lapse of time. Now the rocket is in uniform motion with respect to the distant star or, shifting to the rocket inertial frame, the star is in uniform straight line motion with respect to the rocket. If Special Relativity is correct, we need to make a distance 'correction' because we are dealing with systems in relative motion. To obtain the distance of the star for the traveller, i.e. relative to his frame, we must multiply by the factor $\sqrt{1 - v^2/c^2}$, where $0 \le v < c$. This factor is less than 1 and so we have the celebrated 'Fitzgerald contraction', which applies to all lengths in the other inertial frame and thus to distances also. The 'other' frame is, in this case, that of the rest of the universe, including the Earth, all movements of rotation and expansion being neglected.

Where does this strange factor come from? Fitzgerald originally fished it out of a hat to explain the null result of the Michelson–Morley experiment but Einstein showed that this factor followed from the two premises of Special Relativity, firstly that the 'laws of physics take the same form in all inertial frames' and secondly that 'the speed of light in a vacuum is constant'.

Here $v = \frac{3}{5}c$ and so the factor, which is dimensionless, is $\sqrt{4^2/5^2} = 4/5$. The traveller is supposed to have a clock on board—in a moment we will specify what sort of a clock it is—and, on arrival, once we have converted into Earth year values, it will read as

$$\frac{\text{distance}}{\text{speed}} = \frac{\frac{4}{5}\cdot 6\cdot(9.46\cdot 10^{15})\ \text{metres}}{\frac{3}{5}\cdot(9.46\cdot 10^{15})\ \text{metres/year}} = 8\ \text{years.}$$

Now, even if the voyager twin sends back a radio signal at once to his twin, it will not arrive for another six years. Rather than work out retrospectively what the time was on Earth at the moment the voyaging twin reached the star, it is much more sensible to suppose, as Einstein himself did in his 1905 paper, that a clock C was already installed on the star and that this clock, a very regular pulsing device of some sort, had somehow previously been synchronized with the Earthbound twin's clock. Since the star and the Earth are in the same inertial frame, this starbound clock will, when duly interpreted, read exactly the same as the Earthbound clock does at the same moment (since they share the same inertial frame). Because there is no length contraction involved for a system at rest, we simply divide 'ordinary' distance by the speed of the rocket, i.e.

$$\frac{6\cdot(9.46\cdot 10^{15})\ \text{metres}}{\frac{3}{5}\cdot(9.46\cdot 10^{15})\ \text{metres/year}} = 10\ \text{years.}$$

Now this reading is not the same as the reading on the rocketbound clock. It is quite essential to grasp that this difference is a perfectly objective bona fide physical difference: it is not a question of one twin 'feeling' the passage of time differently.

We may imagine a batch of N_0 radioactive atoms held by twin B on Earth, by twin A on the rocket and already placed on the star. Then, according to the formula for radioactive decay, when twin A arrives at the star

$$N_{\text{rocket}} = N_0\, e^{-8/T} \quad \text{and} \quad N_{\text{Earth}} = N_0\, e^{-10/T},$$

where T is the mean life of the radioactive atoms (in years) measured in a co-ordinate system where the atoms are at rest.[1]

Alternatively, let us suppose that both the twins have very regular blood pressure, that their hearts beat exactly once a second and that each heartbeat of each twin sets off a mini-explosion by way of some electrical device connected up to their bodies. Each mini-explosion leaves a permanent burn-mark on a revolving strip of metal so that the number of burn marks indicates exactly the number of times each twin's heart has beaten. Both the twins agree that the number of burn marks defines what is meant by comparative, and both twins are satisfied that the 'explosion clock' already set up on the star has been exactly synchronized with the rhythm of their heart beats when they were still on Earth. According to Einstein, if we start all three explosion clocks at the same moment, at the end of the two-way trip, the number of burn marks on the strip of metal back on Earth will be the same as that on the strip of metal on the star, and this number is greater than the number of burn marks on the rocketbound strip of metal.

Twin A will, if he alights on the star, notice a discrepancy which really exists. What to conclude? Well, he might decide there was something wrong with one or both clocks, but if he checked his own clock with the one on the star while staying there for a while, he would (according to Einstein) find that they both ran at exactly the same rate.

We discount the time spent on the star (which is the same for all three clocks anyway) and, for the moment, we discount the time spent putting the rocket into reverse motion at the star end of the trajectory. Twin A travels back at exactly the same constant velocity as he came and so his clock will have the same reading for the return trip as it did for the outward bound trip, i.e. will read 8 years. Similarly, twin B's clock, once we have 'backdated' it to restart when the twin leaves the star, must read the same as the clock which was on the star did when twin A arrived, i.e. will read 10 years. Twin A has thus aged by 16 'years' and twin B by 20 years. Now, although a year on the rocket will not be any sort of a naturally perceived interval for the voyager, the number of radioactive atoms remaining, or the number of burn marks on the strip of metal, is there as proof of the amount of 'time' that has slipped by.

We must, of course, completely rule out all qualitative and biological senses to the word 'life'. It may well be that the stress and strain of space-flight will have 'prematurely aged' the traveller—indeed this is almost certain. Secondly, the earthbound twin may have been having such a fantastic life that he doesn't mind looking a bit older. But all this is quite irrelevant since we are merely concerned with the passage of time as measured by *a finite number of punctual regularly repeating events*—an 'objective' phenomenon if ever there was one.

Now, if the situation of the two twins was symmetrical—as Professor Dingle used to claim—then the discrepancy, if observed, would contradict the very Principle of Relativity that made Einstein famous. However, it is not symmetrical since twin B remains in the same inertial frame during the entire process, while twin A starts off in an Earthbound frame, switches to a rocket inertial frame once the desired constant velocity has been attained, goes into reverse when he reaches the star, shifts back into the previous rocket inertial frame and finally ends up stationary in the Earth frame. Twin A has thus been subject to accelerated motion (1) at the beginning of the voyage; (2) at the halfway point and (3) at the end of the voyage.

Even if we have twin A coasting by the Earth and so being in constant motion from the very beginning and have his 'clock' start at the 'very instant' he passes by a space station fixed to the Earth, and stop his clock when he passes by the same spot on the return trip, there still remains the period of deceleration and re-acceleration at the star end on the trajectory.

Now, there is currently much experimental evidence that the effects predicted by Einstein's Special Theory of Relativity really do occur but the vast majority of such tests deal with situations that actually are *symmetrical*—and so should not be cited as direct confirmation of the Twins Paradox (or perhaps better, the Twin Clock Paradox). The most dramatic evidence for time dilation is the mean lifetime of μ-mesons which are produced by cosmic radiation near the top of the Earth's atmosphere as compared with their mean lifetime in laboratory conditions. In the latter typically they live for about $2.2 \cdot 10^{-6}$ seconds; so that even if they travelled at the speed of light they would only cover about 660 metres before decaying. It is found that a fair percentage of such mesons actually reach sea level which, by Earth frame reckoning, is some 20 kilometres or more away. This is perfectly in accord with Special Relativity since the great speed of the mesons relative to the Earth gives them an enormous time dilation factor (over 30). However, the situation is perfectly reversible and so there would be no advantage involved if the μ-meson were a microscopic rocket with a microscopic human being on board. He or she would not 'live any longer' through being in such a state of motion. He would simply judge the height of the Earth's atmosphere to be rather less than how we see it.

We note the interval between two ticks (or any other regular punctual events) as $\delta\tau = \delta t\sqrt{1 - u^2/c^2}$, where u is the velocity of the system relative to a given inertial frame and δt is the 'proper' time interval in the stationary frame. If the frame in which we measure $\delta\tau$ is at rest with respect to the first frame, $\delta\tau = \delta t$ since $u = 0$ throughout. If we start at $T = 0$ and evaluate the 'time' that has elapsed according to the number of ticks, or other strictly punctual events, $\delta\tau$ is a whole number, say T_E. Whether the inertial frame chosen as the 'standard' one is a rocket or an Earthbound laboratory has no bearing on what happens within the frame, provided the chosen frame remains unaccelerated. To make this more precise, suppose the μ-meson were a rocket travelling at uniform speed and bearing a diminutive twin. The heart of the human voyager would beat a certain number of times, say N times, during the equivalent of the life span of the meson and it would beat exactly N times in an Earthbound laboratory or any other frame provided this frame really were inertial. If the agreed way to measure lapse of time and comparative aging is the number of heartbeats, then that is that and no kind of switching of frames is going to change the situation for the person who remains within his own frame. The discrepancy only comes in when we have a double perspective going and make measurements within one's own system and measurements of a different system but from the standpoint of the first system.

Thus if we evaluate time intervals from the outside, for example time intervals within the rocket from the standpoint of the Earthbound system, or, conversely, time intervals on Earth from the standpoint of the rocket, $u \neq 0$ and $\delta\tau \neq \delta t$. If the two systems are moving with uniform speed relative to each other, $u = v =$ constant and $\delta\tau = \delta t\sqrt{1 - v^2/c^2}$. Then, supposing we have some means of synchronizing the two systems at the start—for example by a beam of light which is emitted midway between, say, a floating laboratory fixed to the Earth and the rocket as it passes the Earth—and some acceptable way of synchronizing the concluding events also, then $T_R = T_E\sqrt{1 - v^2/c^2}$ where T_R is the number of ticks in the rocket and T_E the number of ticks in the Earthbound system. We take the former to the nearest whole number. Since $\sqrt{1 - v^2/c^2} < 1$ for $v < c$, $T_R < T_E$.

Finally, what if u is variable? Einstein discusses this very point in his 1905 paper and suggests that a closed arbitrary curve be treated as a polygon and that we consider the accelerated frame, or frames, as a succession of inertial frames in which the physical system is 'instantaneously' at rest (or in constant straight line motion). So $\delta\tau$, the time interval elapsed in the 'moving' frame as judged by an observer on the other frame, becomes an integral $\delta\tau = \int \sqrt{1 - u^2/c^2}dt$ with limits of integration in terms of u being $u = 0$ and $u = v$ (if the rocket starts from rest).

Since $0 \leq u \leq v$ where v is the final velocity which is to remain constant during the outward and homeward voyages, it follows that the above integral $\delta\tau < \delta t$, where the latter is the time interval evaluated in the stationary frame. So the 'acceleration period' of the voyage, whether long or short, does not affect the previous result qualitatively: Einstein in his 1905 paper predicted that any clock following the path of a closed curve and returning to its starting point will always be found to have run slower than an identical clock which remained at that point, or one nearby.

From this we can conclude that, no matter how the rocket system accelerates and decelerates, the total time $T_{\text{rocket}} < T_{\text{Earth}}$ and in terms of heart beats or other punctual events this will be a strict numerical inequality $N_{\text{rocket}} < N_{\text{Earth}}$. Thus the voyager twin 'ages' less.

However, this is only a first look at the subject. Strictly speaking, acceleration is not covered by the Special Theory of Relativity. According to the famous 'Principle of Equivalence' which is the cornerstone of General Relativity, the effects of acceleration on the components of a physical system are the same as the effects on such a system of an 'equivalent' homogeneous gravitational field.

Suppose ν_0 to be the frequency of light emitted in a stationary reference frame $\mathcal{A}$. We now have a second reference frame $\mathcal{A}^*$ which is moving away from $\mathcal{A}$ at uniform velocity v. If we apply the Lorentz transformations we find that the frequency of the light actually received at $\mathcal{A}$ from the moving reference frame is given by

$$\nu_1 \;=\; \nu_0 \frac{\sqrt{1-v^2/c^2}}{1+v/c} \;=\; \nu_0 \sqrt{\frac{1-v/c}{1+v/c}} \;=\; \nu_0 \sqrt{\frac{c-v}{c+v}} \;\approx\; \nu_0(1-v/c)$$

to first order. This is the formula for the Doppler effect when an object emitting light is moving at speed v relative to an inertial frame. Since $1-v/c<1$ there should be a reduction in the frequency of the light received.

Suppose a rocket of length h accelerating relative to an inertial frame $\mathcal{F}$ and bearing two clocks, one at the back of the rocket, clock 1, and the other at the front, clock 2, is is 'instantaneously at rest' at time 0. Now, if the 'clocks' are light sources, the time, t, for the light from clock 1 (measured in $\mathcal{F}$) to reach clock 2 is given by

$$ct \;=\; h + \tfrac{1}{2}at^2,$$

approximately $t = h/c$ if $\frac{1}{2}at$ is very much smaller than c. At this time, t, clock 2 will be moving instantaneously at a velocity $v = at$ and therefore clock 1 will appear to be receding from clock 2 at velocity $v \approx ah/c$. Hence the light at frequency ν_0 from clock 1 will appear shifted to frequency ν_1 given by

$$\nu_1 \;\approx\; \nu_0\left(1-\frac{v}{c}\right) \;\approx\; \nu_0\left(1-\frac{ah}{c^2}\right).$$

The observer alongside clock 2 will thus observe the light shifted towards the red by a difference in frequency of $\nu_0 ah/c^2$ approximately. Conversely, the formula for the light from clock 2 as it reaches clock 1 will be

$$ct \;=\; h - \tfrac{1}{2}at^2$$

and an observer alongside clock 1 will consider that the light frequency is being shifted towards the violet by the same difference as before.

This is nothing new. But according to the Principle of Equivalence, all such effects would also be observed if the rocket were at rest but subject to a gravitational field whose strength was exactly equal to the strength of the force producing the acceleration (but opposite in sign). If the field is that of the Earth and the rocket is pointing upwards ready to be launched, we replace ah by gh on the assumption that the Earth's field is homogeneous over the distance h, or for a varying gravitational field by the difference in gravitational potential $\Delta\phi$. Here h is, or rather was, the height of the rocket but it can be any distance between two points of a rigid body fixed to the Earth and pointing skywards. We thus derive an expression for the difference in frequency of two clocks,

$$\frac{\Delta\nu}{\nu_0} \approx \frac{\Delta\phi}{c^2},$$

where ν_0 indicates the frequency of light emitted at the surface of the Earth (i.e. at the back of the stationary rocket). For two points situated at distance r_0, the radius of the Earth (considered to be constant), and $r_1 > r_0$, $\Delta\phi = \int GM/r^2 dr$ between limits of integration r_0, and r_1 giving us

$$\frac{\Delta\nu}{\nu_0} \approx \frac{1}{c^2}\int_{r_0}^{r_1} \frac{GM}{r^2} dr = -\frac{GM}{c^2}\left(\frac{1}{r_1} - \frac{1}{r_0}\right) \approx \frac{g\,r_0}{c^2}\left(1 - \frac{r_0}{r_1}\right).$$

Since $r_1 > r_0 > 0$ this discrepancy is greater than 0, which means that clock 2, situated at distance r_1, should register more ticks than clock 1. So there is 'more action' the further out you are from the centre of the Earth—and consequently more heart beats and more aging. This positive difference has been measured using very sensitive apparatus situated on different floors of a New York skyscraper, thus confirming the predictions of General relativity (on this issue at least).

Note, however, that in this example the two points are at rest relative to the Earth and so there is no reason to invoke all the effects of Special Relativity such as length contraction and time dilation. In the case of a satellite orbiting the Earth we have not one but *two* relativistic effects and they are opposite in sign: the satellite is in accelerated motion relative to a clock at rest on the surface of the Earth, and so we would expect the satellite clock to run slow, but the satellite clock is also at a point of higher gravitational potential (because work had to be done against the pull of the Earth to put the satellite into orbit) and so, because of the Principle of Equivalence, we would expect the satellite clock to be speeded up. The reader might like to work out the height above sea level when the astronaut starts aging more than his twin who remains at the surface. (Answer given at end of this article.)

What bearing does this have on the original Twin Paradox? The crucial point is what happens during the period of deceleration/acceleration when the rocket turns round—or, better, reverses direction—at the close of its outward journey. According to the Principle of Covariance, there is no specially privileged coordinate system at all and so we can legitimately consider the rocket to be at rest all the time in a succession of 'inertial frames'. From the point of view of the rocket system, it is the Earth, originally moving with a positive constant velocity v with respect to the rocket, which decelerates to zero, then goes into reverse and attains v once more, this time moving in the direction of the (stationary) rocket, if the total time taken for the Earth to reverse is t_2 measured in the Earth frame, the Earth may be considered to have had a constant acceleration of $a = \pm v/(t_2/2) = \pm 2v/t_2$ during this period. By the Principle of Equivalence, we consider the Earth to have been in vertical motion within a uniform gravitational field of strength $2v/t_2$ centred on the base of the rocket. According to this view, the Earth is, during this period, in a region of higher gravitational potential than the rocket and in consequence there is a good deal more going on back on Earth, more ticks, more mini-explosions, more heartbeats. Thus, according to this definition of aging, the Earthbound twin still ages more than the voyager twin during the entire trip, no matter how long or short the acceleration period is, though not by quite the same amount as we obtain using the principles of Special Relativity alone.

We are actually not out of the lion's den yet by a long chalk. The Earth rotates about its axis (or, if you wish, the 'fixed stars' rotate around it) and follows an elliptical course relative to the sun. Also, the course of any actual rocket would not be a straight line. The rates of clocks in rotating reference frames are, of course, dealt with in General Relativity but we soon get into heavy water conceptually. In Newtonian mechanics, if we insist on treating the Earth as fixed, we have to introduce centrifugal and Coriolis forces to explain why a falling body deviates markedly from the vertical. The student at this level is taught to consider these forces, like all Newtonian 'inertial forces', as nothing but mathematical fictions—the implication being, of course, that 'in reality' the Earth reference frame is not fixed but is rotating on its axis. Already, Bishop Berkeley, a remarkably perceptive critic of Newton, questioned the legitimacy of assuming that rotatory motion was 'absolute' rather than relative to the fixed stars (as they were then conceived to be). In General Relativity there is no privileged coordinate system and so it is just as legitimate to consider the Earth to be fixed as rotating on its axis. This means that the centrifugal and Coriolis forces so-called are just as 'real' as any other forces and are ascribed to the rotation of distant celestial masses. To correctly describe the behaviour of a single 'clock' it becomes necessary (in principle) to take into account the rest of the matter in the universe and the 'gravitational fields' due to distant masses may well turn out to be different even for systems moving in uniform straight line motion relative to each other! Non-locality is built into General Relativity from the beginning but it is not the same kind of non-locality that we get in Quantum Mechanics when, for example, paired photons drifting away from each other remain correlated with respect to their spin. The latter is a matter of specific particles being correlated to certain other specific particles, but in General Relativity everything really is 'related to everything else' (at least from the gravitational point of view) since the universe is 'all of a piece'. Interestingly, this is a return to the ancient Roman Stoic conception whereby the cosmos is viewed as a single organic entity: 'Ever consider and think upon the world as being but one living substance, and ... how all things that are concur in the cause of one another's being' as the Emperor Marcus Aurelius put it in his *Meditations*.

It is thus something of an exaggeration to say that the Twins Paradox has long since been laid to rest. If the difference between a so-called 'inertial' and an 'accelerated' system is only a question of degree—strictly speaking inertial systems do not exist at all—one might wonder whether after all we should consider the two trajectories, that of the rocket and the Earth, as being in some sense symmetrical. Certainly, if we have two rockets following an identical circular (or elliptical) course with the same angular velocity but moving in opposite directions, their clocks, initially set at zero, should give the same readings each time they cross—otherwise this would be violating one of the key assumptions of Relativity, the homogeneity of empty space. Sciama has in fact argued that, if we neglect the influence of distant celestial masses, Professor Dingle's view would be perfectly correct!

One sometimes wonders what there is to hold on to in the universe if 'everything is relative': popular books try to reassure the public by telling them that c, the speed of light in a vacuum, is 'absolute' but unfortunately this is not the case in General Relativity. This does not bother me personally since I have always considered c to be a limiting speed, not as a speed actually attained by any particle. To attain c exactly a particle would have to be massless, which means that it would have absolutely *no* resistance to any attempt to change its state of rest or relative straight motion, in which case I do not see how it could be anything at all even for a single instant. The neutrino, long assumed to be massless, is now thought to have a small mass, and the same could be true of the photon.

There are, for all that, certain 'things' which are not relative in the Theory of Relativity, Special or General, namely events by which we must understand 'strictly punctual occurrences which leave traces and thus have observable effects'. The idea of an 'event' such as a flash of light or an explosion is indispensable in Special Relativity and was, for Einstein at least, perfectly unproblematical.[2] Now, the occurrence, or not, of an event is in no sense a subjective matter nor, as far as I can see, does it have anything to do with one's state of relative motion. If an explosion occurs alongside you in your hotel room inertial frame, it occurs (or 'will occur') in all inertial and non-inertial, frames everywhere—to avoid the use of tenses, we can simply say that it 'has occurrence' in all frames. Moreover, if two nearby distinct events are noted by an observer on the spot to have a definite order, this is their true order which they cannot lose. The alleged arbitrary order of 'space-like' separated events is largely an academic issue because such events cannot possibly have any consequences for us. I have read that for one person on Earth an invading fleet may have already set out from Andromeda while for someone in the next town, or even the next street, the decision to invade the Earth may not even have been taken. But so what? When the fleet actually gets here the relevant preceding events on Andromeda will have occurred, and will have occurred in a precise order which is the same for everyone.

What we undoubtedly do lose in Special and General Relativity is the notion of a universal 'now', but I cannot decide whether this is a serious matter or not; if the Big Bang theory is right, presumably there was a period when a 'universal now' did exist and so, in principle at least, as Eddington suggested, it should be possible to define a kind of weighted time slice through the universe as it evolves, in much the same way as we determine the centre of mass of a physical system. Any such chosen 'now' would, of course, be purely theoretical but then this is true of the centre of mass anyway since no such point actually exists in the real world.

Despite the high abstraction of modern science most of us do have some sort of a basic physical picture at the back of our minds, usually quite a crude one: Newton had his apples and billiard-balls and even modern pure mathematicians would be hard pushed if they had to do without images of the 'number line which is everywhere dense' and similar idealizations. I find that I have for most of my life been carrying around in my head a (very) basic schema which at rock bottom consists of only three items: events, a Locality where events can and do occur, and some sort of causal relationship which controls the occurrence (or not) of specific events. These items are certainly amongst the half dozen or so main players in the great game of Relativity and I have sometimes wondered whether it would be possible to describe the world using these concepts alone, to the exclusion of 'object' concepts such as particles and massive bodies—in terms of my primary notions 'objects' are simply relatively repeating event-patterns and their 'mass' is the resistance of a particular pattern to a change of rhythm.

To get going one needs what I call the 'Principle of Occurrence', which in terms appropriate to the present discussion would go something like this:

If an event has occurrence for one observer, it has occurrence *for all possible observers.*

Is this saying anything more than the sort of worldly wisdom dispensed by a typical American celebrity on a chat show, 'if something happens, my God it sure happens!'? I think it probably is: I am more worried about the Principle letting in more than I bargained for than I am about it being tautological. For it could be that by plotting a zigzag course from one possible observer to another, one would end up by allowing in certain 'future' events and once one starts doing this it is not clear where to draw the line. Einstein, towards the end of his life, apparently believed that the 'universe' existed (or exists, rather) in an eternal now: it is not clear whether his motivation was temperamental or mathematical.

Of course, an observable, thus macroscopic, event on closer examination turns out to consist of a number of smaller events and before we know what's happening we find ourselves transported to the wonderland of continuous functions and transfinite set theory. So everything turns out to be infinite since, as we all know, there are just as many points (= events) in a line segment (= interval) two inches long as in one that stretches from here to Mars.

Can such 'infinite regress' be avoided? Yes, very easily, by introducing the idea of an 'ultimate event' which is, by definition, an event which cannot be further decomposed. It readily follows from this that any connected sequence of observable events is made up of a finite number of ultimate events—though it may well be that at the present stage of our technology this number is unknowable. Again, it seems reasonable to suppose that there is a limiting value to the number of events that can possibly occur, or can have occurred, between two arbitrary specific ultimate events, i.e. there is an upper limit to what can occur within an arbitrary region of the Locality.

Such suggestions are perhaps not quite so offbeat today as they were even five years ago. Most nineteenth century physicists thought that matter was continuous. They were completely wrong about this and Einstein was one of the people who showed that they were wrong (not by Relativity as such but by his work on the photo-electric effect and similar phenomena). At present the vast majority of physicists still for some reason insist on seeing space and time as 'infinitely divisible' but this is not the picture emerging from the theory of 'loop quantum gravity', one of the two or three serious rivals to string theory in theoretical physics. According to one of its main spokesmen, Lee Smolin, loop quantum gravity predicts that space comes in discrete lumps about the size of $10^{-99}\,\mathrm{cm}^3$ and time in point-instants of about 10^{-43} seconds.[3] The theory does not speak of 'ultimate events' as such but it does seem to mean that whatever exists at all must either exist momentarily or as a *discontinuous* chain of discrete spacetime configurations. In my terms, I would guess (or postulate) that each of these Plancktime globules can receive at most only one ultimate event. I had, as a first approximation, conceived of the Locality as being a uniform four-dimensional grid: in loop quantum gravity the building blocks of spacetime are naturally more complicated in structure (more like irregular polyhedra) but the basic point is that there is a given (whole) number of blocks to any bounded region of spacetime. If, then, we attribute at most one ultimate event to one spacetime globule, this amounts to my principle that there must be a maximum number of events that can occur in any bounded region of the Locality.

As I envisage things, ultimate events have uniform and negligible (but not zero) extent; they are not elastic and not subject to such things as Lorentz contractions and time dilations. The legitimate divergence of opinion about the 'length' of spatial and temporal intervals according to states of motion (which is today an unquestionable experimental fact) must, then, apply only to the gaps between the events. These gaps do not have a fixed length because they do not have a proper length at all. Viewed like this some of the 'paradoxes' of Relativity are perhaps more acceptable. Imagine two successive screams of someone falling into a Black Hole. It has been said that for someone on the outskirts of the Black Hole the interval between the screams would be 'infinitely long' (an exaggeration, of course, for if this were so the observer would not hear the second scream) whereas for the person falling into the Black Hole the interval would be of perfectly normal length. Now reduce these screams to two ultimate events: both 'observers' would agree that the screams took place and took place in a given order. They would only disagree on the duration of the interval which is, in a sense, neither here nor there.

I hasten to add that this distinction between 'events' and the 'gap between events' is no part of the theory of loop quantum gravity. Lee Smolin does discuss the question of what there is between what he calls 'ticks' (the nearest he gets to the idea of an ultimate event) and writes, 'Time does not exist in between the ticks; there is no in between', in the same way that there is no water in between two adjacent molecules of water' (Smolin, *op. cit.*). I would say that there is something between two molecules of water, namely a gap, for if there was not a gap the two molecules would not be two but one. Similarly, the gap between 'ticks' is, if you like, 'nothing' but this only means that it is not measurable: it is that portion of the Locality where no events have occurrence. If there is to be a competition between the Locality and events, it is the former and not the latter that must be considered to be the more fundamental: for I can imagine a Locality entirely empty of events (in much the same way as the de Sitter universe is empty of matter) but I cannot imagine events occurring without there being somewhere for them to occur.

Notes

[1] This example is taken from W. G. V. Rosser, *Introducing Relativity* (Butterworth), an extremely useful book pitched at about first year undergraduate level.

[2] The idea of an 'event' is also indispensable in Quantum Mechanics since it is a punctual, objective occurrence with observable consequences that collapses the diffuse wave function and obliges a 'particle' to take up a specific position and the cat to unequivocally die. However, the phenomenon is nothing like so clear-cut as in Relativity since there is the troublesome question of the observer's active and usually deliberate involvement in the 'event'—which is why the literature prefers to speak of a 'measurement' rather than an 'occurrence' even if nothing is being measured. Also, the wave function in its uncollapsed state cannot by any stretch of imagination be considered to be composed of events which is why one has to speak of 'probabilities', 'potentialities' and so forth.

[3] Lee Smolin, Atoms of Space and Time, *Scientific American, Special Edition, A Matter of Time*, vol. **16** no. 1 (2006). This special edition also includes a very good article by Ronald Lasky on the Twins Paradox from the standpoint of Special Relativity, to which I am indebted (though the author's views on the issue are not necessarily the same as mine).

The answer to the question of when the astronaut twin starts to age more than his earthbound brother is: about 3,200 kilometres above sea level. We first obtain an expression for $\Delta\nu/\nu_0$ (where $\Delta\nu$ is the difference in frequencies of the clock in the satellite compared with the frequency ν_0 of an Earthbound clock) according to Special Relativity. This change is approximately $-\frac{1}{2}v^2/c^2$ and using $mv^2/r = mGM/r^2$ we arrive at

$$\frac{\Delta\nu}{\nu_0} \approx -\frac{\frac{1}{2}\,g\,r_0}{c^2}\,\frac{r_0}{r}.$$

Adding this to the expression already obtained for $\Delta\nu/\nu_0$ according to the difference in gravitational potential we obtain a rough result

$$\frac{\Delta\nu}{\nu_0} \approx \frac{g\,r_0}{c^2}\left(1 - \frac{3\,r_0}{2r}\right),$$

where r_0 is the mean radius of the Earth and r is the radius of the satellite's orbit considered to be circular. See pp. 272–3 of Rosser, *op. cit.*

Solution 212.3 – 100 seats

> There are 100 seats on a plane, and 100 people have booked (different) seats. They form a queue. The first person to board the plane ignores the instructions on his ticket and chooses a seat at random. Thereafter each passenger goes to his/her allocated seat if it is unoccupied and otherwise chooses an unoccupied seat at random. What is the probability that the last person gets her booked seat?

Tony Forbes

This is an interesting problem but on reflection I now think that we were probably guilty of making it unnecessarily difficult by not telling you the answer. Certainly in my own experience probability problems are quite often impossible to solve with any degree of confidence if you do not already have the solution. Proceeding in ignorance, let us see what happens on 100000 typical flights of the plane.

957	99025	99031	98957	98997	98986	98997	98977	98910	98911
98925	98891	98837	98912	98831	98805	98789	98796	98792	98797
98747	98689	98754	98710	98692	98646	98714	98626	98597	98620
98596	98542	98590	98472	98534	98469	98423	98398	98399	98395
98346	98410	98334	98300	98128	98183	98213	98101	98193	98059
98015	97893	97905	97939	97730	97835	97830	97757	97708	97577
97545	97531	97380	97311	97276	97249	97215	97070	96955	96918
96682	96596	96576	96366	96284	96220	95955	95972	95649	95542
95130	94941	94684	94385	94066	93857	93467	92950	92332	91709
90954	89948	88859	87629	85754	83168	79868	74974	66489	49977

These are the numbers of times the i th passenger gets correctly seated. So the last person gets her seat about 49977 times in 100000 trials; hence the answer to the problem is obviously 1/2. Also it seems even more is true. If you look at the rational numbers obtained by dividing the counts by 100000, you will see that apart from the first they approximately form the sequence

$$\frac{99}{100},\quad \frac{98}{99},\quad \frac{97}{98},\quad \ldots,\quad \frac{3}{4},\quad \frac{2}{3},\quad \frac{1}{2}.$$

Notice, by the way, that by choosing a seat at random the first passenger does not necessarily occupy a seat other than the one he booked. Therefore with probability 1/100 all passengers get their booked seats.

Now that we know the answer we can argue with confidence. The number 100 is not particularly relevant. So we assume there are n passengers and it is just as easy to demonstrate that the same sequence of rational numbers occurs but starting with $(n-1)/n$ instead of 99/100. We wish to prove that

$$\mathbb{P}\,(\text{passenger } i \text{ correctly seated}) \;=\; \begin{cases} \dfrac{1}{n} & \text{if } i = 1, \\ \dfrac{n-i+1}{n-i+2} & \text{if } 2 \le i \le n. \end{cases} \qquad (*)$$

Clearly $(*)$ holds if $n = 2$ or if $i = 1$. Assume $(*)$ is true for all n from 2 to $N-1$, say, and we shall prove $(*)$ for $n = N$.

Consider passenger i, $2 \le i \le N$, and suppose passenger 1 chooses seat s. If $s = 1$, everybody gets their booked seats—probability $1/N$. If $s = i$, passenger i definitely does not get correctly seated. If $s > i$, passenger i definitely gets correctly seated—probability $(N-i)/N$.

If $s < i$, then passenger i is in the same situation as in the original problem but with only $N-s+1$ seats, namely $1, s+1, s+2, \ldots, N$, and the passengers renumbered by subtracting $s-1$. Thus passenger s is now playing the role of the original passenger 1. Since $2 \le N-s+1 < N$ we can use $(*)$ to compute the probability of passenger i (who is now called $i-s+1$) getting correctly seated:

$$\frac{(N-s+1)-(i-s+1)+1}{(N-s+1)-(i-s+1)+2} \;=\; \frac{N-i+1}{N-i+2}.$$

Putting everything together, we find that the probability of passenger i getting his/her booked seat is

$$\frac{1}{N} + \frac{N-i}{N} + \frac{1}{N}\sum_{s=2}^{i-1} \frac{N-i+1}{N-i+2} \;=\; \frac{N-i+1}{N-i+2},$$

and thus $(*)$ is proved.

Problem 217.1 – Another 100 seats

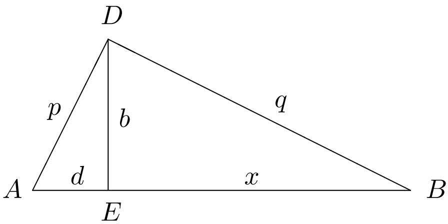
This interesting variation on Problem 212.3 was communicated to me [ADF] via Peter Cameron.

There are 100 seats on a plane, numbered 1–100. There are 100 passengers, also numbered 1–100. Each passenger has a preferred seat. The passengers arrive in order. On arrival, a passenger sits in his/her preferred seat if possible; otherwise he/she goes to the next available seat in the numerical ordering. If no such seat exists, the passenger leaves the plane. In how many of the 100^{100} possible choices for the preferred seats do all passengers embark successfully?

The ladder problem yet again: Problem 204.9

John Bull

A ladder of length 1 stands against a vertical wall just touching a shed of height and width b. Find the distance of the ladder bottom from the shed.



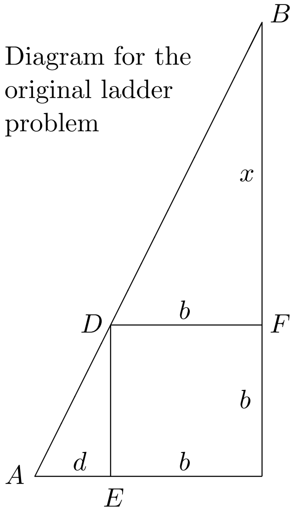


Diagram for the original ladder problem

Rotate B about D until B is on AE extended. We then transform Figure 1 to Figure 2; ADB is now a right triangle with ED an altitude to the right angle at D. This construction can be found in Euclid VI.8 and in elementary geometry text books [1, 2]. Labelling would normally follow a different convention, but we shall stay with that of the original problem with, additionally, $AD = p$. So in our case $p + q = 1$.

In Figure 2, triangles ADB, DEB and AED are all similar. So

$$\frac{AD}{DE} = \frac{DB}{EB} = \frac{BA}{BD}, \quad \frac{DE}{AE} = \frac{EB}{ED} = \frac{BD}{DA}, \quad \frac{AE}{AD} = \frac{ED}{DB} = \frac{DA}{BA},$$

Or

$$\frac{p}{b} = \frac{q}{x} = \frac{d+x}{q}, \quad \frac{b}{d} = \frac{x}{b} = \frac{q}{p}, \quad \frac{b}{p} = \frac{p}{q} = \frac{p}{d+x}.$$

This gives

$$p^2 + q^2 = d(d+x) + x(d+x) = d^2 + 2dx + x^2 = (d+x)^2,$$

which offers a proof of Pythagoras' theorem on triangle ADB. Hence

$$(d+x)^2 = p^2 + q^2 = (p+q)^2 - 2pq = (p+q)^2 - 2b(d+x)$$

and with $p + q = 1$ implies $(d + x)^2 + 2b(d + x) - 1 = 0$. Also $dx = b^2$. The solution now follows previous M500 articles: Steve Moon (**212**, p.12) and Nick Hobson (**214**, p.18). But note that we have proved Pythagoras' theorem along the way rather than have simply used it.

[1] Clement V. Durrell, *A New Geometry for Schools*, 1957, page 504.

[2] Hall & Stevens, *A School Geometry*, 1902, page 268.

Dick Boardman

I expect you are fed up with ladder problems. However, I may as well add my halfpenny contribution. Referring to the notation in the diagram on the previous page, we start with equations

$$d\,x \;=\; b^2 \tag{1}$$

and

$$d^2 + 2\,b\,d + x^2 + 2\,b\,x + 2\,b^2 \;=\; 1 \tag{2}$$

given in Nick Hobson's solution on page 18 of M500 **214**.

Using a d-axis and an x-axis, equation (1) is a rectangular hyperbola and equation (2) is a circle, centre $(-b, -b)$. Both of these curves are symmetrical about the line $d = x$. So rotating the axes by 45 degrees produces a quartic with no cubic or linear terms. Substitute $d = X + Y$ and $x = X - Y$ and the result falls out.

Problem 217.2 – Chords and regions

Sebastian Hayes

We have n points situated irregularly on the circumference of a circle. They are joined by straight lines in all possible ways. What is maximum number of regions into which the lines divide the circle?

Problem 217.3 – Integral

Show that

$$\int_1^{\sqrt[3]{3}} t^2 dt \times \cos\frac{3\pi}{9} \;=\; \log\sqrt[3]{e},$$

and then translate the problem into limerick form. [We have left 3/9 as is to make our version of the fourth line work, but if you can do the job without this artifice, so much the better. Similar contributions welcome.]

Solution 214.5 – 1000000 tarts

> There are one million tarts; all weigh the same except for 100, which are too light. There is also a weighing machine that will indicate whether or not a batch of tarts has the correct weight. Devise a testing strategy to identify the 100 defective tarts with a small maximum number of weighings.

John Smith

A bound for the best possible algorithm

There are $C_{100}^{1000000}$ ways of ordering 100 bad tarts in a collection of 1000000 tarts; n weighings can, at best, distinguish 2^n orderings. So the best scheme must require at least n weighings where n is the smallest integer satisfying $2^n \geq C_{100}^{1000000}$. A bit of electronic computing shows that

$$\log_2 C_{100}^{1000000} = 1468.384722,$$

so that, in the worst case, any weighing scheme will require at least 1469 measurements.

A simple algorithm

A simple algorithm would use repeated bisection.

Until all bad tarts are found, repeatedly:

Divide the batch into two equal sized batches:

weigh one batch;

if it is bad, put the other batch aside;

if it is not bad, put it aside;

until the batch size is 1.

Then the tart in the remaining batch must be bad.

Put the bad tart to one side, and start again.

This way we can locate a bad tart in a sample of 2^n tarts in n weighings. For an initial $N = 1000000$, we need 20 weighings. By repeating the process 100 times, we can find all 100 bad tarts in 2000 weighings.

The inefficiency of this method lies in the fact that, for the first few weighings of each cycle, it is almost certain that the batch is bad. Thus very little information is gained from the measurement.

Some hand-waving analysis

A good first weighing should divide the $C_{100}^{1000000}$ possible orderings into two roughly equal sized orderings. If we choose a batch of n tarts at random, the batch will contain, on average, $100n/1000000$ tarts. If n is neither too large nor too small, the number of bad tarts will roughly follow a Poisson distribution with mean $100n/1000000$. The probability of a batch of n tarts having no bad tarts is thus roughly $\exp(-100n/1000000)$. An ideal measurement will be equally likely to show bad or correct weight. So we should choose n such that

$$\exp \frac{-100\,n}{1000000} = \frac{1}{2},$$

which gives

$$n = \frac{1000000}{100} \log 2.$$

In the more general care where we have N tarts of which k are bad, then the first weighing should be of a batch of about $0.7N/k$ tarts.

A better algorithm

We use analysis to give the following possible algorithm.

Suppose we have N tarts, of which k are bad.

If $N = k$, then all the tarts are bad.

Otherwise, select 2^n of the tarts, where n is the largest integer satisfying $2^n \leq (N-1)/k$. Weigh the 2^n tarts.

If they have the correct weight, then put them to one side, and solve the remaining problem of $N - 2^n$ tarts, of which k are bad.

If they do not have the correct weight, then repeatedly divide the 2^n sample into 2, and so find a bad tart in another n weighings. Then solve the problem of $N-1$ tarts, of which $k-1$ are bad.

In symbols, let $W(N,k)$ be the number weighings required to find k tarts in a batch on N. A recurrence relation for $W(N,k)$ is

$$W(N,k) \;=\; \begin{cases} 0 \text{ if } N = k; \quad \text{otherwise} \\ \max(1 + n + W(N-1, k-1), 1 + W(N-2^n, k)), \end{cases}$$

where n is the largest integer satisfying $s^n \leq (N-1)/k$.

Analysis of this recurrence relation to give a formula for $W(N,k)$ is too hard (for me). But for a computer, it only requires 10^8 iterations, plus storage for 2000000 integers. My computer analysis suggests that $W(1000000, 100) = 1521$. This number of 1521 is fairly close to the ideal figure of 1469, and certainly much smaller than the figure of 2000 weighings from the simple algorithm.

Six-region sudoku puzzles

Tony Forbes

As usual we begin with a sudoku puzzle. Complete the array to make a Latin square on $\{1, 2, \ldots, 9\}$ such that each of the 3×3 boxes into which the array is divided contains all of the symbols 1–9. The solution is unique.

				3	7			
		6					1	
8					2	7	6	
		4						2
1			5		3			7
2						9		
	8	3	2					6
	5					8		
			3	9				

Puzzle A

Four coordinates and six regions

Consider the 9×9 grid of a standard sudoku puzzle. Instead of labelling rows and columns 0–8, let us imagine that we are using base 3 numbers, 00, 01, 02, 10, 11, 12, 20, 21, 22, instead. Or, looking at it another way, we use a two-coordinate system. For rows, the first digit, or coordinate, is the row block number and the second coordinate is the row number within the row block. Similarly for columns.

Now a cell has *four* coordinates, (a, b, c, d): a = row block number; b = row number within row block a; c = column block number; d = column number within column block c. To recover the original coordinates you just multiply by 3 and add; $(a, b, c, d) \to (3a + b, 3c + d)$.

A *region* is the set of cells obtained by fixing a certain pair of coordinates, and moreover it can be identified by the fixed coordinates. Thus row ab is the set of cells where a and b are fixed and c and d each run from 0 to 2, column cd is the set of cells where c and d are fixed and a and b vary, and box ac is the set of cells where a and c are fixed and b and d vary.

So we have identified rows, columns and boxes by fixing two out of the four cell coordinates. Using $*$ to denote a coordinate that varies from 0 to

2, we have rows: $(a, b, *, *)$, columns: $(*, *, c, d)$ and boxes : $(a, *, b, *)$. But there are three further ways of fixing two coordinates, and we can therefore introduce three new region types according to the chosen pair. We call the regions defined by $(*, b, c, *)$, $(a, *, *, d)$ and $(*, b, *, d)$ *split rows*, *split columns* and *split boxes* respectively—in agreement with the explanation given in M500 **214**.

region type	coordinates	region type	coordinates
row	$(a, b, *, *)$	split row	$(*, b, c, *)$
column	$(*, *, c, d)$	split column	$(a, *, *, d)$
box	$(a, *, c, *)$	split box	$(*, b, *, d)$

A split row consists of three rows of three, spaced three apart in one of the three column blocks (like the cells marked 'r' and 'x' in the array, below). A split column consists of three columns of three, spaced three apart in one of the three row blocks ('c' and 'x'). A split box is a 3 x 3 square array of cells, spaced three apart in both directions ('b').

		0			1			2			c
a	b	0	1	2	0	1	2	0	1	2	d
	0	.	.	c	.	.	c	.	.	c	
0	1	r	r	x	.	.	c	.	.	c	
	2	b	.	c	b	.	c	b	.	c	
	0	.	.	.	.	.	.	.	.	.	
1	1	r	r	r	.	.	.	.	.	.	
	2	b	.	.	b	.	.	b	.	.	
	0	.	.	.	.	.	.	.	.	.	
2	1	r	r	r	.	.	.	.	.	.	
	2	b	.	.	b	.	.	b	.	.	

We define a *6-region sudoku square* as a 9×9 Latin square where every row, column, box, split row, split column and split box contains the symbols $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. A *6-region sudoku puzzle* is a sudoku puzzle with the additional constraint that the split rows, split columns and split boxes must also contain the symbols 1–9.

[Acknowledgement. This is based on an idea of Robert Connelly of Cornell University and communicated to me by Peter Cameron of Queen Mary College, London.]

Symmetry

As well as any I haven't thought of, the valid symmetry operations for a 6-region sudoku square include the following:

(i) rotation by 90 degrees anticlockwise;
(ii) reflection in the middle row or the middle column;
(iii) reflection in a main diagonal;
(iv) swap any two row blocks; swap any two column blocks;
(v) any permutation of the symbol set;
(vi) any permutation of the four coordinates.

On the other hand, swapping two rows is never a legitimate operation.

The last one in the list, (vi), is especially interesting because it can change the shapes of the regions. We are already familiar with one special case: transposition is equivalent to interchanging the row and column coordinates, $(a, b, c, d) \to (c, d, a, b)$. But now we have four coordinates and hence 24 permutations; so there is plenty of scope for experimentation. Take the permutation of the coordinates that sends the symbol at position (a, b, c, d) to position (b, c, a, d), and on examining the effect it has on various regions, you can verify that, for example, you should get successively row, box, split row and back to row.

1	2	3	4	5	6	7	8	9
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

→

1	2	3	.	.	.	.	.	.
4	5	6	.	.	.	.	.	.
7	8	9	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

→

1	2	3	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
4	5	6	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
7	8	9	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

Symmetric 6-region sudoku puzzles remain symmetric under any permutation of the four coordinates.

To prove this, let π be any permutation of the four coordinates. We know that the operation $(a, b, c, d) \mapsto \pi((a, b, c, d))$ preserves the 6-region sudoku property. Now define another operation on the four coordinates,

$$\mu : (a, b, c, d) \mapsto (2 - a, 2 - b, 2 - c, 2 - d)$$

and observe that applying μ to a cell moves it to its diametrically opposite position (rotation by 180°). Hence after applying μ to a symmetric 6-region puzzle, the result is also a symmetric 6-region puzzle.

But π and μ commute; that is $\mu(\pi((a,b,c,d))) = \pi(\mu((a,b,c,d)))$ for any cell coordinates (a,b,c,d). For example, if π maps (a,b,c,d) to (b,c,d,a), we have:

$$\begin{tikzcd}
(a,b,c,d) \arrow[r, "\pi"] \arrow[d, "\mu"] & (b,c,d,a) \arrow[d, "\mu"] \\
(2-a,2-b,2-c,2-d) \arrow[r, "\pi"] & (2-b,2-c,2-d,2-a)
\end{tikzcd}$$

How to solve 6-region sudoku puzzles

An interesting feature is the influence of a single digit on the rest of the array. For instance, suppose the middle cell is a 1. Applying the basic rules eliminates 1 from all cells marked 'x' in the array on the left, below. But with a little more work you can prove—as explained in issue **214**—that 1 is also forbidden in the cells marked 'z' (right-hand array).

1-free cells

.	.	.	.	x	.	.	.	.
.	x	.	x	x	x	.	x	.
.	.	.	.	x	.	.	.	.
.	x	.	x	x	x	.	x	.
x	x	x	x	1	x	x	x	x
.	x	.	x	x	x	.	x	.
.	.	.	.	x	.	.	.	.
.	x	.	x	x	x	.	x	.
.	.	.	.	x	.	.	.	.

More 1-free cells

z	.	z	.	x	.	z	.	z
.	x	.	x	x	x	.	x	.
z	.	z	.	x	.	z	.	z
.	x	.	x	x	x	.	x	.
x	x	x	x	1	x	x	x	x
.	x	.	x	x	x	.	x	.
z	.	z	.	x	.	z	.	z
.	x	.	x	x	x	.	x	.
z	.	z	.	x	.	z	.	z

I leave it for you to construct similar patterns where the 1 is placed in other positions, and as a consequence you can prove that the two main diagonals of a 6-region sudoku square contain the symbols 1–9. By the way, this feature only works for 9×9 puzzles. If you try it with a 16×16 grid, you will find that you can restrict the occurrence of 1s to 9 positions in 6 boxes, but since $6 < 9$, the corresponding deduction cannot be made.

As with standard sudoku, because any permutation of the symbol set is a valid symmetry operation a 6-region sudoku puzzle must have at least eight starter digits. But but now the extra restrictions on the solution allow the minimum to be attained. Indeed, the most elegant puzzles have eight starter digits and 2-fold rotational symmetry, like Puzzles B, C, D and E, below. And because 8 is even, the central cell, at coordinates $(1,1,1,1)$, will not contain a starter digit. However, it is easy to show that only the missing symbol is valid in this position.

There is also a general principle which halves the work involved in solving a symmetric 6-region sudoku puzzle, but I expect you can figure that out for yourself while you attempt Puzzles B, C, D and E. And of course Puzzle A, a truly fiendish standard sudoku puzzle, whose unique solution actually conforms to the 6-region rules!

	7							2
					4	9		
		1	8					
6							3	

Puzzle B

	9							
				1		7		
8								4
		3		5				
							2	

Puzzle C

							6	
			1				2	
	4						7	
	3				5			
	9							

Puzzle D

	5							
3				6				
	7						8	
				9				2
							4	

Puzzle E

Problem 217.4 – n^2

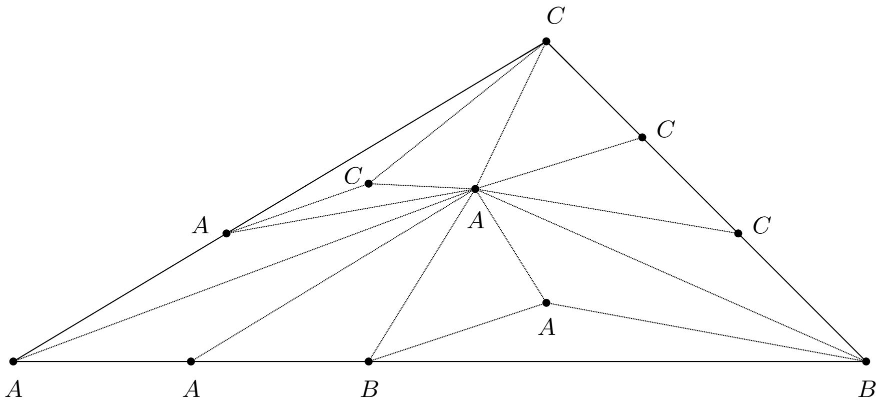
John Bull

Where n is a positive integer, prove that $(n+1)^n - 1$ is divisible by n^2.

Problem 217.5 – Triangulating a triangle

Take any triangle and label its vertices A, B and C. Subdivide the triangle into smaller triangles by adding new points in its interior or on its boundary and adding new edges. Label the new points A, B and C in any way you like subject to the restriction that points on the edge opposite vertex x of the original triangle must not be labelled x.

Prove that the number of new triangles labelled ABC is odd (and therefore positive).



Problem 217.6 – Triangle

Take any triangle and label its vertices A, B and C. Let a, b, c denote the lengths of the sides opposite A, B, C respectively. Show that

$$\log c \;=\; \log a - \frac{b}{a}\cos C - \frac{b^2}{2a^2}\cos 2C - \frac{b^3}{3a^3}\cos 3C - \dots .$$

Problem 217.7 – Exponents

Find interesting positive rational numbers x and y such that x^y/y^x is an integer. For instance,

$$\frac{2^4}{4^2} \;=\; \frac{\left(3\frac{3}{8}\right)^{2\frac{1}{4}}}{\left(2\frac{1}{4}\right)^{3\frac{3}{8}}} \;=\; 1.$$

Contents M500 217 – August 2007

04.01

Cover: A decomposition into four mutually disjoint caps of size 20 of the 80 non-zero points of AG$(4, 3)$, the affine geometry of dimension four over GF(3). Points in AG$(4, 3)$ are four-vectors of elements taken from $\{0, 1, 2\}$. Lines are triples of points $\{X, Y, Z\}$ where $X + Y + Z = 0$ with coordinate-wise addition modulo 3. A cap is a set of points no three of which form a line. Twenty is the maximum possible size of a cap in AG$(4, 3)$. **ADF**