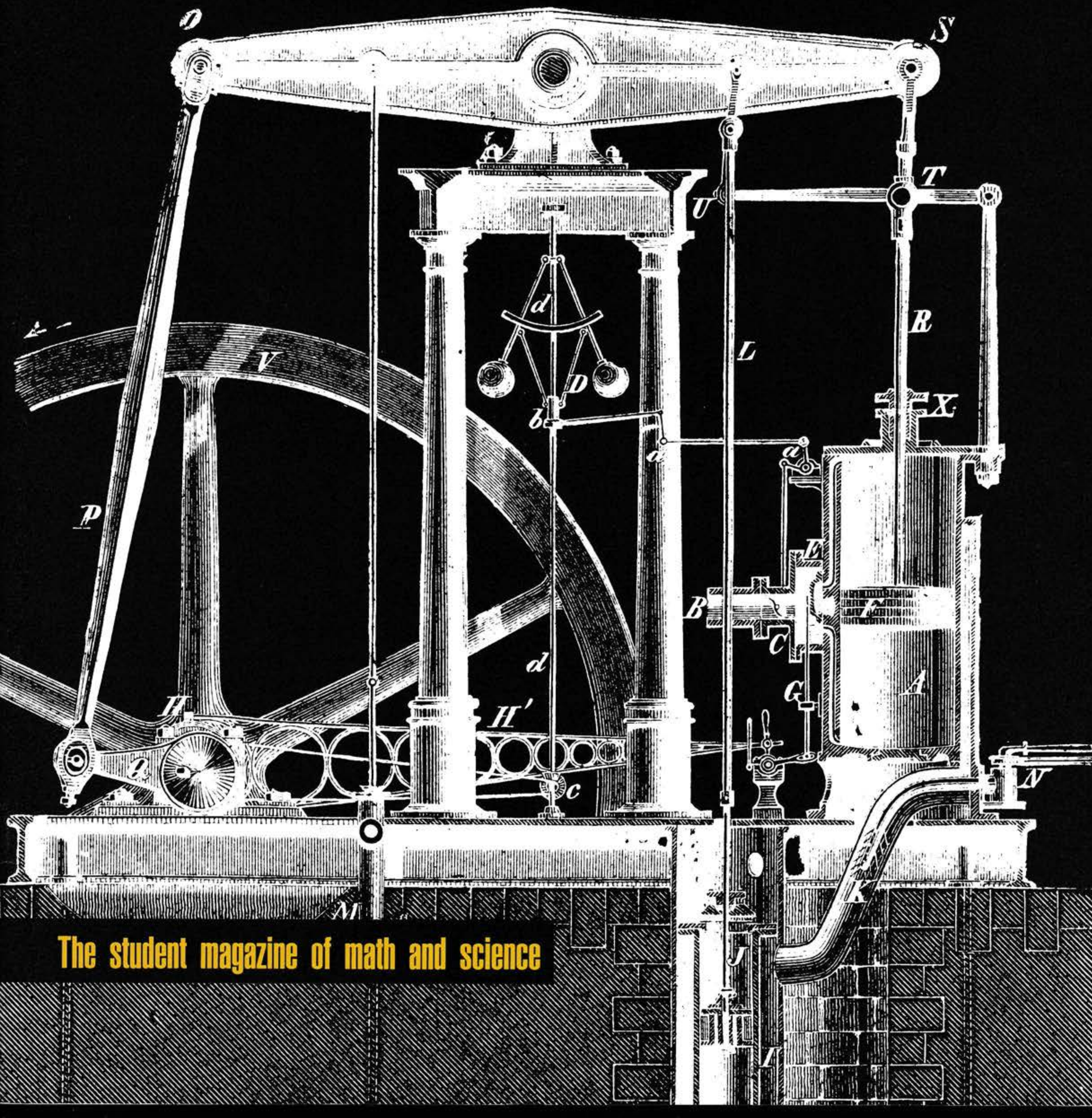
QUANTUM

NOVEMBER/DECEMBER 1990

$5.00



The student magazine of math and science

GALLERY Q

National Gallery of Art, Washington (Gift of Mr. and Mrs. Stephen M. Kellen) © NGA

Siberian Dogs in the Snow (1909–1911) by Franz Marc

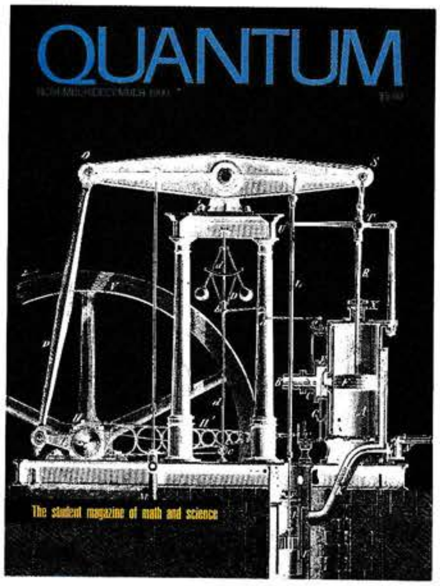
When we decided to go to extremes in this issue of *Quantum* (see A.L. Rosenthal's article on page 8), all kinds of extremes sprang to mind. As we enter the cold season in the Northern Hemisphere, extremes of temperature were at the forefront. And when we're almost blinded by a field of snow, we perceive white as one extreme of a continuum of color (whether we think of it as all colors combined or removed, depending on the medium). Franz Marc (1880–1916) touches on these extremes in his expressive portrait of his sheepdog Russi, seen from two angles.

Marc spent much of his brief career painting animals. (His life was cut short near Verdun during World War I.) He developed a profound nature mysticism that, combined with an urge toward abstraction and a symbolism of colors, tended to produce intensely colored canvases of animal and vegetable life. Marc believed this was the best way to express the conflicts and resolutions of natural forces that civilization shields from us or teaches us to ignore.

Some elements of his later style are absent from *Siberian Dogs in the Snow*—most obviously, the color symbolism. At this point in his development Marc was more concerned with the interaction of color and light. In a letter to a fellow artist, Marc described how the painting arose out of an experiment in the use of a prism to clarify tonal relationships.

QUANTUM

NOVEMBER/DECEMBER 1990 VOLUME 1, NUMBER 2



For the Scottish engineer and instrument maker James Watt, the 1780s were a very productive decade. Fifteen years earlier, while working on a Newcomen steam engine, he greatly improved its efficiency by adding a separate condenser chamber. But in 1781 a business partner urged him to invent a rotary steam engine for use in corn, malt, and cotton mills, and Watt went to work. In that year he devised the sun-and-planet gear, which allowed a shaft to produce two revolutions for each stroke of the engine. In 1782 he patented the double-acting engine, in which the piston pulled as well as pushed. This engine required a new method of rigidly connecting the piston, engaged in linear motion, to another part, engaged in rotary motion. So in 1784 he came up with the required linearizing device. Watt considered this "one of the most ingenious, simple pieces of mechanism I have contrived," and it's the subject of "Making the Crooked Straight" on page 20. In 1788 he added a centrifugal governor to automatically control the speed of the engine, and with his invention of the pressure gauge in 1790, the Watt engine was all but ready to make its dramatic contribution to the Industrial Revolution.

For a look at a cleaner, quieter device at the forefront of modern technology, turn to "Lightning in a Crystal" on page 12.

FEATURES

DEPARTMENTS

Notes of a traveller

And news of a partnership

THESE ARE EXCITING TIMES we live in, and more of us than ever before are finding ourselves on planes heading to or from Moscow. *Quantum* staff will be spending a week in the Soviet capital, planning future issues and working to improve the logistics of our bihemispheric production.

One recent visitor to the USSR was Lynn Arthur Steen, who teaches mathematics at St. Olaf College in Minnesota and serves as a member of the Mathematical Sciences Education Board of the National Research Council. During his week-long stay he investigated the Soviet approach to math education. What follows are Professor Steen's impressions and thoughts about math and science education in the two countries, which both students and teachers will no doubt find of interest.

In the past few years Americans have learned quite a bit about the Soviet Union. We know that their economy is deteriorating, national strife is increasing, and their military empire is crumbling. We can also see that the USSR is seeking to integrate itself into the world economy and expand contacts in all areas.

One thing America did not learn from news coverage of recent US-Soviet relations, however, is how the USSR has managed to produce so many talented mathematicians and scientists, who shocked us with Sputnik and continue to impress us in olympiads and scientific exchanges. The answer lies in one of the Soviet Union's best-kept secrets: a system of mathematics education that produced a tradition of excellence in research that is as good as any produced in Western countries.

Even as Gorbachev was touring the United States this past spring, a small delegation of US mathematicians visited Moscow at the invitation of Yevgeny Velikhov of the Soviet Academy of Sciences to explore means of cooperation in mathematics education. The invitation was especially timely, since math and science education in the United States is currently under siege.

Many parallels between mathematics education in the Soviet Union and in the US can be seen, but the differences are more striking. The US can learn much from both the similarities and the differences.

Just as President Bush has laid out national goals for mathematics and science education for the United States, so Gorbachev has established a commission to improve mathematics education in the Soviet Union. The emphasis in the USSR is to increase the role of computers in education at all levels.

In the Soviet Union, just as in the United States, there is great unevenness from school to school, and from teacher to teacher, in the quality of mathematics education. Both nations have responded with similar interventions: special high schools for math and science and university-based enrichment programs for students who can benefit from greater challenges.

Both countries debate how best to deploy limited resources for math education. Conservatives (mostly university professors) prefer programs that nurture highly talented students, wherever they can be found; reformers seek to "raise the water table" by improving mathematics education for everybody.

In one important area, however, there is a striking contrast between the US practice and the Soviet tradition: testing. US students go through sixteen years of short-answer, multiple-choice tests in mathematics, beginning with number facts in primary school and continuing right through a multiple-choice Graduate Record Exam administered to college seniors. In the USSR, mathematics tests are often given in oral or written (essay) form, emulating the type of environment in which mathematical ideas are used in the working world.

Bite-sized test items eviscerate education as surely as TV sound bites trivialize politics. In contrast, open-ended tests requiring holistic responses encurage higher-order thinking and creative problem solving.

Students in the USSR learn from their experience with school tests to think before answering. US students instead train for rapid response, learning how to take tests rather than how to solve problems. In Soviet schools tests are used as an intrinsic part of the curriculum, and the teacher's responses focus on each individual student in order to prevent failure.

The mathematics curriculum in the USSR is, for the most part, more formal and traditional than that becoming common in the United States. The mathematical tools of academia predominate; those of the state or business (for instance, statistics, discrete mathematics) are almost invisible. So in this respect US schools appear better attuned to the real needs of society.

But we have a lot to learn from the USSR in the area of testing. Tests should be part of the curriculum—an opportunity to learn and be taught—not separate from it. They should enable students to reveal what they can do, not merely what they don't know or can't quickly recall. If we are to be number one in mathematics and science, as President Bush has urged, we need tests that measure what's important, not just what's easy and cheap to grade.

As part of NSTA's efforts to reform the scope, sequence, and coordination of secondary science education in this country, we are developing a prototype interactive digital video disk teaching system for high-level ability assessment. Rather than requiring students to recall isolated facts about phenomena, this exciting technology will allow measurement of a student's understanding of scientific concepts. The interactive optical disk may prove to be an important element in a new approach to teaching and testing.

I'M HAPPY TO ANNOUNCE that *Quantum* has entered into an agreement with the international publisher Springer-Verlag, which is based in Heidelberg, Germany, and has offices in New York, Tokyo, London, and elsewhere. The National Science Teachers Association will retain editorial control over *Quantum*, and our working relationship with Quantum Bureau in Moscow will remain the same. Springer will handle our printing, subscriptions, and mailing; NSTA and Springer will both engage in promotion and solicitation of advertisements.

As part of the agreement, *Quantum* will be published bimonthly throughout the year beginning with the September/October 1991 issue, so those who have subscribed at full price (as opposed to the introductory price of $9.95) will receive six issues, not four. Those who renew will, of course, receive six issues per year.

We welcome Springer to the *Quantum* venture. We are confident that the impressive resources of Springer-Verlag will help make *Quantum* available wherever English is spoken or taught in schools. With that kind of exposure, *Quantum* is more likely to attract high-quality submissions, and our readers will share in the excitement of being part of an international experience.

—Bill G. Aldridge

Be a factor in the QUANTUM equation!

Have you written an article that you think belongs in *Quantum*? Do you have an unusual topic that students would find fun and challenging? Do you know of anyone who would make a great *Quantum* author? Write to us and we'll send you the editorial guidelines for prospective *Quantum* contributors. Scientists and teachers in any country are invited to submit material, but it must be written in colloquial English and at a level appropriate for *Quantum*'s predominantly high school readership.

Send your inquiries to:

Managing Editor
Quantum
1742 Connecticut Avenue NW
Washington, DC 20009-1171

QUANTUM

THE STUDENT MAGAZINE OF MATH AND SCIENCE

A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the USSR Academy of Sciences
in conjunction with
the American Association of Physics Teachers (AAPT)
& the National Council of Teachers of Mathematics (NCTM)

Quantum (ISSN 1048-8820) contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by the Academy of Sciences of the USSR and the Academy of Pedagogical Sciences of the USSR, as well as original material in English.

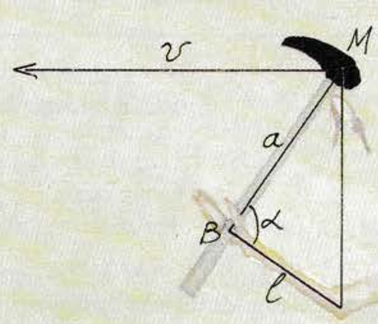
Subscription price for one year (six issues): individual, $18; student, $14; institution/library, $28; foreign, $26. Bulk subscriptions for students: 20-49 copies, $12 each; 50 + copies, $10 each. Correspondence about subscriptions, advertising, and editorial matters should be addressed to *Quantum*, 1742 Connecticut Avenue NW, Washington, DC 20009-1171.

This project was supported, in part, by the
National Science Foundation
Opinions expressed are those of the authors and not necessarily those of the Foundation

Tomahawk throwing made easy

No, it's not "all in the wrist"...

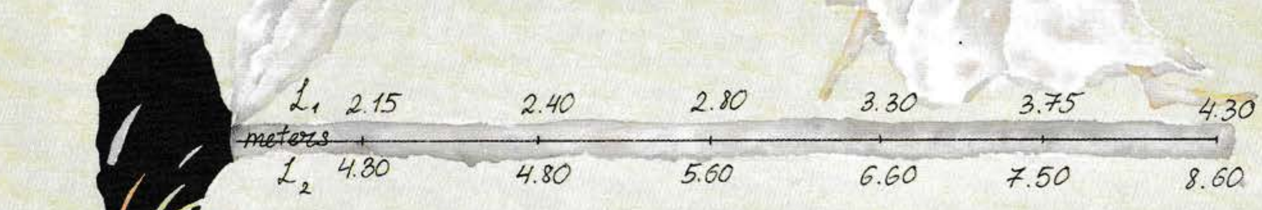
by V. A. Davydov



WHEN I WAS A BOY, BACK in the 1960s, my friends and I were fascinated by the novels of James Fennimore Cooper and others. We dreamed of the adventurous life among the Indian tribes of North America. Making a lasso out of a clothesline, we tried to catch a bush, a tree branch, or even an unfortunate cat chancing to emerge from the basement to doze in the sun. But most of all we envied the skill with which Indians wielded their menacing weapon—the tomahawk.

Just about every author of "Indian" novels devotes some pages to the wonderful art of tomahawk throwing. Our interest in the problem was kept at a fever pitch by the movies. At that time Indian films were very popular, and their heroes never missed a chance to throw a tomahawk. Take the following scene. An Indian tribe decides to punish a paleface. He's tied up and thrown against the wall of his bungalow, and each Indian throws his tomahawk at him. The last tomahawk cuts the rope, and when the unconscious victim slides to the ground the moviegoers see the outline of a human body formed by tomahawks stuck in the wall. After the film was over everyone was eager if not to master hatchet throwing (we realized that it was beyond the capacity of a paleface boy) then at least to understand the technique Native Americans used to throw tomahawks.

The younger generation isn't as interested anymore in "Indian questions." My own children, for instance,



"The axe cleaved the air in front of Heyward, and cutting some of flowing ringlets of Alice, buried itself and quivered in the tree above her head."
—James Fennimore Cooper, *The Last of the Mohicans*

Art by Sergey Ivanov

can't tell a Huron from a Comanche and hardly know who Osceola was. Nevertheless, they're still impressed by the fantastic ability of North American natives in manipulating their traditional weapon.

The basic idea behind the theory of hatchet throwing was born when we started hiking regularly. Finding a dry tree trunk near our campsite (and there are lots of dead trees in our forests), we'd try to hit it with a hatchet. We immediately discovered an interesting fact: if the person throwing the hatchet stands at a certain distance from the tree, the probability that the hatchet will stick in the tree (and not fall back after the hitting the tree with the butt or handle) suddenly increases. Only a little practice was needed to ensure that you'd hit the target tree, say, a hundred times out of a hundred—provided, of course, you were standing at the proper distance. My attempts to understand this phenomenon led me to formulate a model, which I'll now try to explain. You'll see that in order to master tomahawk throwing, you don't have to be an Indian. What's really needed is skill in estimating distances. Once you know how to do that, the rest is a cinch.

So let's take a look at the model. The problem is obviously divided into two parts. First, you have to be able to at least hit a pole or a tree trunk with a hatchet; second, you have to hit it with the cutting edge and not the butt or handle. I'll assume you can manage the first problem on your own.

While throwing, you move your hand in the following way. The arm holding the hatchet rotates at the elbow with an angular velocity ω, and the throw takes place when the velocity of the hatchet's center of mass is directed horizontally. Strictly speaking, if we want the hatchet to hit a certain spot on the pole, its direction at the moment of release might be something other than horizontal. But we've posed a more modest problem: how to embed the tomahawk in a vertical pole at any spot whatsoever. In this case we can ignore the effect of gravity. In an actual experiment the point of impact would be lower.

Art by Sergey Ivanov

In this model we also assume (and this is very important) that the hand doesn't give the hatchet any additional rotation. Try it yourself and you'll see that it's practically impossible to add rotation to the tomahawk's motion by moving your palm. You can only release the tomahawk and let it move freely.

Let's introduce the following parameters (see the figure on the facing page): l is the distance from the arm's center of rotation (the elbow) to point B on the handle where we hold the tomahawk; a is the distance from point B to point M, which is the center of the hatchet's mass; α is the angle between the arm and the handle. The angle α can have different values, but it's easier to throw when $\alpha \cong \pi/2$. The velocity of the hatchet's center of mass is then described by the following equation:

$$v = \omega\sqrt{a^2 + l^2}\,.$$

Now let's define the angular velocity of rotation of a thrown tomahawk. The simplest way to do that is to shift to a reference system that moves with the hatchet's center of mass with velocity v. Point M of this reference system (the center of mass) doesn't move, whereas point B (like every other point of the tomahawk) rotates around M; the velocity u of point B is at any moment directed perpendicularly to the handle and equals $\omega_B a$, where ω_B is the angular velocity of rotation of the flying tomahawk. In a stationary reference system the velocity ωl of point B at the moment of release is directed perpendicularly to l—that is, along the handle. So

$$u \equiv \omega_B a = \sqrt{v^2 - (\omega l)^2}\,.$$

Substituting

$$v = \omega\sqrt{a^2 + l^2}\,,$$

we get

$$\omega_B a = \omega a\,,$$

$$\omega_B \equiv \omega$$

—that is, a flying tomahawk rotates with an angular velocity ω equal to the angular velocity of the hand during the throw. This conclusion is valid even if angle α isn't $\pi/2$. In spite of its simplicity, this result is very important. It means that the ratio of the translational velocity of the tomahawk's center of mass to its angular velocity of rotation doesn't depend on the "force of the throw" (the momentum transferred to the hatchet at the moment of release) and equals

$$\frac{v}{\omega} = \sqrt{a^2 + l^2}\,.$$

This means that the distance L_n covered by a flying hatchet after n full rotations also doesn't depend on the throwing force. Since the time needed for n rotations is equal to $2\pi n/\omega$, we get the following distance:

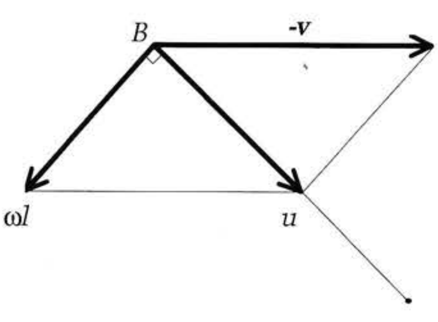
$$L_n = 2\pi n\sqrt{a^2 + l^2}\,.$$

This is really a remarkable conclusion: there's a range of distances L_n, and to make a successful throw you have to position yourself at the following distance from the target:

$$L_n + \arctan\left(\frac{l}{a}\right)\sqrt{a^2 + l^2}$$

(the second term arising because the hatchet's handle makes an angle of $\arctan(l/a)$ to the vertical at the moment of release).

Now let's estimate the magnitude of the elementary "quantum" L_1—that is, the distance covered by the hatchet after one full turn. Let l = 33 cm, a = 20 cm (measurements taken from my own arm and my own carpenter's hatchet). The calculation gives us L_1 = 2.42 m. So if I throw my hatchet from a distance of 2.82 m (don't forget to add the arc tangent term to L_1), it hits the target after one full turn.



My experience has shown that, with hardly any practice, you can hit the target from a distance of L_1 and L_2. Mastering the subsequent "quantum levels" is more difficult, but many friends of mine were able to hit a tree trunk from a distance of L_4 (more than 10 meters).

"This is all very nice," you may be thinking. "But an Indian can hit his target from other distances as well, not just from those equal to L_n. How do you explain that?"

It seems to be quite simple. There's a parameter a in our model that can be easily altered: all you have to do is shift the palm of your hand to another position on the tomahawk's handle. This shift modifies the whole range of throwing distances. It also changes the location of point B, which results in a "blurring" of L_n "levels" and the appearance of a "zonal structure." Inside each of the zones we can adjust point B (that is, the value of the parameter a) to a position ensuring a successful throw. But we can do more than that. If the tomahawk's handle is long enough, we can get the zones corresponding to adjacent levels L_n and L_{n+1} to overlap. Let's estimate the handle length b for which the nth level L_n equals the $(n + 1)$th level corresponding to the minimum possible distance a_{min} from point B to the center of mass M. This condition gives us the following equation:

$$2\pi n\sqrt{l^2+b^2} = 2\pi(n+1)\sqrt{l^2+a_{min}^2}\,.$$

My experience has shown that it's hard to throw the hatchet if a is less than 10 cm. So let's assume that a_{min} equals 10 cm. Solving the last equation, we get

$$b=\frac{1}{n}\sqrt{(2n+1)l^2+(n+1)^2a_{min}^2}\,.$$

An examination of this equation shows that the longest handle is needed to ensure the overlap of the first and second zones—that is, for $n = 1$. Substituting $l = 33$ cm, $a_{min} = 10$ cm, and $n = 1$, we get a handle length of $b \cong 60$ cm. Overlap of the second and third zones ($n = 2$) is achieved if the tomahawk's handle is 40 cm long, and so on. So to be able to hit any target from any distance, b has to be rather large. That's why Indian tomahawks have such long handles!

Actually, though, even shorter handles will do: the hatchet can hit the tree trunk at either the upper or lower part of the cutting edge, which brings the boundaries of the zones still closer. My experience suggests that a handle length of about 50 cm is quite sufficient.

Our model shows that there's no difficulty in mastering the art of tomahawk throwing. You just have to be able to judge the distance to the target and hold the hatchet at the right place. A good idea is to cut marks on the handle showing the respective target distances.

But how do *real* Indians throw tomahawks? It's quite possible they do it just the way I've described. Or maybe they know how to give the tomahawk an additional rotation with a flip of the hand? I have no idea, since I've never met a single Native American. I certainly hope that I will someday, and that I won't pass up the opportunity to learn more secrets of this remarkable art.

Tomahawk throwing is an exciting sport. Maybe in the not-so-distant future its practitioners will organize an association and sponsor tournaments. And—who knows—maybe one day this sport will even be included in the Olympic Games!

Just for the fun of it

Problems offered for your enjoyment by G.A. Halperin, V.V. Proizvolov, N.A. Rodina, L.M. Salakhov, and L.A. Steingraz

B16
Is the pattern shown in figure 1 symmetrical?

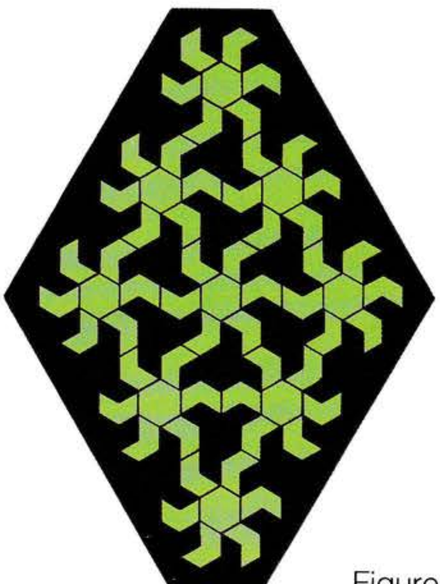
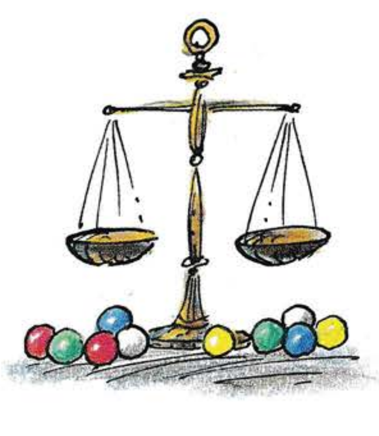


Figure 1

B17
You have two red balls, two blue ones, two green, two yellow, and two white. A number of balls of different colors are placed on the left pan of a balance while the other balls of the same colors are put on the right one. The balance tips to the left. If you exchange any pair of balls of the same color, however, the balance either tips to the right or stays even. How many balls are there on the balance?



B18
Move a single match in each row of figure 2 to get a true equality.

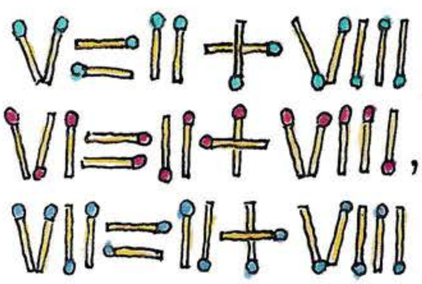


Figure 2

B19
A square is cut into a number of rectangles in such a way that no point of the square is a common vertex of four rectangles. Prove that the number of points of the square that are the vertices of rectangles is even.

B20
A steel ball floats in mercury. Will the depth of immersion increase or decrease as the temperature rises?

Art by Edward Nazarov

SOLUTIONS ON PAGE 59

Going to extremes

Sometimes an "end run" is more direct than a "dive up the middle"

by A. L. Rosenthal

IF YOU WANT TO ACQUIRE some skill in solving mathematical problems, you should try to master the more or less common approaches, techniques, and methods of mathematical reasoning. Here's one very general approach, which we'll call the "extremity rule."

The extremity rule can be succinctly stated in four words: "Consider the extreme case!" This is actually a recommendation to consider an object having extreme—or as mathematicians say, "extremal"—properties. If we're considering a set of points on a straight line, the rule tells us to focus our attention on the extreme left or extreme right point of the set. If the problem concerns a set of numbers, the extremity rule recommends that we consider its maximum and minimum. Here are some examples.[1]

Problem 1. *A set of points* M *is given in a plane such that each point in* M *is the center of an interval connecting a pair of points in* M. *Prove that set* M *is infinite*.

A good way to start is to consider a simpler but similar problem. So before doing problem 1 let's try this one.

Problem 2. *A set of points* M *is given on a straight line such that each point in* M *is the center of an interval connecting two other points belonging to* M. *Prove that set* M *is infinite.*

[1]Other examples of applying the rule can be found in recent issues of *Quantum*—for instance, in problems M10 and M15.—*Ed.*

Let's assume that M is a finite set and apply the extremity rule. If M is finite, it has extreme points—the extreme left and the extreme right. Consider one of them—for instance, the left one—and denote it by A. Point A is an extreme one and, consequently, can't lie inside the interval connecting two other points of the set M. The contradiction proves that M isn't a finite set.

There's another solution to the same problem, also based on the extremity rule. Assuming again that M is a finite set, consider the lengths of intervals connecting pairs of points in M. This set of numbers is finite. Applying our rule, consider the longest interval BC. Clearly, there are no points of M outside BC; otherwise, there would be longer intervals. Therefore, all the points of M lie on the interval BC, which implies that neither B nor C satisfies the above condition—again a contradiction.

Now let's return to problem 1. Assuming that M is a finite set, apply the extremity rule this way. Fix an orientation of the plane and consider the extreme left point of set M. If there are several "extreme left" points, choose the lowest one. You can easily see that this point (denoted by A) can't lie within an interval connecting two points of M. Indeed, if such an interval exists one of its ends is either to the left of point A or on the same vertical line with A but below it. Both situations contradict the choice of point A.

As with problem 2, there's another approach here. Consider the set of distances between pairs of points of M. If set M is finite, there is only a finite number of paired distances, so that the largest among them can be found. Let it be the distance between points A and B. But point B is the center of an interval CD whose ends, according to our assumption, belong to M (fig. 1). Now it's easy to prove that either AD or AC is longer than AB (do it yourself, making use of the fact that the median m drawn to one of the triangle's sides is less than half the sum of the two other sides).

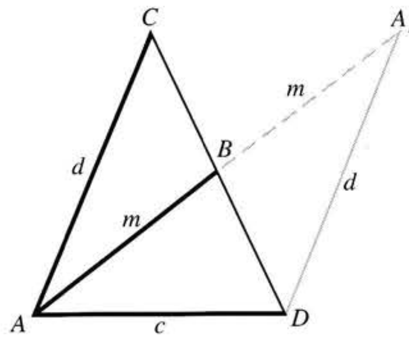
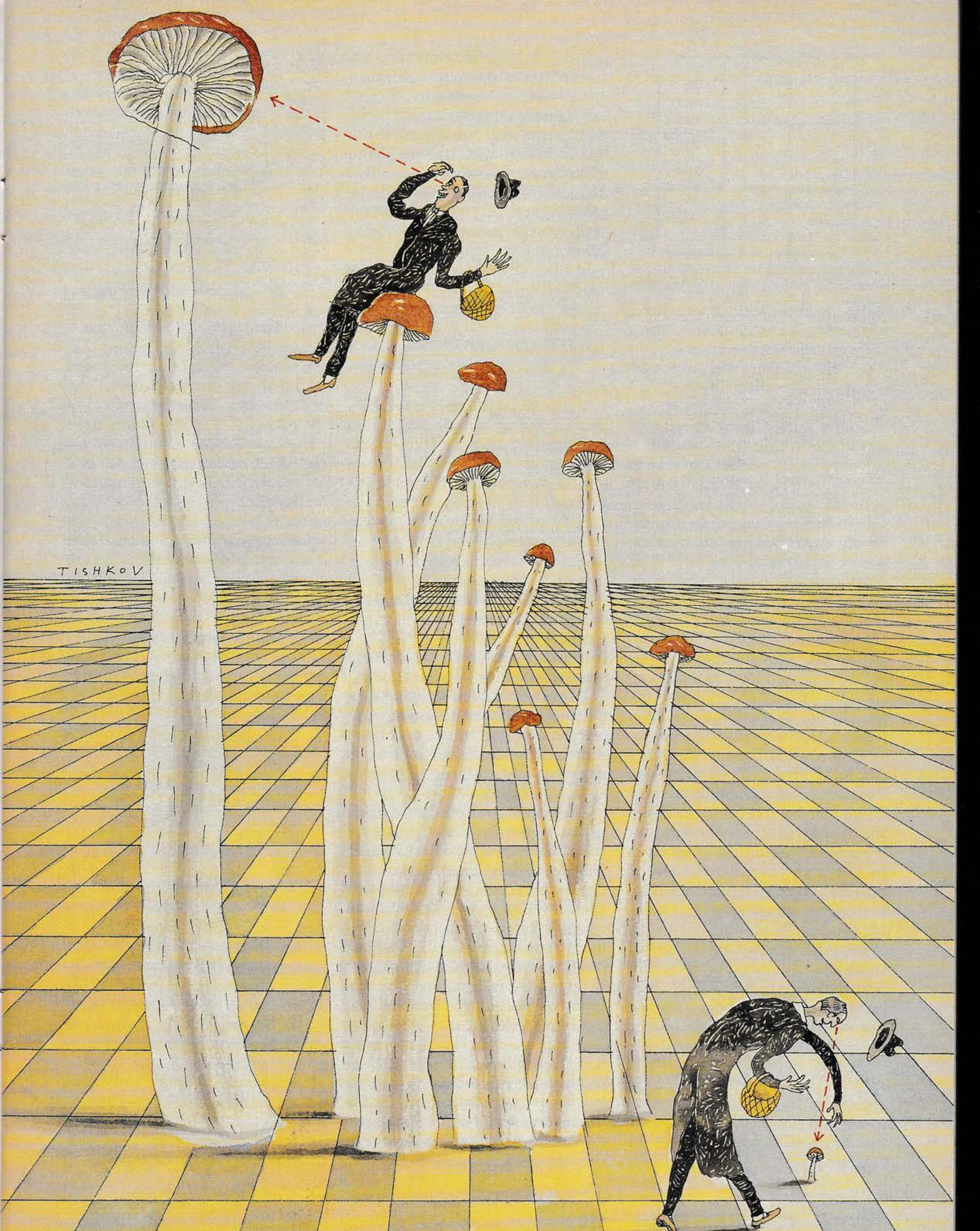


Figure 1

Problem 3. *The squares of an infinite chessboard are marked by natural numbers in such a way that each number is equal to the arithmetic mean of the four adjacent numbers—the upper, lower, right, and left ones. Prove that all the numbers written on the chessboard are equal.*

The extremity rule is helpful here in one of its variations: "Consider the smallest number!" Among the numbers written on the chessboard there's the smallest one. This is easy to prove. Let k be one of the numbers. If 1 is one of the numbers on the chessboard, then 1 is the minimum number (since there are no natural numbers less than 1). If 1 isn't on the chessboard, see whether 2 is on it. If it is, then 2 is the smallest number. Otherwise, look for 3, and so on. In no more than k steps the smallest number will be found. Denote it by m and the square in which it's written by P. Denote the numbers in the adjacent squares by a, b, c, and d (fig. 2). According to our condition, $m = (a + b + c + d)/4$, or $a + b + c + d = 4m$. Because

Art by Leonid Tishkov


TISHKOV

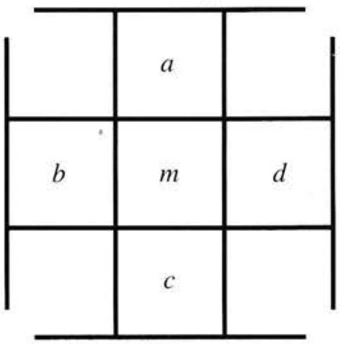


Figure 2

of the choice of m we have $a \geq m$, $b \geq m$, $c \geq m$, $d \geq m$. If at least one of these inequalities is a strict one, we get $a + b + c + d > 4m$, contradicting the assumption. This means $a = b = c = d = m$.

So if a square of the chessboard contains the smallest number m, then the four adjacent squares also contain m. By moving to an adjacent square again and again, we can travel from square P to any other square on the chessboard. Therefore, all the numbers on the chessboard are equal to m.

Problem 4. *A number of rooks are placed on an* n *by* n *chessboard so that the following condition is observed: if a square of the chessboard is free, the total number of rooks standing on the horizontal and vertical lines crossing this square is not less than* n. *Prove that there are at least* $n^2/2$ *rooks on the chessboard.*

This is a tough problem. But a skillful application of the extremity rule dramatically simplifies the situation. Consider a line on the chessboard (which may be either vertical or horizontal) with the least number of rooks on it. There may be several such lines "equally loaded" with rooks. In that case, choose any one of them. Let this line be a horizontal one (or else rotate the chessboard 90 degrees). Denote the number of rooks on this horizontal line by k. If $k \geq n/2$, there are no fewer than $n/2$ rooks on each of n horizontals, and the chessboard contains at least $n^2/2$ rooks.

Now let k be less than $n/2$. There are $n - k$ unoccupied squares on the chosen horizontal, and each vertical line passing through a free square on that line contains, according to the statement of the problem, no less than $n - k$ rooks, so that all $n - k$ vertical lines contain no fewer than $(n - k)^2$ rooks. The remaining k verticals contain no fewer than k rooks each (because of the choice of the number k). So the total number of rooks on the chessboard is no less than $(n - k)^2 + k^2$. It remains to be proved that $(n - k)^2 + k^2 \geq n^2/2$. This can be done in various ways—for instance,

$$\left[(n-k)^2+k^2\right] - \frac{n^2}{2} = \frac{n^2}{2} - 2nk + 2k^2$$

$$= 2\left(\frac{n^2}{4} - nk + k^2\right)$$

$$= 2\left(\frac{n}{2} - k\right)^2 \geq 0.$$

If n is an even number, there's a pattern that satisfies the condition and contains precisely $n^2/2$ rooks: the rooks are all standing on black squares (or all on white ones). If n is odd, it's impossible to position $n^2/2$ rooks in such a way as to satisfy the statement of the problem since $n^2/2$ isn't an integer, but there *is* a pattern containing $(n^2 + 1)/2$ rooks: one rook is placed in one of the corner squares and the others are placed on squares of the same color.

The next problem is also solved by the extremity rule.

Problem 5. *A number of points are given in a plane, not all contained in one straight line. Prove that there exists a circle passing through three of them that contains none of the given points inside.*

Drawing all possible circles through triples of given points, we get a set of circles (some of which may coincide). We have to prove that at least one of them doesn't encircle any of the given points. The extremity rule tells us to consider the smallest circle, but figure 3 shows that one of the given points may remain inside such a circle. Although we can get a solution this way (see exercise 2 below), we'll do something different. Let's try to solve a simpler problem first: let's find a circle passing through *two* of the given points that doesn't contain any of the given points. Measure the distances between each pair of points and use the extremity rule in the form "Consider the smallest one!"—that is, take

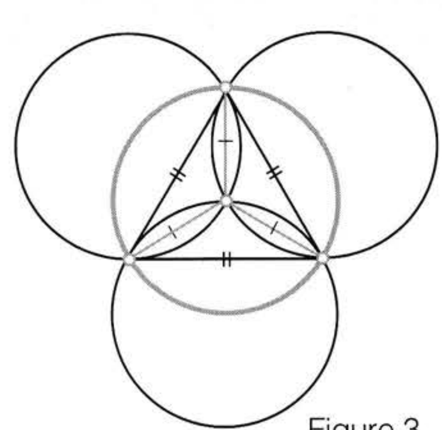


Figure 3

a pair of points A and B that are closest to each other. It's easy to show that the circle constructed with the interval AB as a diameter satisfies the following condition: the distance to any of the other $(n - 2)$ given points from either A or B is no less than AB, so each of the remaining $(n - 2)$ points is located outside the circle. Now draw circles through A, B, and each of the other $(n - 2)$ points and choose the smallest among them (prompted again by the extremity rule). Let it be the circle passing through A, B, and C. This is the circle we're looking for, since any circle going through A, B, and a point C' lying inside the shaded "sickle" (fig. 4) is smaller than the circle passing through A, B, and C (prove it yourself).

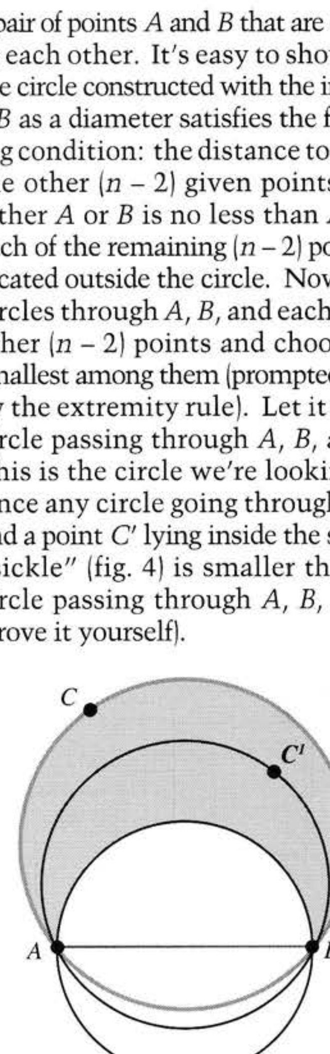


Figure 4

Problem 6. *You are given* n *lines* (n $\geq$ *3) in a plane, no two of which are parallel and no three of which have a point in common. The lines cut the plane into several regions. Prove that for any line at least one of the regions adjacent to it is a triangle.*

Let l_1 be one of the lines. Applying the extremity rule, choose from among the intersection points a point P lying at the shortest distance from l_1.

Denote the lines intersecting at P by l_2 and l_3 and consider the triangle formed by l_1, l_2, and l_3 (fig. 5). No other line intersects this triangle (otherwise there would be an intersection point Q on either l_2 or l_3 that is closer to l_1 than P is).

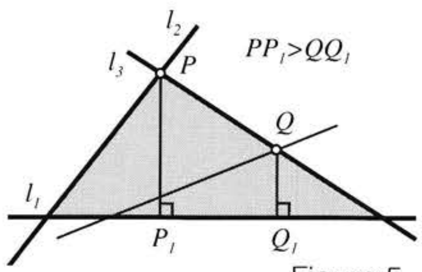


Figure 5

Problem 7. *Prove that there are no natural numbers* x, y, z, u *that satisfy the equation* $x^2 + y^2 = 3(z^2 + u^2)$.

Let's assume that the equation can be solved. Consider a solution for which $x^2 + y^2$ takes the least value (if there are several such sets of four numbers, take any of them). Denote the four numbers by a, b, c, d. The equation $a^2 + b^2 = 3(c^2 + d^2)$ implies that $a^2 + b^2$ is a multiple of 3. But $a^2 + b^2$ is divisible by 3 if and only if both a and b are divisible by 3 because the square of a number that isn't a multiple of 3 always leaves a remainder of 1 when divided by three.

Consequently, $a = 3m$, $b = 3n$, so that

$$a^2 + b^2 = 9m^2 + 9n^2 = 3(c^2 + d^2).$$

Dividing the last equality by three, we get

$$c^2 + d^2 = 3(m^2 + n^2).$$

So we've found four natural numbers c, d, m, n that satisfy the given equation such that

$$c^2 + d^2 < a^2 + b^2,$$

contradicting the choice of a, b, c, d.

Problem 8. *You are given* n *lines (*n *≥ 3) in a plane. Any two lines intersect, and at least three of the lines pass through each intersection point. Prove that all the lines intersect at one and the same point.*

Let l be one of the lines. If not all the lines intersect at one point, then there's at least one intersection point that doesn't lie on l. Choose from among such points the point M closest to l. There are at least three lines l_1, l_2, l_3 that pass through M. These lines intersect l at points A_1, A_2, A_3. Let A_2 lie between A_1 and A_3 (fig. 6). The statement of the problem implies that besides l and l_2 at least one more line passes through A_2. It has to intersect one of the intervals MA_1 or MA_3 at some point N. Then N lies closer to l than M does, which contradicts the choice of M.

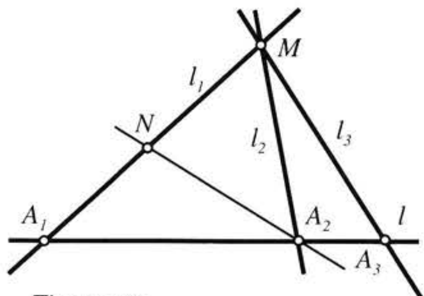


Figure 6

A further development of the extremity rule is the "ordering rule," which reads: "Arrange the elements of your set any old way—in increasing, decreasing, or any other order!"

Problem 9. *Seven mushroom gatherers collected 100 mushrooms, but no two of them picked the same number of mushrooms. Prove that there are three people who together picked at least 50 mushrooms.*

Write down the people's names, putting the most productive gatherer first and working down the list to the least productive. It's clear that we should consider the persons with the three highest ratings since they gathered more mushrooms than any other group of three. Let's prove that their joint total is at least 50 mushrooms. If the third person on the list picked 16 mushrooms or more, then the second has at least 17 and the first at least 18 mushrooms. Altogether they collected at least 16 + 17 + 18 = 51 mushrooms. If the person in third place collected no more than 15 mushrooms, the rest of the gatherers (in positions four through seven) collected at most 14 + 13 + 12 + 11 = 50 mushrooms, which again leaves at least 50 mushrooms for the first three.

Now it's time for you to try your hand at "going to extremes"!

Exercises

1. There are n^2 integers arranged in an n by n table in such a way that for each zero the sum of the numbers in the corresponding row and column is at least n. Prove that the sum of all n numbers is at least $n^2/2$.

2. (a) There is a point D inside a circle circumscribed around a triangle ABC such that the radius of the circumcircle is not greater than the radii of the circles ABD, BCD, CAD. Prove that the triangle ABC is acute, D is its orthocenter (the common point of its altitudes), and the radii of the four circles are equal.

(b) Find another solution to problem 5 in this article, starting with the choice of the smallest circle passing through three of the given points.

3. You are given n points ($n \geq 3$) in a plane. Each line passing through a pair of the points contains at least one more given point. Prove that all the n points lie on a single line.

4. Find all triples of natural numbers x, y, z such that $x + y + z = xyz$.

5. Prove that in any tetrahedron there is an edge forming acute angles with all the edges emerging from its end points.

6. A number of checkers are placed on a checkerboard. A move can take any of them to one of the four adjacent squares (rather than along the diagonal, as is usually the case). After several moves all the checkers return to their initial positions and each of them has been to all the squares of the checkerboard exactly once. Prove that there was a moment when none of the checkers was positioned on its initial square.

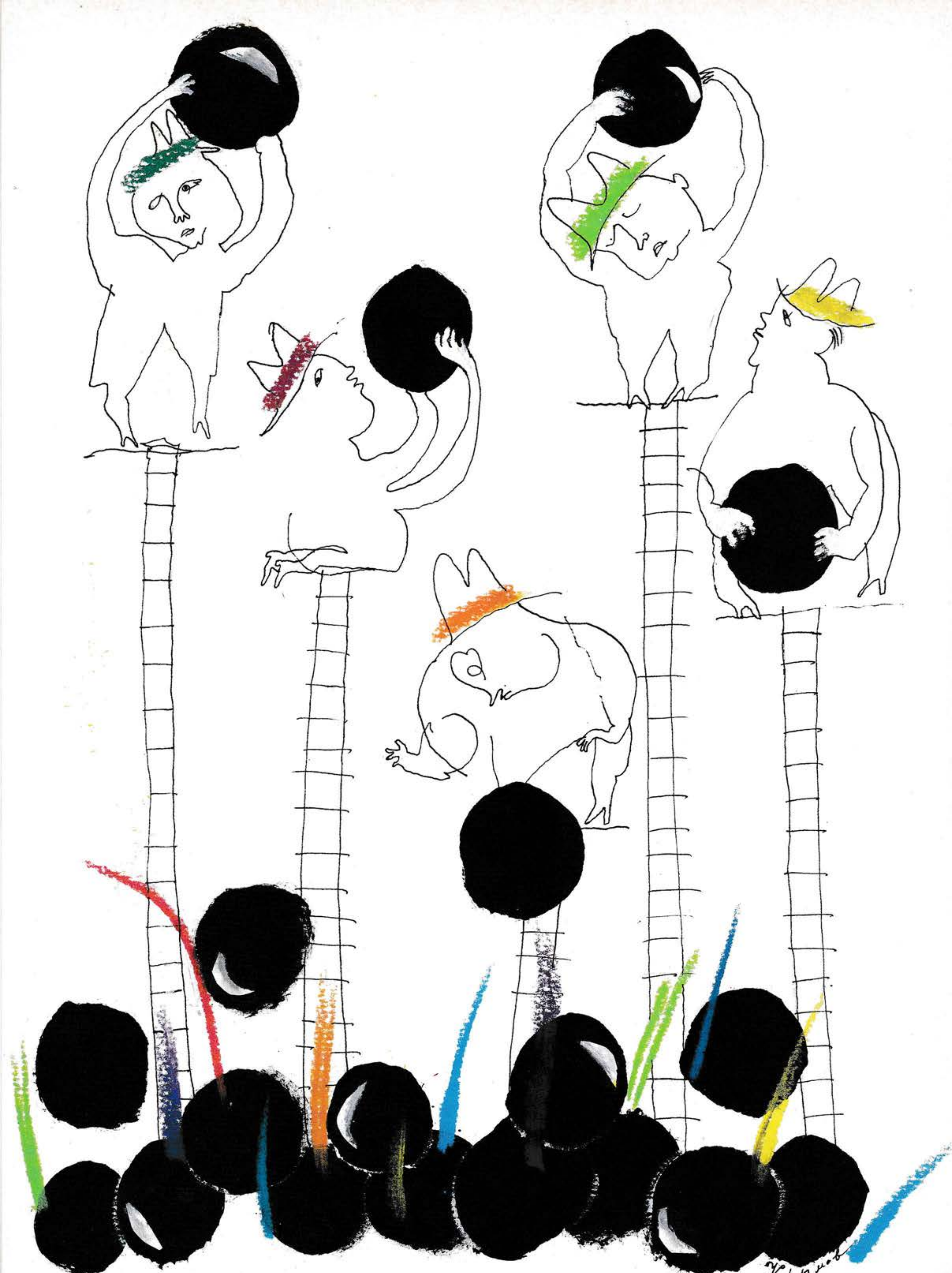
7. Solve the following system of equations:

$$\begin{cases} x_1 + x_2 = x_3^2, \\ x_2 + x_3 = x_4^2, \\ x_3 + x_4 = x_5^2, \\ x_4 + x_5 = x_1^2, \\ x_5 + x_1 = x_2^2. \end{cases}$$

8. A cube is broken down into smaller cubes. Prove that at least two equal cubes emerge from this process.

9. In a certain country all distances between airports are different. An airplane took off from each airport and headed for the nearest one. Prove (a) that no more than 5 airplanes arrived at each airport; (b) that if the number of airports is odd, then there was an airport at which none of the airplanes landed.

HINTS AND SOLUTIONS ON PAGE 60



Fa

Su

Fo

Na

Ad

Ci

Quantum, 1742 Connecticut Avenue NW, Washington, DC 20009
Please allow 4 - 6 weeks for delivery of first issue

Lightning in a crystal

How the LED grew up to be a laser

by Yury R. Nosov

IF YOU ASK AN EXPERT IN electronics—an engineer, a scientist, or the head of an electronics company—what shows the most promise in this area, eight out of ten will answer: electronic optics.

The old idea of using light signals for information transfer instead of electricity (as is the case in traditional microelectronics) turned out to be a very fruitful one. The marriage of electronics and optics may improve the operational parameters of computer equipment: operating speed would be increased by a factor of hundreds or thousands, and it would be more reliable, noise-free, and miniature.

This was already well understood in the 1960s. So why do most of the potential advances envisaged here still await realization? Well, there are quite a number of hurdles to overcome. In order to "harness" light we have to be able to handle it as easily as electric current. We must be able to amplify and transform light signals, transmit them from one location to another without significant loss, develop recording and storage devices. But first of all, we have to learn how to generate them. Whatever the importance of the other elements of an electronic optics system, the basic component is the light generator. It's the alpha and the omega of the system. Of course, an ordinary light bulb is of no use here. The source must be at least as small, reliable, and long-lasting as conventional transistors and integrated circuits.

Art by Dmitry Krymov

The natural place to look for a solution was semiconductor technology.

The diode that glowed

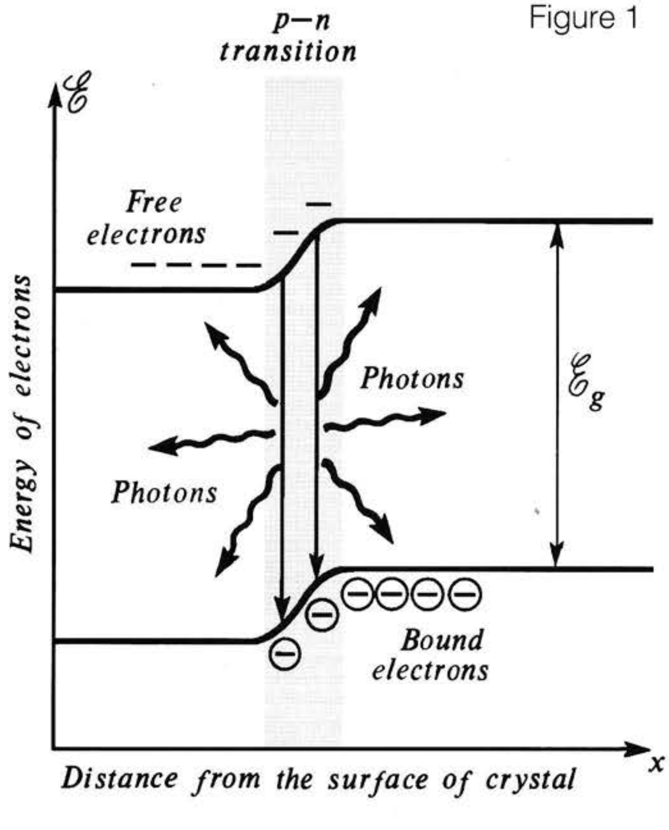
Let's briefly review the situation in this area as it was thirty years ago. At that time the main concern of semiconductor science was to satisfy the needs of transistorized instrumentation. The whole future of electronics seemed to depend totally on their development. The first transistors were made of germanium, but it was clear that better results could be obtained by using silicon or the then new semiconductor gallium arsenide (GaAs). The "silicon way" quickly achieved success and since 1960 it has constituted the mainstream of microelectronics.

The gallium arsenide transistors, however, persistently refused to appear. Millions spent on developing perfect GaAs monocrystals could almost have been written off as a complete loss, but . . . sometimes a loss turns into a real find. And so it was in the gallium arsenide "deadlock." Hope still glimmered, and then it glimmered in the literal sense of the word.

In 1956 it was discovered that electric current passing through GaAs diodes causes them to emit light! So the first light-emitting diode (LED) appeared. Physicists and engineers started to scrutinize the effect. It was immediately established that the semiconductor crystal of an LED did not heat up, which meant that the radiation was caused by luminescence, the phenomenon known as "cold radiation."

The operating principle of the light-emitting diode was quickly explained. The GaAs crystal of the diode isn't homogeneous. Its different regions vary in their properties. By introducing different kinds of impurities, you can enrich one of the crystal's halves (the left one in figure 1) with mobile electrons and deprive the other half of them. The energy of the electrons is higher on the right and drops sharply at the boundary, called the "*p–n* transition" (which plays an exceptionally important role in semiconductor electronics). This energy barrier "prohib-

Figure 1



its" electrons from crossing the p–n transition from left to right "at will." But if an external voltage is applied to the crystal, the barrier lowers a bit and some of the electrons are injected into the right half—that is, they're injected from the emitter to the base. It's in the vicinity of the p–n transition that our phenomenon takes place. After getting to the right side, the electrons fall from the mobile state into the bound one and lose the acquired energy. The lost energy may be emitted as a quantum of radiation, the photon. In this way a light-emitting diode transforms the energy of electric current into radiation energy!

It's as if a heavy stone were first rolled to the top of a mountain and then fell into an abyss. Hitting the rocky bottom of the abyss, the stone produces a spark. The height of the mountain determines the color of the spark: the greater the energy gap E_g between the mobile and bound states of an electron, the greater the energy of the quantum and the shorter the wavelength of the emitted light. With an increase of E_g, the color of the radiation shifts to the blue–violet end of the spectrum. When a sufficiently strong current passes through the diode, the "stone fall" becomes so intense that separate "sparks" merge into a continuous glow.

Of course, a metaphor never coincides perfectly with the phenomenon it's meant to clarify. The true quantum picture of electron transition that causes photon generation can't be reduced to any other process. Actually, once it's understood, the picture becomes as simple and clear as any other physical process (in any case, no more complicated than the fall of a real stone in real mountains).

A little color, please . . .

In the years under discussion here, the theory of luminescence was already well developed, which made it possible not only to calculate the processes in crystals with known properties but also to predict new effects. And there certainly was something to calculate and predict here.

The problem was, the first GaAs light-emitting diodes radiated in the infrared band of invisible wavelengths. Of course, infrared light can be registered by various photodetectors and has numerous technical applications. Still, it seemed like a nice idea to have diodes emit light the human eye could see, since the eye is our main instrument for apprehending the world. Why not light-emitting diodes that glow in all the colors of the rainbow, bright and clear? To achieve this one had to find semiconductors with energy gaps greater than that of gallium arsenide. As usual when the physical mechanisms are understood and the problem is precisely formulated, the means of solving it were readily found.

Soon[1] no one was surprised to see gallium phosphide LEDs emitting intense red or green light, depending on the type of impurities introduced into the crystal. A triple compound of gallium, arsenic, and phosphorous made it possible to obtain any wavelength from dark-red to orange or almost yellow. Silicon carbide emitted yellow-green and pale blue light, though very faintly. Only blue light, like Maeterlinck's evasive blue bird, couldn't be captured by the scientists. The brightest were the red light-emitting diodes, so it was under a "crimson sail" that electronic optics sailed into technology and into our daily life.

Numerous instruments use arrays of LEDs positioned in a specific order on a panel. By selectively turning on appropriate light-emitting diodes in the array, one can generate a digit, letter, or graph. This naturally led to the following thought: why cut the semiconductor plate into individual little crystals and then bring them all back together again in an array of LEDs? In response, character-synthesizing indicators appeared on the scene—plastic casings enclosing several crystals, or a single one, with several points that light up independently of one another.

Light-emitting diodes and numeric displays began to be produced commercially at the end of the 1960s and were quickly put to use in a broad range of applications. Worldwide production approaches 10 billion pieces a year! These bright-red glowworms and numerals can be found in electronic watches and pocket calculators, on laboratory and industrial instrument panels, in the keys and buttons of radio and electronic equipment, in the cockpits of airplanes and submarines . . . just about everywhere.

It's true, the use of light-emitting diodes is restricted by the short distance required between the display and the user's eye. But there is already talk about using superintense LEDs in automobiles as taillights. Of course, it'll be a long time before light-emitting diodes light our homes—although, given the rapid advance of technology, we shouldn't be too rash in our predictions.

Unfit for computer duty

It's time to catch our breath and sum up. Everything I've talked about so far has to do with the use of light-emitting diodes to display information, numerical or otherwise. They turn the electrical impulses of computer-generated information into a visually perceived image that is quickly and easily apprehended by the user. Undoubtedly, such devices are of the utmost importance. But this is only one area in which electronic optics can (and should) help information science. What about processing, transmitting, and storing information? Can a light-emitting diode be of any help in these areas? Alas, the bright rainbow of colors seems to fade here . . .

The first stumbling block, as I already mentioned, is the low intensity of the light emitted by LEDs. Even if it can be perceived by the human eye, it's not always detected by a light-sensitive device (especially if it's located at a distance from the LED). Another problem is that the radiation of light-emitting diodes isn't monochromatic. We'll look at the quantitative side of the matter later; the crucial point here is that the emission bandwidths are too broad for use in many electronic optics devices.

Finally, and most important of all,

[1]It's easy to say "soon" nowadays, but that "soon" meant almost a decade of elaborate analytical studies, hard work on synthesizing superpure semiconductors, development of new equipment and technologies . . .

light-emitting diodes radiate almost homogeneously in all directions. It's impossible to concentrate its energy in a sharply focused beam. They're of no use in performing the simplest task in electronics—sending a signal from point A to point B. The greater part of the emitted energy is not only uselessly squandered, it irradiates the surrounding space and may even jam other sources. The light-emitting diode is a careless chatterbox incapable of keeping a secret. It's obviously not suitable for use in information science, where all operational features must be precise and trustworthy and where each bit of information must use only the amount of energy it actually needs.

Fortunately, there's a good alternative to the light-emitting diode as a radiating source. It's the laser, which emits intense, almost monochromatic, very focused light. Let's digress for a moment and look at the quantitative side of laser operation.

Its directionality is characterized by a solid angle α containing the beams generated by the source; if the beams diverge, symmetrically deviating from a certain axis (the direction of emission), this divergence is measured in radians, or degrees and minutes as in conventional plane geometry.

There are no strictly monochromatic waves in nature. Any light source always has some range of color, or wavelength. Quantitatively, this range is described by the notion of monochromaticity, which is defined as the ratio of the bandwidth of the wavelengths of the generated radiation $\Delta\lambda$ to the wavelength λ_0 of the center of the band: the smaller $\Delta\lambda/\lambda_0$, the better the operational features of the laser. A good example is the typical helium–neon laser, for which $\alpha < 1'$ and $\Delta\lambda/\lambda_0 < 0.000001$.

Such a light source would be quite suitable for computational electronic optics were it not for the fact that the helium–neon laser has a glass discharge tube almost half a meter long and a high-voltage power supply unit weighing several kilograms. Now place beside it an integrated circuit the size of a postage stamp containing about a million transistors and requiring only 5 volts of power. Are these two units compatible? Obviously not! And, indeed, numerous attempts to use conventional lasers in microelectronic computer devices came to naught. As the Russian saying has it, "You can't hitch a bull and a doe to the same wagon."

Obviously, there's only one way to make lasers and microcircuits compatible: make both of them semiconductors.

The birth of a new laser

The story of the semiconductor diode laser is typical of scientific discoveries in the 20th century. After the solid-body (ruby) and gas (helium–neon) lasers were almost simultaneously invented in 1960, scientists predicted that a semiconductor laser could be made as well. It was expected that, like other semiconductor devices, it would be small, cheap, durable, resistant to outside influences, flexible in its parameters, and useful in a wide range of applications. It was quite a challenge to create such a device, and leading laboratories throughout the world vied with each other to catch this "beautiful butterfly." Theoreticians were able to describe the desired quantum structure of the crystals, thus narrowing the list of potential candidates. The butterfly's fate was sealed. On the eve of 1963, almost simultaneously, the first semiconductor lasers were created in the US and USSR.

The pioneering semiconductor was again gallium arsenide. The only difference was that it contained more impurity elements, which created a greater number of free electrons. After the p–n transition is achieved on a sheet of gallium arsenide, the large piece is broken with a scalpel into tiny rectangular crystals. The sheet splits strictly along its crystallographic planes, so that the opposite facets of the crystals are parallel and highly reflective. These two mirrors form a resonator, which is necessary for the laser feedback effect. The crystal's lower facet is then soldered to a massive copper substrate (to increase heat transfer), and a second, thinner electrode is connected to its upper facet.

When an electric current is applied, the crystal starts to emit infrared light as a light-emitting diode does—weakly and in all directions. But as soon as the current reaches a certain value (called the threshold current), the picture changes dramatically: the radiation power suddenly jumps and intense light is emitted from the strips on the side facets where the p–n transition plane intersects the resonator's facets. A spectral analysis of the radiation revealed that this phenomenon resulted in the substantial narrowing of the band of generated wavelengths. There was no longer any doubt—it was a laser!

The operating principle was explained without much delay. As with light-emitting diodes, the external voltage applied to the crystal "drives" electrons up the "energy barrier," except that this barrier is a bit higher and the number of electrons much greater than in a light-emitting diode. The electrons gather near the p–n transition, creating a so-called active zone. "Falling from the barrier," they give rise to quanta of radiation (that is, photons). It's at this point that the analogy with light-emitting diodes breaks down. The light wave propagating along the p–n transition plane is reflected off the mirrored faces of the crystal and, repeatedly passing through the active zone, forces more electrons to "drop from the barrier." It turns out that a huge quantity of electrons simultaneously and identically undergo the prescribed quantum transition (shown by the two-headed arrow in figure 1). As a result, the laser beam has a high degree of monochromaticity and a specifically determined polarization. Because of the way it is created, such radiation is called stimulated or induced radiation, whereas the radiation of a light-emitting diode is spontaneous (random in its direction, polarization, and, to a certain extent, wavelength).

Another problem surmounted

The new device aroused great interest. It seemed obvious that fundamental changes were about to begin in electronic optics. But time passed and there was no serious application of the new laser. The initial euphoria gave way to bewildered disappointment.

The laser operated only at low, "nitrogen" temperatures (–196°C) and only if the external current was supplied in short, infrequent impulses. Even then, its lifetime was exceedingly brief—several dozen hours at most. If it was operated in any other manner, it would immediately overheat and fail completely. It also turned out that its degree of monochromaticity was only marginally better than that of light-emitting diodes, by a factor of merely 10 to 20 ($\Delta\lambda/\lambda_0 \cong 0.005$ compared to 0.05 for an LED), and it was still far worse than that of a gas laser (by a factor of thousands). In its directionality ($\alpha \cong 30°$), the semiconductor laser seemed like just an improved light-emitting diode. "What kind of laser is *this*?!" frustrated electronic optics engineers might have exclaimed. Gradually experiments with the new device gave way to renewed speculation about the sunny prospects of "ideal" electronic optics.

The worst part of it was that the semiconductor laser's drawbacks were provided with a rigorous and apparently insurmountable theoretical basis. The electrons injected into the thin active zone weren't willing to stay there but scattered all over the crystal. The same thing happened with the light wave. Instead of contributing to the laser effect, the lost electrons and quanta only caused useless overheating of the crystal. What could make the mobile charge carriers (electrons) and light radiation (photons) stay in a specific area of the crystal? You can't put a shield or a mirror inside what is, after all, a monocrystal, in which all the atoms are positioned in an ideal predetermined order. When researchers began their chase after the "beautiful butterfly," somehow none of this ever came up. Now it began to look as if the butterfly was destined for a jar of formaldehyde in a museum of physics curiosities.

But the nimble human mind once again emerged triumphant. And what did it come up with? Heterostructures. If some of the gallium atoms in a gallium arsenide crystal are replaced with aluminum atoms, the structure of the crystal lattice isn't changed because the atoms of the two elements are so similar in their physical properties. But this results in the creation of a new semiconductor, gallium–aluminum arsenide, with a larger energy gap than that of pure gallium arsenide. The area between the two semiconductors inside a single monocrystal is called a heterojunction. In addition to the energy barrier it also includes an optical barrier because the two semiconductors have different refractive indexes. The active zone has a higher refractive index and, sandwiched between heteroboundaries, makes an ideal trap for electrons as well as a waveguide for light beams.

Further refinements

"Her Majesty Technology" took over from here. In virtually no time at all scientists learned how to set up pairs of heterojunctions parallel to each other inside a monocrystal and separated by the fantastically precise distance of a few atomic layers. The threshold current was lowered to several dozen milliamperes, the upper limit for the laser's operating temperature reached 100°C, and accelerated aging tests showed that the new laser diodes should last several decades. And so the renaissance of the semiconductor laser began. The industry was flooded with inventions and discoveries. You want to lower the threshold current? Okay—sandwich the active zone between heterojunctions not only above and below but also between two other heterojunctions on the sides. The microfilament of the active semiconductor can then be excited by a current of only 1 mA! You want to narrow the spectrum of emitted wavelengths? Just give one of the heteroboundaries a wavy shape. The resonator's selectivity increases sharply, and the degree of monochromaticity reaches values typical for gas lasers. To increase monochromaticity even more, use a structure with two "coupled" crystals (fig. 2).

Heterolasers are now manufactured by the millions each year, and their use has displaced our entrenched notions in many areas of technology. A fiber optic communication line is capable of transmitting all 30 volumes of the Great Soviet Encyclopedia in digital form in a single second! A desktop CD ROM (read-only memory) unit can "hold" more than a million books. Optical integrated circuits using heterolasers are being developed with a view to creating supercomputers that will be thousands of times as productive as the current models.

"Lightning"?

The title of this article included the word "lightning." Why? Because the current density in light-emitting diodes and semiconductor diode lasers may be several times (and sometimes scores of times) greater than that in a lightning discharge. Scores of "microlightning" discharges flash in our crystals, but they're under human control; instead of bringing destruction, they breathe life into electronic optics.

This area is now on the front line of solid-state physics. Equipment has already been developed that's capable of growing multilayered semiconductor structures whose composition and properties vary in each monatomic layer. Such structures make it possible to control almost every electron in the lattice: one can be "planted" in a "quantum well," another can be "walled up" inside a "quantum box," another can be set free to "wander" over the whole crystal . . . But it's much more thrilling to carry out such projects with your own hands than to write about them. If you don't believe me—try it yourself!

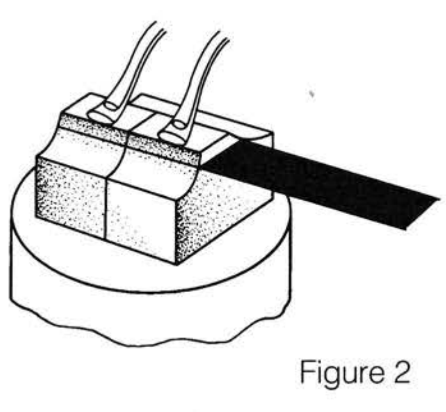


Figure 2

Physics for fools

Need we say "Kids, don't try this at home . . ."?!

by V.F. Yakovlev

TO TELL YOU THE TRUTH, I had a difficult time at the university. That's probably why I'm especially infuriated by all those remarks in textbooks on physics and mathematics that go: "It is well known that . . .," "Simple calculation readily yields . . .," "One can easily see . . . ." Where is it well known from? Why is the calculation simple? Usually, for me, it was very difficult and sometimes even impossible! Such remarks not only mislead students but also contribute very effectively to the development of an inferiority complex.

Isn't their purpose really to mask the authors' incompetence? I mean, good students will find their way through a text even if it's full of mistakes. I'm sure that if someone had forced me to write a textbook for differential analysis without any access to lecture notes or books on the subject, Euclid or Archimedes could have understood what I was trying to say. It's very easy to write textbooks for clever people. Even fools can handle that. To write a textbook for people of more middling talent—now *that's* a challenge. But what a noble task it is!

Just imagine: "Quantum mechanics for the feebleminded," "Differential calculus for utter fools." Books like that would surely top any best-seller list! You'd have to litter the text with remarks like "One can barely derive from this . . .," "It is very difficult to understand that . . .," and so on. The readers who got through all this would glow with enthusiasm.

Now you can easily understand my triumph when, reading the book *Matter, Earth, Heaven* by the famous physicist George Gamow (published in the US in 1959), I came across a reference to another book published in 1908 in St. Petersburg (now Leningrad) under the title (according to Gamow) *Physics for Fools*. I was on the verge of jumping out of my chair and shouting *"Eureka!"* in the library reading room. Unfortunately, my joy was shortlived since Gamow failed to mention the author's name.

After a long and tedious search in many catalogues and reference works, I finally found the book. Its title page reads:

Published by
The Society for the Encouragement of Stupidity

New Physics Without Instruments

A Complete Description of Popular Experiments
Easily Performed At Home

The Best Leisure for
Persons Longing for Physics and Astronomy

Compiled from the Latest Sources and Discoveries

by
Sergey Olympov

The author's name was a pseudonym, and only after more searching in libraries did I find his real name: Sergey Maximovich.

Mr. Maximovich was born in St. Petersburg in 1876. In the 1930s he was still living in Leningrad and was employed at the State Institute of Geodesics and Cartography, working on aerial photography and the physics of light-sensitive materials. He also studied how to measure various characteristics of photographic materials (there's a branch of physics called "sensitometry" or "photographic metrology").

Sergey Maximovich was an extremely ingenious person with artistic talent, as you can see from his drawings that follow and the explanatory notes he wrote for them.

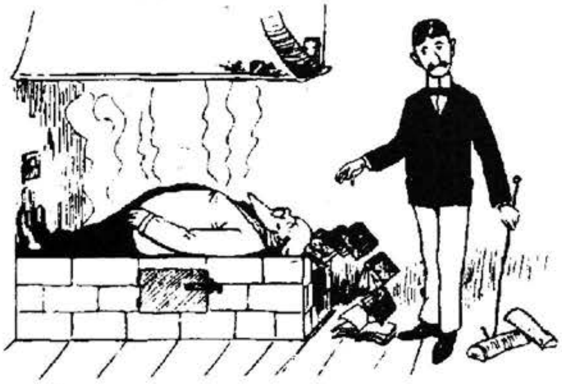


Figure 1
Thermal expansion.

"All bodies get longer when heat is applied. So, for example, rails are always made shorter than necessary. The following experiment can be performed by anyone at home.

"Have an older and more forgiving member of the family lie down on a cold stove so that his feet touch a wall while his head touches a stack of books positioned at the edge of the stove. Make a fire in the stove. Soon you'll see that, as the temperature rises, your relative will stretch and push the stack of books with his head until they fall on the floor. The nature of this phenomenon is pretty obvious, but let's continue the experiment anyway. As the temperature increases even more, the phenomenon enters its next phase: your relative will begin to deform until, finally, he jumps up and runs away. This is an exceptionally convincing demonstration of the law known in the scientific community as the 'transformation of heat into motion.'

"If you immediately place your relative on a red-hot stove, he might enter a spheroidal state and the experiment will fail."

Figure 2
A simple electrical machine.

"Experiments with large electrical machines are not cheap and are not without danger. But anyone can build such a machine from simple household items.

"Ask a friend to sit on a two-gallon bottle and have him hold a fork in his hand. Rub rubber galoshes with a fox coat and bring it toward the fork. Soon you'll hear a characteristic hissing sound, and your friend's nose will start to emit long, bright sparks (which are especially impressive in the dark). With this simple instrument you can carry out all the experiments described in physics textbooks—you can charge a Leyden jar, light a small light bulb, and even run a sewing machine.

"From time to time it's useful to grease your friend and the bottle with a thin layer of warm petroleum jelly."

Figure 3
Electrolysis.

"What could be more durable than a gold watch? The eternal glittering of the noble metal, the motion of its hands as if personifying Time itself—everything suggests stability and perpetuity. But, in actual fact, that's not the way it is. Take a particularly massive gold watch with an anchor escapement and carefully lower it into a big ceramic cup containing a mixture of nitric and hydrochloric acids. By the next morning the watch will have disappeared—only the crystal and dial will remain. They should be taken out, rinsed with water, dried, and stored in absorbant cotton. Don't worry about the watch: in Nature, nothing is lost! Pour the greenish liquid into a bottle with a tightly fitting cork and store it in a dark place.

"Our next volume, *Chemistry Without Instruments*,[1] will include detailed instructions on how to get the watch back. The reader must have already guessed that this is done with the help of that wizard of the 20th century, Electricity. The machine described in the second experiment will be of inestimable help here."[2]

Figure 4
Refraction.

" 'What are you doing?!' your hostess will probably exclaim in horror when she sees you approach a mirror with an uplifted stick. 'You'll break the mirror!'

"Nothing of the sort. From the laws of optics you know that the angle of incidence is related to the angle of refraction by a specific formula—you only have to hit the mirror with the end of the stick such that this relation is satisfied. The stick breaks with no consequences for the mirror—to no small surprise on the part of those present.

"A regular pane of glass (not a mirror) would, of course, have been smashed to bits."

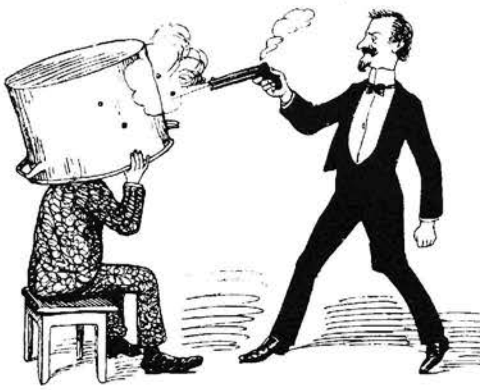


Figure 5
Propagation of sound.

"It's known that sound propagates in a solid body along its surface. A very interesting experiment is based on this phenomenon. Put a thin-walled tin pot on a volunteer's head and then shoot at it with a pistol, machine gun, or mortar. To the person under the pot, the deafening shots will sound like the snapping of your fingers."

Figure 6
Man on eggs. (Toward a physiology of birds)

"This experiment never fails to create a great sensation, especially among those running households. 'Who can hatch a chicken in a quarter of an hour?'

"Throwing a triumphant glance at the silent gathering, go to the chicken coop, find a brood hen, and collect two dozen eggs ready to be hatched. Upon returning to the room, put the eggs upon a chair and under it, unobserved by the others, put a burning kerosene lamp, hiding it under your frock coat. The latent heat[3] of the kerosene rapidly develops the chicks, and they'll soon announce their arrival into this world with their happy cheeping. The only thing you have to be careful about is not to crush the eggs."

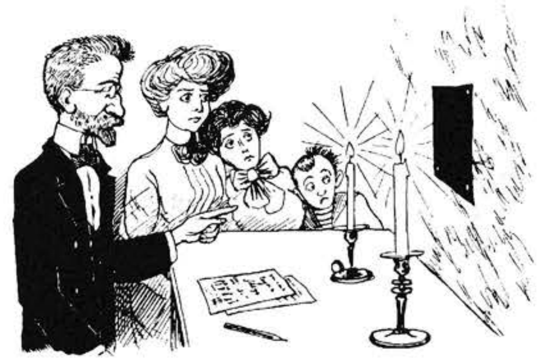


Figure 7
Interference and diffraction.

"Attach a sheet of white paper to a wall (a marble wall is preferable) and illuminate it with a candle. Now light another at, of course, a strictly determined distance from the first. You'd expect there to be more light, but as it turns out the sheet of paper gets darker. This is the phenomenon of interference, which the great Newton called the 'golden key of Providence.' "

[1]Which never appeared.—*Auth.*

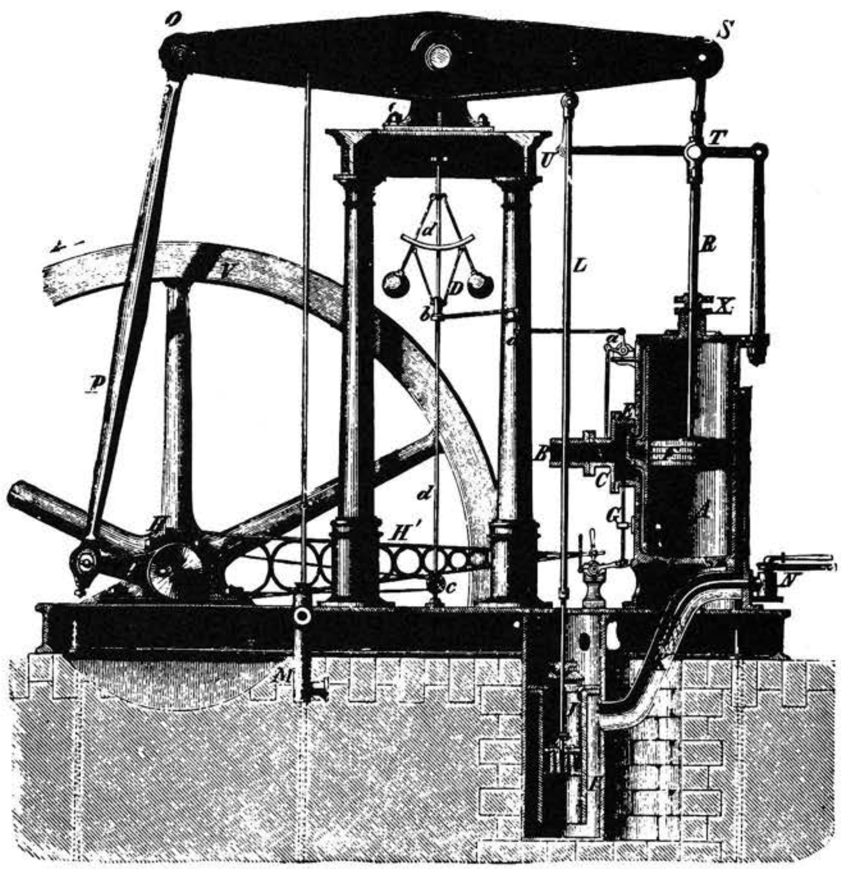
[2]During World War II the French Nobel Prize–winner Jolio-Curie dissolved his gold Nobel medal according to nearly the same technique in order to save it from the Nazis. He kept the bottle with the solution throughout the Nazi occupation, and after the liberation of France used electricity to extract the dissolved gold. Jolio-Curie sent it to Sweden and obtained a new Nobel medal.—*Auth.*

[3]A pun in the original—*skrytaya teplota* could be translated literally as "hidden warmth."—*Ed.*

Making the crooked straight

Inversors and Watt's steam engine

by Yury Solovyov



WHEN STEAM ENGINES AND steam pumps were invented, the theory of articulated mechanisms—systems of rigid links connected by hinges in such a way that the motion of one or more links is transformed into the motion of other links—began its rapid development. For almost a hundred years progress in this area was determined by the problem faced by the English mechanical engineer James Watt (1736–1819) in his attempts to improve his steam engine.

Watt's original design is schematically shown in figure 1. He put a piston inside a steam cylinder, where it could move back and forth. The piston was connected to a rod passing through the top cover of the cylinder. The rod was rigidly fastened to the piston and could, therefore, perform only linear motion. A rocker arm *AF* was attached to a hinge on top of the pillar *OP*, and the hinge *F* coupled the connecting rod *FE* with the rocker *AF*. This connecting rod was, in turn, attached to the crankshaft *QE* by the hinge *E*. A flywheel was attached to the crankshaft.

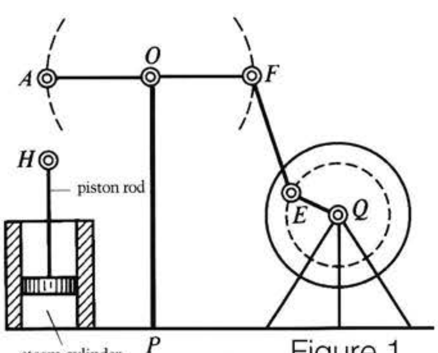


Figure 1

If one could connect the head *H* of the piston rod to the rocker *AF*, the motion of the piston would be directly transformed into rotation of the flywheel. But point *H* is in linear motion whereas point *E* makes a circular arc with radius *OA* and center at point *O*. Consequently, it's impossible to connect points *H* and *A* rigidly without breaking the machine.

So this was Watt's problem: to develop a linearizing mechanism that would drive point *H* along a straight line and point *A* along an arc. Watt solved it by devising an articulated mechanism that drove point *H* along a curve having a small deviation from a straight line.

Many scientists subsequently developed linkages that drove point *H* with a smaller deviation, but it wasn't until the 1860s that a technique for driving point *H* exactly along a straight line was discovered.

Watt's simple linearizing mechanism

Here is Watt's reasoning. Consider two rockers AO and BO' rotating around fixed centers O and O'. If the ends A and B of the rockers AO and BO' are hinged to a segment AB, which Watt called a "shackle," a point of the shackle undergoes a motion very close to linear (fig. 2). In order to

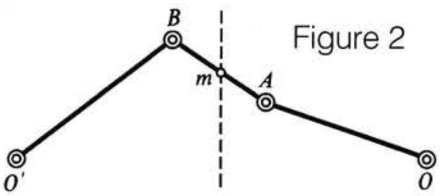


Figure 2

define the most suitable position of the fixed center O' and the length of the rocker BO', consider three positions of the rocker OA (fig. 3): the middle OA and the two extremes OA' and OA''. There should be a point m

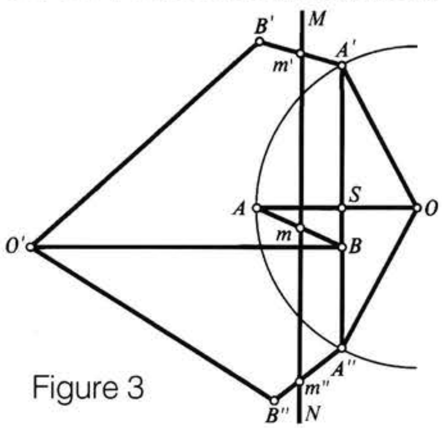


Figure 3

of the shackle that stays on the same straight line MN in all three positions. Watt took as that line the perpendicular to the segment OA passing through the midpoint of the altitude SA of the circular segment $A'AA''$.

Take a shackle ab of fixed length and choose a point m on it (fig. 4). The arcs drawn from points A', A, and A'' with radius am intersect the straight line MN at points m', m, and m'', yielding three positions of point m of the shackle (fig. 3). Plotting on the extensions of $A'm'$, Am, and $A''m''$ segments equal to mb, we get three

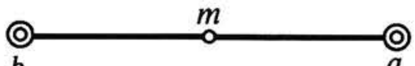


Figure 4

positions at the other end of the shackle denoted by B', B, and B''. These three points define a circumference passing through them. To find its center we drop perpendiculars to the centers of the segments $B'B$ and BB'', which meet at point O'. The center O' defines the length of the second rocker $BO' = B'O' = B''O'$.

Connecting the end b of the shackle with end B of the rocker by a hinge ensures that at least in the middle and at the two extreme positions of the rocker OA the point m of the shackle stays on the straight line MN.

Watt hoped that, moving from m' to m'', point m of the shackle would experience only a small deviation from a straight line. He was right: the trajectory is indeed quite close to a straight line, the precise trajectory being a sixth-order curve looking like an elongated figure eight (fig. 5).

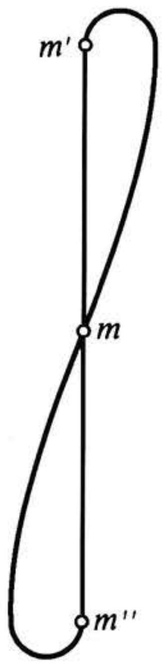


Figure 5

Watt's parallelogram

Watt had one more problem. In addition to the rod driving the piston of the steam cylinder, he had to provide a linear trajectory for another rod attached to the piston of a pump used to fill the condenser (fig. 6). Watt modified his mechanism so that it included two points, each of them moving approximately along a straight line.

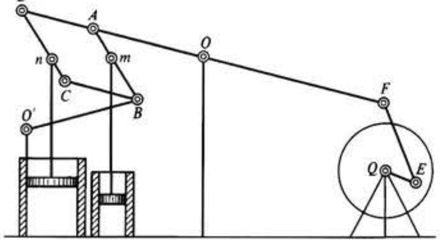


Figure 6

Extend the rod OA (fig. 7) and then complete the parallelogram $ABCD$. Plotting the straight line through points O and m, denote by n the intersection of this line and CD. Point n then

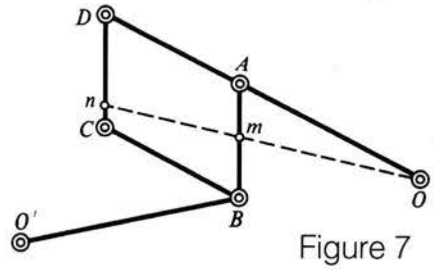


Figure 7

moves along a curve similar to that of point m and, consequently, also has a small deviation from a straight line. Since the steam cylinder is higher than the pump cylinder, Watt attached the head of the steam piston rod at point n, which has a greater amplitude, while the head of the pump's rod was attached at m.

Figure 6 is a schematic drawing of a steam engine with Watt's parallelogram as it appeared in 1784.

Watt himself considered the discovery of linearizing mechanisms his greatest scientific achievement (and not the governor now bearing his name, which is the cornerstone of automatic control theory).

Chebyshev's linearizing mechanism

A number of remarkable linearizing mechanisms were invented by P. Chebyshev, the outstanding Russian mathematician and mechanical engineer. He used his theory of functions with the least deviation from zero, developed in 1858. I won't go into the details of his theory here, but I'll describe one of the most practical Chebyshev mechanisms.

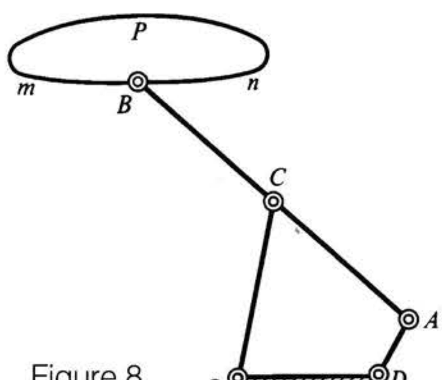


Figure 8

This mechanism (fig. 8) consists of a link AB with a hinge C at its center. The second link OC equal to $AB/2$ is attached to the hinge, so that OC =

$AC = BC$. The other end O of OC is attached to an immobile hinge O. Point A is attached to a third link DA attached to an immobile hinge D. If

$$OD = \frac{OC + CA + AD}{3},\ OC = AC = BC,$$

then point B of the Chebyshev mechanism describes a curve mPn, the portion mn of which has a very small deviation from a straight line. Chebyshev showed that the maximum deviation of the curve fragment mn from a line parallel to OD is given by the formula

$$\delta = \frac{1}{2}\left(\sqrt{\frac{4}{9}(r-a)(2r+a) + \frac{(4a-r)^3 r}{12(2r+a)^2}} - \sqrt{\frac{4}{9}(r-a)(2r+a)}\right),$$

where $r = AB$, $a = 2AD$. It's a very small value indeed. For example, for $AC = OC = BC = 32$ inches (81.3 cm), $OD = 25$ inches (63.5 cm), $DA = 11$ inches (27.9 cm), we get $\delta = 0.032$ inch (0.081 cm).

Rigorous linearizing mechanisms

All the linearizing articulation mechanisms I've described so far are approximate: a straight line is approximated by a suitable curve. The theory of rigorous linearizing mechanisms is based on an important geometrical transformation called "inversion."

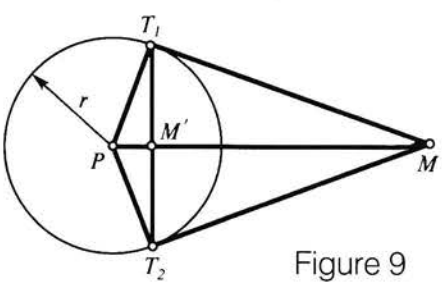


Figure 9

Consider a circle with center P and radius r (fig. 9). Take a point M lying, for example, outside the circle. Plot tangents MT_1 and MT_2 and find the point M' where chord T_1T_2 intersects the line PM. The right triangle PMT_1 yields

$$PM \cdot PM' = r^2. \qquad (1)$$

Conversely, for each point M' lying inside the circle, we can easily find the corresponding outer point M.

Points M and M' lying on the same ray radiating from the center P of a circle of radius r are called *inverses* of each other with respect to this circle if their distances from the center satisfy equation (1). It's obvious that the inverse of a point lying on the circumference coincides with the point and that there is no inverse of the center.

A transformation that produces an inverse M' for each point M is called an *inversion* with respect to the given circle. The circle itself is called the *circle of inversion*, and its center is said to be the *pole of inversion*. The square of its radius is the *degree of inversion*.

An inversion defines (the center P being the sole exception) a one-to-one transformation of the points of the plane. The relation between points and their inverses is a reciprocal one: if M' corresponds to M, then M corresponds to M'. Each point of the circle of inversion is a fixed point.

Let's take a look at one property of inversion that's very important for our purposes.

THEOREM 1. A straight line that does not contain the pole of inversion is mapped by inversion into the circle passing through the pole.

PROOF. Let A be the projection of the pole of inversion on the given line (fig. 10), B an arbitrary point of this line, A' and B' inverses of points A and B. By definition, $PA \cdot PA' = PB \cdot PB'$, or $PA{:}PB = PB'{:}PA'$. This relation ensures that triangles PAB and $PB'A'$ are similar. Since angle PAB is a right one, angle $PB'A'$ is also right. So point B' lies on the circle with diameter PA', which is what we set out to prove.

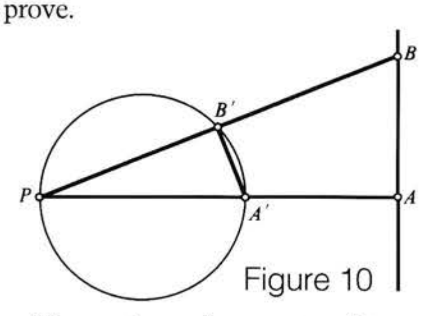


Figure 10

The reciprocal property of inversion immediately yields another assertion.

THEOREM 2. A circle passing through the pole of inversion is mapped by inversion onto a straight line perpendicular to the line through the pole of inversion and the center of the circle.

So, if we could design a mechanism that applies inversion, rotational motion would be transformed precisely into linear motion. Mechanisms that make use of inversion are called "inversors."

Peaucellier's inversor

In 1864 the French engineer A. Peaucellier constructed the following inversor. Four links of the same length are connected by hinges to form a rhombus $ABCD$ (fig. 11). Two other links of equal length BO and DO, but longer than the sides of the rhombus, are attached to opposite vertices of the rhombus. Hinges are put at points B, O, and D.

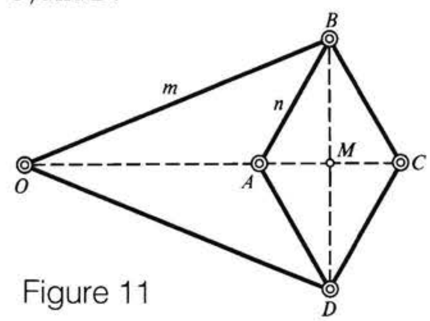


Figure 11

THEOREM 3. For any position of Peaucellier's inversor, the product of lengths AO and OC is a constant value.

PROOF. Denote the length of the long links by m, so that

$$OB = OD = m,$$

and the length of the short links by n, so that

$$AB = BC = CD = DA = n.$$

Now plot the diagonals of the rhombus. One of them will pass through point O (since the vertices of isosceles triangles DOB, DAB, and DCB with a common base BD belong to the same straight line). Let $OA = r$, $OC = \rho$. Considering the triangle OBM, we have

$$BM^2 = m^2 - OM^2. \qquad (2)$$

The triangle BCM yields

$$BM^2 = n^2 - CM^2. \quad (3)$$

Subtracting (3) from (2), we get

$$\begin{aligned} m^2 - n^2 &= OM^2 - CM^2 \\ &= (OM + CM)(OM - CM) \\ &= OC \cdot OA, \end{aligned}$$

or

$$\rho \cdot r = m^2 - n^2,$$

which means that the product

$$\rho \cdot r = OC \cdot OA$$

doesn't change when OC and OA vary, and our proof is done.

Consequently, if point O is fixed and point A moves along a curve, then point C follows the image of that curve under inversion. So if point A moves along a circle passing through the pole of inversion, point C moves along a straight line. (It turns out that Chebyshev's student Lipkin at St. Petersburg University devised this same inversor independently in 1872.)

Let's look at one more inversor before we leave the subject.

Hart's inversor

Soon after the appearance of Peaucellier's inversor an English mathematician and mechanical engineer named Hart constructed an inversor based on an antiparallelogram. A quadrilateral $ABCD$ is called an antiparallelogram (fig. 12) if its opposite sides are equal and two of them (sides AB and CD in fig. 12) intersect each other. The fact that a hinged antiparallelogram produces inversion stems from the following two theorems.

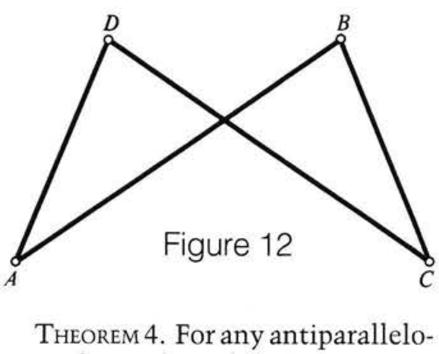


Figure 12

THEOREM 4. For any antiparallelogram the product of its diagonals DB and AC (fig. 13) is a constant value.

PROOF. We'll begin by denoting the relationships

$$AB = DC = m, \; AD = BC = n.$$

Take a segment BL parallel to AD and draw a circular arc with center B and radius BL. This arc passes through point C since

$$BL = DA = BC.$$

Now draw the line AM tangent to this arc. Its square equals the product of the secant and its outer segment. Consequently,

$$AM^2 = AL \cdot AC = DB \cdot AC. \quad (4)$$

Considering the triangle ABM we have

$$\begin{aligned} AM^2 &= AB^2 - BM^2 \\ &= AB^2 - BC^2 \\ &= m^2 - n^2. \end{aligned}$$

Comparing this with (4), we get

$$DB \cdot AC = m^2 - n^2 = \text{constant},$$

as asserted.

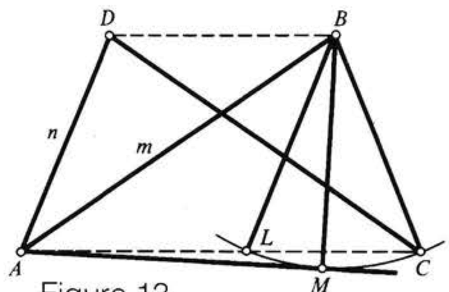


Figure 13

THEOREM 5. Choose any two equal sides of a hinged antiparallelogram and fix a point on a third side. Draw a straight line through this point parallel to the diagonals of the antiparallelogram. The product of the distances from the fixed point to the intersections of the line with the chosen sides remains the same for all positions of the antiparallelogram.

PROOF. In the notations of figure 14 the product in question is one of the following four: $MN \cdot NQ$, $MN \cdot NP$, $PQ \cdot PM$, $PQ \cdot QN$. All these products are evidently equal. It's therefore suf-

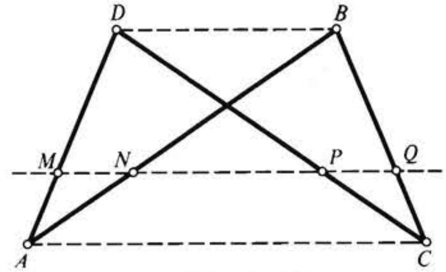


Figure 14

ficient to consider the product $MN \cdot NQ$, where the point N is fixed. The similarity of triangles AMN and ADB yields

$$MN = BD \cdot \frac{AN}{AB},$$

while the similarity of triangles ABC and NBQ implies

$$NQ = AC \cdot \frac{BN}{AB}.$$

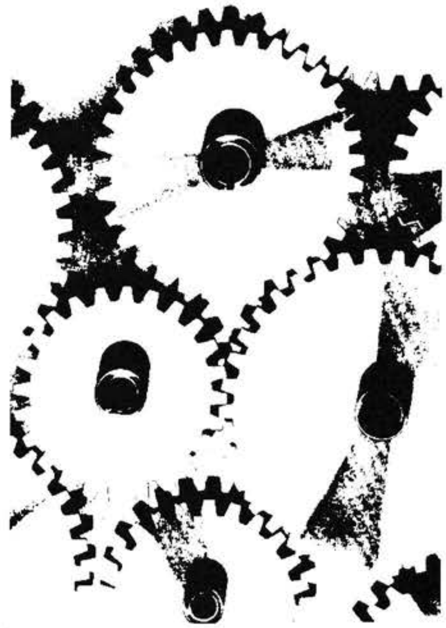
Multiplying these equalities we get

$$MN \cdot NQ = BD \cdot AC \cdot \frac{AN \cdot BN}{AB^2}.$$

The ratio $(AN \cdot BN)/AB^2$ is a constant since all its terms are constant values. The product $BD \cdot AC$ is a constant by theorem 4. Consequently,

$$MN \cdot NQ = \text{constant}.$$

This is how Hart's inversor works. Taking any of the above four points as the pole of inversion, we move the second point along a circle passing through the first point. Then the third point traces a straight line.



Challenges in physics and math

Math

M16

Virus versus bacterium. A colony of n bacteria is invaded by a single virus. During the first minute it kills one bacterium and then divides into two new viruses; at the same time each of the remaining bacteria also divides into two. During the next minute each of the two newly born viruses kills a bacterium and then both viruses and all the remaining bacteria divide again, and so on. Will this colony live infinitely long or will it eventually perish? (R. Kovtun)

M17

All isosceles. On straight lines AB and BC containing two sides of a parallelogram $ABCD$ points H and K are chosen so that the triangles KAB and HCB are isosceles ($KA = AB$, $HC = CB$; see figure 1). Prove that the triangle KDH is also isosceles. (V. Gutenmacher)

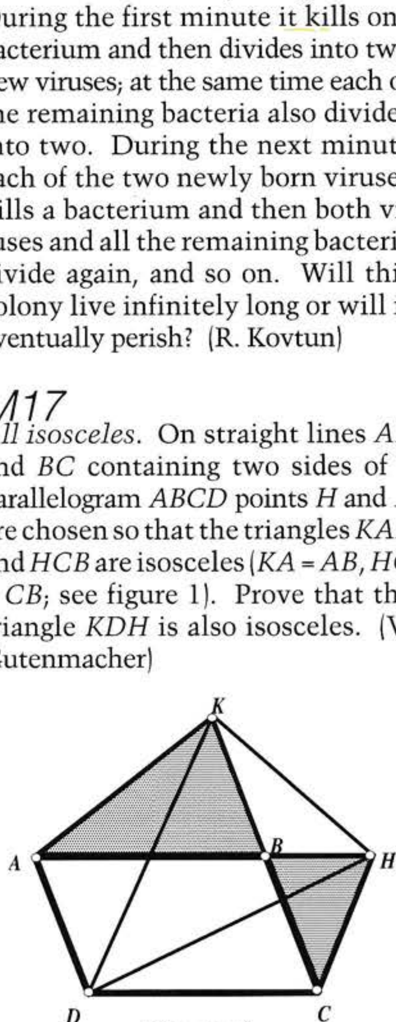


Figure 1

M18

Numismatics. At a trial 14 coins were produced as physical evidence. An expert found seven of them counterfeit and the other seven genuine, and he knows which are which. But the judge knows only that the counterfeit coins all weigh the same, as do the genuine ones; and, in addition, that the latter are heavier than the former. How can the expert convince the judge of the correctness of his expertise by three weighings on a pan balance? (R. Freiwald)

M19

Summing chords. A number of chords are drawn in a circle of radius 1 so that each diameter crosses no more than k chords. Prove that the sum of the lengths of all the chords is less than πk. (A. Kolotov)

M20

Internal depreciation. (a) When a number N is multiplied by 8, the sum $S(N)$ of its digits can decrease (for example, $S(75) = 12$, whereas $S(8 \cdot 75) = S(600) = 6$). Prove that it can't decrease by a factor of more than 8. In other words, prove that

$$\frac{S(8N)}{S(N)} \geq \frac{1}{8}$$

for any natural N.

(b) What are the other natural numbers k for which a positive c_k can be found such that

$$\frac{S(kN)}{S(N)} \geq c_k$$

for any natural N? What's the greatest suitable value of c_k for a given k? (I. Bernstein)

Physics

P16

High-stepping hoop. A ring of radius R rolling along a horizontal surface with velocity v hits a step of height h ($h << R$). The collision is absolutely inelastic. What will the velocity of the ring be after it "climbs" the step? At what minimum velocity can the ring climb the step? (There is no slippage.)

P17

Air strain. Two weightless pistons connected by a thin weightless string of length l (fig. 2) are positioned in two cylinders with cross sections S_1 and S_2. The space between the pistons is filled with water. Find the strain on the string if each vessel opens up into the atmosphere. (The density of water is ρ.)

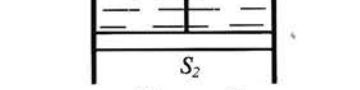


Figure 2

P18

A warm impression. A coin is pressed tightly against a frost-encrusted window. The ice under the coin first starts to melt along the edge and only later under the center of the coin. Why is that?

P19

Through thick and thin. In 1815 the English scientist Children staged the following set of experiments. Two platinum wires of the same length but different diameters were connected to the Volt battery. In the first experiment the wires were connected in series, whereas in the second the connection was parallel. In the first case only the thin wire was heated, and in the second the hot wire was the thick one.

For almost 25 years scientists were unable to explain why. Maybe you can come up with the answer in a bit less time!

(Hint: assume that the quantity of heat radiated by the conductor into the surroundings is proportional to the conductor's surface area and to the temperature difference between the conductor and the surroundings.)

P20

Water specs. What eyeglasses should be prescribed for a person whose eyesight is normal under water?

SOLUTIONS ON PAGE 57

The natural logarithm

What's so "natural" about 2.71828... anyway?

by Bill G. Aldridge

YOU KNOW WHAT A LOGArithm is—it's just another word for an exponent that represents a number. You choose a base—for instance, 10—and, by assigning an exponent x, you can represent the number n in the form $\log_{10}n = x$ (or more commonly lg n). The interesting thing about numbers in logarithmic form is how calculations are simplified.

The great mathematician Laplace (1749–1827) said, "The invention of logarithms shortens calculations extending over months to just a few days and thereby, as it were, doubles the life-span of the calculators." (Back in those days, *calculator = person*.) To take a simple example, to multiply two numbers logarithmically, you just add their exponents. What's 5,673 times 1,347? Referring to a table of common (that is, base 10) logarithms, we find that lg 5673 = 3.75381 and lg 1347 = 3.12937. Our multiplication problem then becomes $10^{3.75381}$ times $10^{3.12937}$. Adding the exponents, we get $10^{6.88318}$. Working backward in the log table, we find that if lg n = 6.88318, n equals approximately 7.6415 times 10^6, or 7,641,500. If we multiplied it out the long way, we'd get 7,641,531. As you can see, the use of logarithms produces approximations because tables of logs are carried out only to a certain number of decimal places. For most purposes, though, the precision achieved is perfectly adequate.

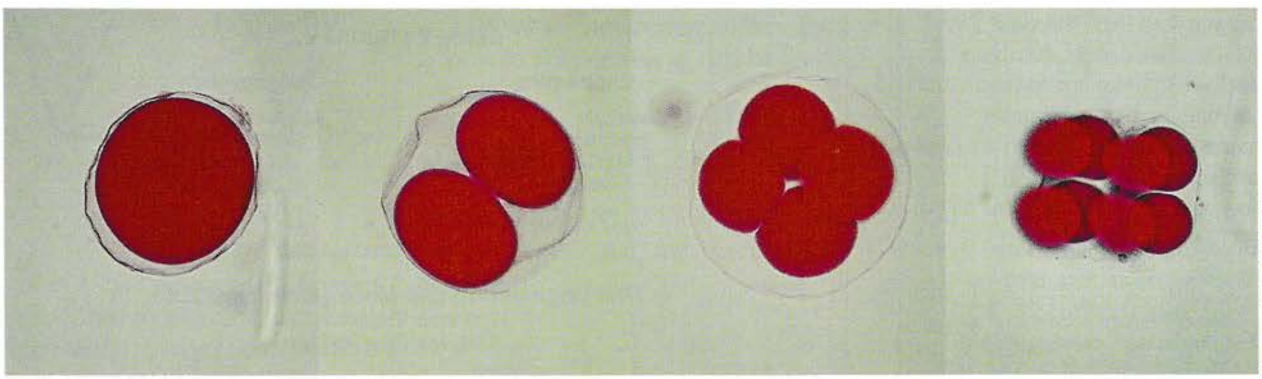
Maybe that didn't look like much of a simplification, but the importance of logarithms and their laws in performing complicated calculations—for instance, those involving roots—was immense. I say "was" because pocket calculators now perform in fractions of a second what used to take hours or days, even with logarithms. But other, more important uses of the logarithm persist in science, math, and engineering.

Base 10 logarithms arose out of our number system, based as it is on the number 10. All logarithmic relationships that occur in the natural world, though, have a different base. Because of that, such logarithms are called "natural." Unlike the "common" base, the natural base is a transcendental number. Now, does that sound "natural" to you?

The number is designated e, and it is usually written as the rational approximation 2.71828... .

Although I used natural logarithms in high school and college and had derived the value of e mathematically, I never really knew where it came from. You can find a derivation of e in many math texts, but it's always presented as an abstraction. That always bothered me.

Finally, long after leaving school, I decided to work through for myself to see how e can be deduced from an actual process in the natural world. I



Photographs by Ed Reschke

Another case of e *at work: early cleavage of a fertilized starfish egg (magnified 100x).*

could have chosen radioactive decay, or the discharge of a capacitor in an electrical circuit, or the abstract concept of entropy in statistical thermodynamics. My specialty happens to be physics, but I decided to use a biological phenomenon in my pursuit of e.

The biology of bacterial growth

I looked at the growth of staphylococcus bacteria in what is called a selective culture medium (Staphylococcus 110 agar). This bacterium has a diameter of 0.5 to 1.5 micrometers and splits every 20 minutes or so. It does this in the nutrient culture, and the process is called *transverse binary fission*. The process goes like this: A newly formed cell undergoes a gradual increase in volume, as it prepares for cell division. After some time it forms a *septum* that ultimately divides the enlarged cell into two identical *daughter* cells. Cellular components are divided equally between the two developing cells. Each of the daughter cells then begins to increase in volume in preparation for the next division cycle. Each time the cells divide the population doubles.

The *generation time*, defined as the time it takes the population to double, varies from 20 minutes to several days, depending on the species of organism and the culture in which it grows. The generation time can be found just by watching a few cells divide, using the average time needed to divide as your estimate of the generation time. (You'd need a microscope to do this, which I didn't have handy, so I got all this from a book.)

It turns out that actual bacterial growth is exponential (doubling each 20 minutes) only for a certain period of time. At first the bacteria must adjust to the medium (*lag time*); then the growth is exponential; then it levels off to a stationary period, when all of the nutrient has been used; finally, the cells begin to die, and the curve drops. I restricted myself to the phase of exponential growth.

If we started our culture with daughter cells all of the same size, just after they have formed, we could observe them dividing in a synchronous fashion, at least for a while, until they began to get out of phase. But if we just select a random sample of the bacteria, some are ready to divide, others have just started to grow, and still others are at some point in the growth phase. Since the bacteria are in various stages, a given bacterium might divide at any time. If there are enough bacteria, cell division in this asynchronous mode occurs almost continuously. I looked into this growth pattern because it fulfills the assumptions needed for the math.

The mathematics of bacterial growth

Each bacterium divides into two bacteria after a certain period of time. Each of these two daughters grows and then each one divides into two more, and the process continues for as long as there is nutrient and space available for new cells. If we start with 5,000 bacteria, and the generation time is 20 minutes, how many will there be in two hours? Let's say we start at 8:00 o'clock. At 8:20, we have 10,000 bacteria; at 8:40, we have 20,000; at 9:00, we have 40,000; at 9:20, we have 80,000; at 9:40, we have 160,000, and at 10:00, two hours later, our population of 5,000 bacteria has increased to 320,000. How many would we have at the end of the next two hours?

Next, I tried to find an equation that describes the relationship between the time t required for a certain number N of bacteria to be produced from a small initial number N_0. (We can assume that we're starting with so many bacteria and at such different stages of growth that cell division is occurring randomly and continuously.) I could then divide the time t into a large number n of small intervals, each having the same size Δt. Because each interval is Δt long and there are n of them all together, the total time t is given by $t = n\Delta t$. Since the interval Δt is just the total time divided by the number of intervals, we have the expression $\Delta t = t/n$.

During any time interval, ΔN of the cells divide. Suppose that the time interval Δt is 0.01 second and we get a certain number of divisions in that 0.01 second. If the interval is increased to, say, 0.02 or 0.03 second, the number of cells ΔN produced in that interval is also greater by a factor of two or three. If we provide twice as many cells at the beginning of that time interval, then there will also be twice as many cells produced. In other words the number of cells ΔN that divide during the time interval Δt is proportional to that time interval and to the number of cells N present when the interval starts. I expressed this relationship mathematically by the proportion

$$\Delta N \sim N\Delta t.$$

This proportion means, for example, that if Δt or N is doubled, the number of cells that can divide doubles. If either factor is halved, only half as many cells can divide.

Writing the proportionality as an equation by including a proportionality constant k, I got

$$\Delta N = kN\Delta t.$$

I've said there is an extremely large number n of these very small time intervals Δt. The increase in the number of bacteria ΔN given by this equation can be used for each of several time intervals. For the first interval,

$$\Delta N = kN_0\Delta t,$$

where N_0, the number of bacteria present at the beginning of the first interval, is merely the number of cells with which we started.

The number of bacteria at the end of the first interval is $N_0 + \Delta N$. But if the value of ΔN from the previous equation is used, this total must be $N_0 + kN_0\Delta t$. If we factor out N_0, we have just $N_0(1 + k\Delta t)$ for the number of bacteria at the end of the first time interval. Let's call that number N_1. So

$$N_1 = N_0(1 + k\,\Delta t).$$

Now I had to find the number of bacteria at the end of the second time interval. The increase ΔN again had to be proportional to the number N_1 of bacteria I started with in that interval and the length of the interval. Again, using k as a constant of proportionality, we have the equation

$$\Delta N = kN_1 \Delta t$$

for the increase in the number of bacteria during the second interval. The total number of bacteria present at the end of the second interval is obviously the number present when it started plus the increase—$N_1 + kN_1\Delta t$. Stating it as an equation, at the end of the second interval we have

$$N_2 = N_1 + kN_1\Delta t.$$

Factoring out N_1, we get

$$N_2 = N_1(1 + k\Delta t).$$

Since I had already found the number of bacteria at the end of the first interval N_1 in terms of the starting number N_0, I simply replaced N_1 in this equation with that number, which gives

$$N_2 = [N_0(1 + k\Delta t)](1 + k\Delta t),$$

or more simply,

$$N_2 = N_0(1 + \Delta t)^2.$$

By now I'm sure you've caught on and know what my next task was: to find the number of bacteria at the end of the third time interval. I started with N_2, so that the increase is given by

$$\Delta N = kN_2 \Delta t.$$

As before, the number N_3 at the end of the third time interval is given by $N_2 + \Delta N$—what we started with plus the increase. So we have

$$N_3 = N_2 + kN_2\Delta t,$$

and factoring out N_2 we get

$$N_3 = N_2(1 + k\Delta t).$$

Replacing N_2 by the value we had in terms of N_0, we now have

$$N_3 = [N_0(1 + k\Delta t)^2](1 + k\Delta t),$$

or more simply

$$N_3 = N_0(1 + k\Delta t)^3.$$

You see the pattern that results from these steps, and maybe you're bored by them. But we're on the verge of generalizing the result, and that's always fun.

If we continue to look at the number of bacteria at the end of successive time intervals, the total at the end of each interval is equal to the number present at the end of the preceding interval times the quantity $(1 + k\Delta t)$. When we do the various substitutions for the starting numbers, back to the initial amount, we'll have, at the end of n time intervals,

$$N_n = N_0(1 + k\Delta t)^n.$$

The quantity N_n is the total number of bacteria that have been produced after a time t has elapsed. There were n intervals of size Δt each making up that time t. So $t = n\Delta t$ and $\Delta t = t/n$. (Sorry if this seems like beating a dead horse, but I'm exposing all the steps in my thinking, including the ones we usually just buzz past.)

I used this value of Δt in my equation for N_n. I then got

$$N_n = N_0\left(1 + \frac{kt}{n}\right)^n.$$

You'll notice that the quantity $(1 + kt/n)^n$ is a binomial that can be expanded in a binomial series, giving

$$\left(1 + \frac{kt}{n}\right)^n = 1 + n\frac{kt}{n}$$
$$+ n(n-1)\frac{(kt/n)^2}{2!}$$
$$+ n(n-1)(n-2)\frac{(kt/n)^3}{3!} + \dots .$$

If we've made Δt small enough, a very large number n of time intervals is involved. We want the value of the terms of this binomial expansion when n is very large. From elementary calculus I knew how to "take the limit of this expression as n approaches infinity," which is written

$$\lim_{n\to\infty}\left(1 + \frac{kt}{n}\right)^n.$$

In the limit, as n becomes infinitely large, the binomial series simplifies considerably. This is because all factors involving n, like $n(n-1)(n-2)$, become simple exponentials of n—in this case, simply n^3. As such, they are all canceled by identical powers of n in the denominator of the quantity (kt/n), which is always raised to the same power. So, in the limit, our particular binomial expansion is just

$$1 + kt + \frac{(kt)^2}{2!} + \frac{(kt)^3}{3!} + \dots + \frac{(kt)^m}{m!}.$$

In this limiting case, the number N_n of bacteria after time t is then given by

$$N_n = N_0\left[1 + kt + \frac{(kt)^2}{2!} + \frac{(kt)^3}{3!} + \dots + \frac{(kt)^m}{m!}\right].$$

Suppose we let kt equal 1 in this series. Then for $m = 8$ the series becomes

$$1 + 1 + \frac{1}{2} + \frac{1}{6}$$
$$+ \frac{1}{24} + \frac{1}{120} + \frac{1}{720}$$
$$+ \frac{1}{5040} + \frac{1}{40320} + \dots .$$

or, in decimal form,

$$1 + 1 + 0.5 + 0.1666666667 + 0.0416666667 + 0.0083333333 + 0.0013888889 + 0.0001984127 + 0.0000248016 + \dots .$$

When I added these numbers, I got 2.71828... . *Eureka*! (Or is it *déjà vu*?)

A handy little nonrepeating constant ...

What's more interesting is that when we let $kt = 2$, the value of this expansion is 7.389056..., which is just $(2.71828...)^2$. If $kt = 3$, the sum of this series becomes the cube of 2.71828..., and so on. This endless, nonrepeating decimal number we get as a base for the exponential isn't even a rational number. So let's call it e. (That's nice and irrational, isn't it?) Its value, to 16 places, is 2.718281828459046. Then our equation for the number of bacteria after some time t becomes quite simple:

$$N = N_0 e^{kt}.$$

Now if we define (which someone already had) the *natural logarithm* as "the exponent needed on e to give a certain result," we can write this exponential expression as a base e logarithm. (Just as base 10 logarithms are abbreviated to "lg," base e logarithms are shortened to "ln.") Writing our equation for bacterial growth in logarithmic form, we get

$$kt = \ln(N/N_0),$$

or in terms of the time t

$$t = (1/k)\ln(N/N_0).$$

This equation tells us how long it takes to produce N bacteria when we start with N_0 of them. To use this equation, you would need a table of natural logarithms. But these are readily available in books and on almost all electronic calculators.

Needless to say, I was quite pleased when all my calculations worked out correctly and I discovered for myself the connection between the base of the natural logarithm, 2.71828... , and a natural phenomenon. But nothing would have "clicked" if I hadn't admitted to myself that something was "stuck." So if something doesn't seem to make sense to you, don't be afraid—or ashamed—to work it out for yourself, no matter how trivial it might seem to someone else.

In the meantime, here are a few problems that involve working with e. (And maybe one of you can tell me: why "e"?!)

Exercises

1. Suppose you know the generation time t_g for a given bacterium—say, 20 minutes. Find the constant of proportionality k.

2. Starting with 10 bacteria and the same generation time as above, how long will it take to get 1,000,000?

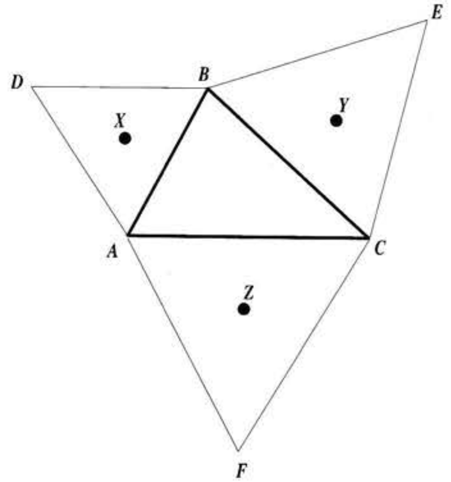
3. Prove that e is not the rational number 271,801/99,990.

SOLUTIONS ON PAGE 63

Readers write . . .

From Richard G. Brown of Phillips Exeter Academy: "I enjoyed the 'Botanical Geometry' article in your September/October 1990 issue. Your reference to Napoleon's triangle reminded me of the special relationship posed in the problem below. Some years ago, a geometry class and I discovered this relationship. Our proof used vectors."

Problem: ABC is an arbitrary triangle with equilateral triangles built on its sides as shown. X, Y, and Z are centroids of these equilateral triangles. (XYZ is known as Napoleon's triangle and is equilateral.) The problem is to discover a relationship involving the centroids of triangles ABC, XYZ, and DEF.



MATHEMATICAL SURPRISES

Play it again . . .

and again . . .

by John Conway

On 1089

When I was a little boy, my father taught me something that really puzzled me. You start with a three-digit number (say, 379), reverse it (to 973), and take the difference:

$$\begin{array}{r} 973 \\ -\ \underline{379} \\ 594 \end{array}$$

Then you reverse *that*, and add:

$$\begin{array}{r} 594 \\ +\ \underline{495} \\ 1089 \end{array}$$

We get 1089. What's so surprising about that? Well, the surprising thing, my Dad said, is that you *always* get that *same* answer: 1089.

Well, he wasn't quite right. If you start with a number whose first and last digits are equal (say, 585), you'll get zero:

$$\begin{array}{r} 585 \\ -\ \underline{585} \\ 000 \end{array} \qquad \begin{array}{r} 000 \\ +\ \underline{000} \\ 000 \end{array}$$

But it *is* true that you'll always get either 0 or 1089. Can you explain why?

On 6174

The Indian mathematician Kaprekar discovered what to my mind is a more surprising result of this kind. You start with any four-digit number whose digits aren't all equal, arrange its digits to form both the largest and the smallest numbers you can, and take their difference:

$$\begin{array}{r} 4321 \\ -\ \underline{1234} \\ 3089 \end{array}$$

Then you just keep on doing that same thing, and quite a strange thing happens:

$$\begin{array}{rrrrrrrrr} 4321 & 9830 & 9441 & 9972 & 7731 & 6543 & 8730 & 8532 & 7641 \\ -\ \underline{1234} & -\underline{0389} & -\underline{1449} & -\underline{2799} & -\underline{1377} & -\underline{3456} & -\underline{0378} & -\underline{2358} & -\underline{1467} \\ 3089 & 9441 & 7992 & 7173 & 6354 & 3087 & 8352 & 6174 & 6174 \end{array}$$

After a time, the answer you get is always Kaprekar's magic number 6174, which, as you can see, leads immediately to itself. Can you show that this indeed always happens?

On 153

It seems a bit artificial to work with a fixed number of digits, so here's something that works with arbitrarily large numbers. Start with any positive multiple of 3 and repeatedly replace it with the sum of the cubes of its digits. Why do you always get to the magic number 153?

It helps to know the cubes of the ten digits:

$$0^3 = 0,\ 1^3 = 1,\ 2^3 = 8,\ 3^3 = 27,\ 4^3 = 64,\ 5^3 = 125,$$
$$6^3 = 216,\ 7^3 = 343,\ 8^3 = 512,\ 9^3 = 729.$$

An example:

$$\begin{aligned} &999999 \longrightarrow 6 \cdot 729 = 4374, \\ &4374 \longrightarrow 64 + 27 + 343 + 64 = 498, \\ &498 \longrightarrow 64 + 729 + 512 = 1305, \\ &1305 \longrightarrow 1 + 27 + 0 + 125 = 153, \\ &153 \longrightarrow 1 + 125 + 27 = 153 \ldots \end{aligned}$$

On RATS

Here's a digital game I invented that contains an unsolved problem. I call it "RATS," which is an acronym for **r**everse, **a**dd, **t**hen **s**ort. You take any positive number, reverse it, add the result to the original, then sort the digits of the answer in increasing order, deleting any initial zeros. You just keep on doing that, and watch what happens:

3	6	12	33	66	123	444	888	1677	3489
3	6	21	33	66	321	444	888	7761	9843
6	12	33	66	132	444	888	1776	9438	13332

12333	44556	(000)111	222	444	888	1677
33321	65544	111	222	444	888	7761
45654	110100	222	444	888	1776	9438

and we're in a cycle.

Lots of other numbers give cycles, but some don't—for instance, the number 1 starts off a sequence that gets larger and larger:

1	2	4	8	16	77	145	668	1345	6677
1	2	4	8	61	77	541	866	5431	7766
2	4	8	16	77	154	686	1534	6776	14443

13444	55778	133345	666677	1333444
44431	87755	543331	776666	4443331
57875	143533	676676	1443343	5776775

5567777	12333445	66666677	133333444
7777655	54433321	77666666	444333331
13345432	66766766	144333343	577666775

556667777	12333334444	55666667777
777766655	44443333321	77776666655
1334434432	56776667765	133443334432

123333334444	556666667777	1233333334444	
444433333321	777766666655	4444333333321	
567766667765	1334433334432	5677666667765	...

and now alternate numbers are very similar, but increasing in length.

Now here's the unsolved problem. Every number that we've tried (and we've tried all numbers with fewer than 15 digits!) either gets into a cycle or enters this particular sequence. Does this continue forever? We don't know.

On 0, 0, 0, 0

Here's a slightly more mathematical problem. Start with any four whole numbers *a*, *b*, *c*, *d* and replace them with the difference of *a* and *b*, *b* and *c*, *c* and *d*, *d* and *a*. Repeat this process and watch:

5	98	67	13
93	31	54	8
62	23	46	85
39	23	39	23
16	16	16	16
0	0	0	0
0	0	0	0

We always get to four zeros, and often surprisingly quickly. Can you prove this? What happens if you take three numbers instead of four, or indeed any other number of numbers?

A very mysterious sequence

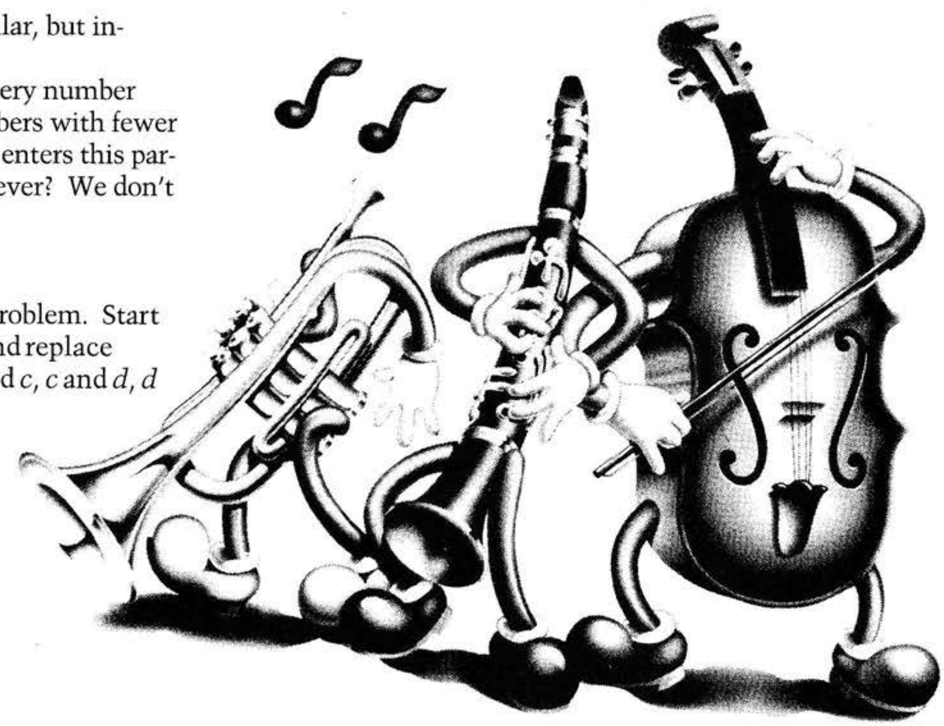
I end with a problem that I don't expect many of you to solve completely. What's the rule that governs the sequence of digit-sequences

1
11
21
1211
11221
312211
13112221
1113213211
31131211131221
13211311123113112211
...

and how rapidly does the length of the *n*th sequence tend to infinity?

The first half of the question is easy—if you can't do it yourself, ask somebody younger than you are for some help. The second half is quite surprising.

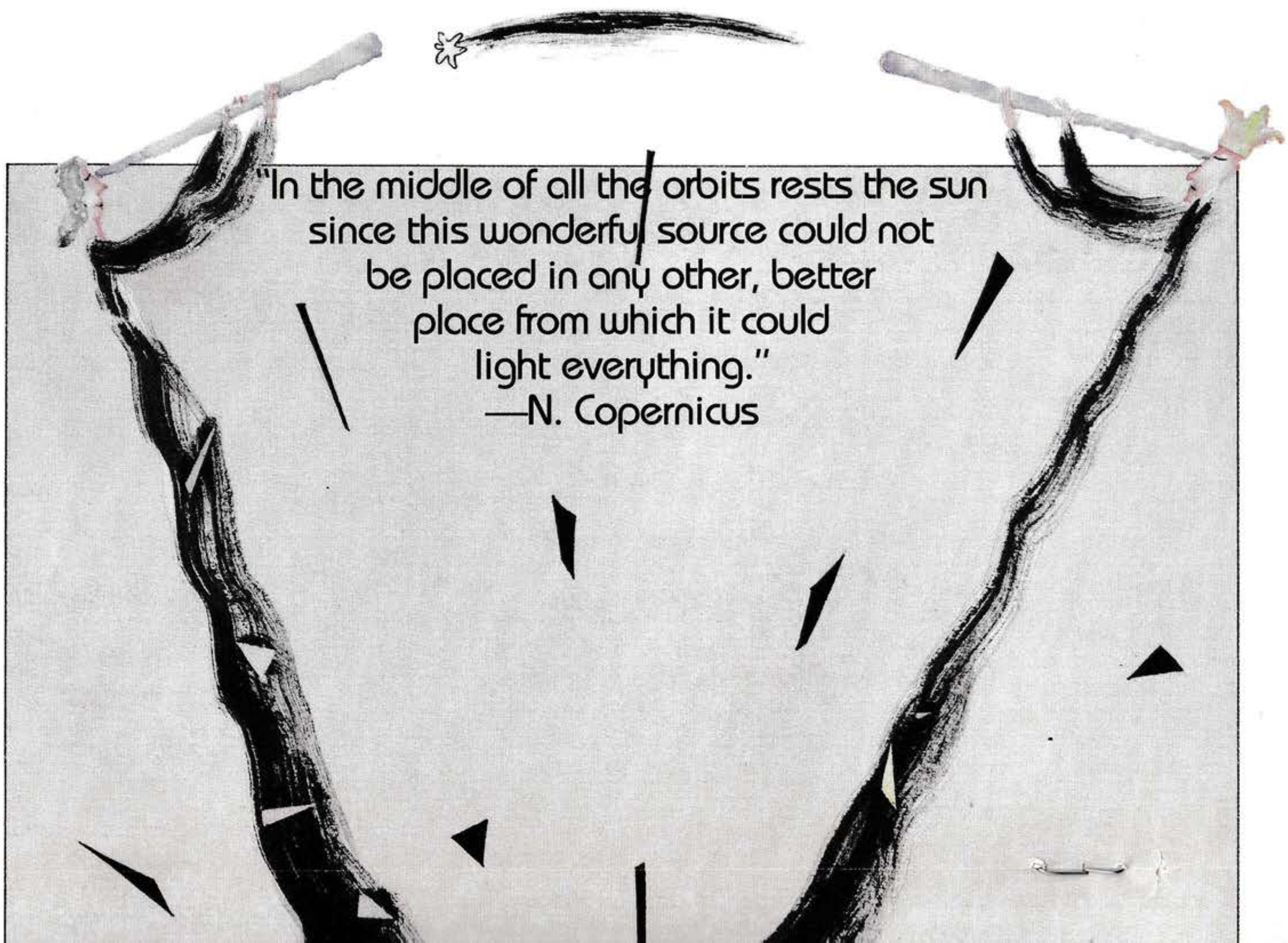
SOLUTIONS ON PAGE 63



What's new in the solar system?

A lot—but the old laws of orbital motion still apply

CENTURIES SEPARATED the first attempts of ancient scholars to find regularity in the motion of planets—"wandering stars"—and the work of Nicolaus Copernicus (1473–1543). The revolution begun by Copernicus and accelerated by his followers influenced all further development of astronomy. Since then mankind has dramatically increased its research capabilities. We can now study the solar system directly from space vehicles. But space science didn't bring any revision of the fundamental laws established by observational astronomy. On the contrary, using these laws we can solve new problems that have arisen only with the advent of the space age.



Questions and problems

1. When do you move faster around the Sun, at midday or at midnight?

2. When does the Earth move faster in its orbit around the Sun, in winter or in summer?

3. Astronomers have found that the velocities of different

Sun, in winter or in summer?

3. Astronomers have found that the velocities of different parts of Saturn's ring are not proportional to their respective distances from the rotation axis. What does this say about the ring's structure?

4. The Sun attracts the Moon almost twice as strongly as it does the Earth. So why doesn't the Moon—the Earth's satellite—become a separate planet?

5. Can a planet or satellite move in an elliptical orbit at a constant speed?

6. Why isn't it possible to launch a satellite that would constantly hover over a region of the Earth situated at a certain latitude?

7. Can we apply Kepler's third law to compare the rotational periods of the Earth and the Moon?

8. The angular velocity of rotation of a hypothetical planet is such that bodies are weightless at the equator. What speed does a body have to attain to go into orbit around the planet?

9. Imagine that the Earth is set on a table rotating along the orbit around the Sun. What is the force with which the Earth acts on the table?

10. How does air resistance change the speed of a satellite moving in the rarefied upper layers of the Earth's atmosphere?

Art by Sergey Ivanov

Mental microexperiment

Imagine you're an astronaut returning to your spacecraft after a space walk with a bit too much velocity. Is it possible to get a bruise hitting the spacecraft? Remember, you're weightless . . .

It's interesting that . . .

. . . Mark Twain was born in 1835, two weeks after the appearance of Halley's comet, and died the day after its closest approach to the Sun in 1910. Shortly before that he joked that since he was born in the year when Halley's comet made its scheduled visit, he would die right after its next scheduled appearance.

. . . the Ptolemaic system is fundamentally flawed, but it's still capable of predicting some celestial phenomena to any accuracy. It may sound paradoxical, but the Ptolemaic system can be used to solve several problems in modern astronautics—for instance, calculating the trajectories of space vehicles visible in the sky.

. . . the twin-Sun hypothesis has gotten to be quite popular in recent years. It suggests that the Sun has a "sibling" rotating around a common center of mass along an extremely elongated elliptical orbit. This notion is supported by paleontologists, who have found a certain cyclicity in the fossil record of extinctions of animal and plant species. These natural catastrophes are thought to be linked with comet showers caused by the other sun's approach to "our" Sun. The round trip would take the Sun's sibling at least 26 million years. Right now they say it's at an extremely distant point in its trajectory, which is why it has never been detected.

SOLUTIONS ON PAGE 60

Shapes and sizes

Specifically, convex polygons with integer sides inscribed in a circle of integer radius

by George Berzsenyi

IT'S NOT DIFFICULT TO SEE that if (a, b, c) is a Pythagorean triple (that is, positive integers such that $a^2 + b^2 = c^2$), then the right triangle with sides $2a, 2b, 2c$, as well as the quadrangles with two sides of length $2a$ and two sides of length $2b$, are inscribable in a circle of radius c. Since there are well-known methods for the generation of Pythagorean triples, it's easy to characterize all such triangles and quadrangles.

This month's problem asks the more general question: **What other convex polygons with integer sides can be inscribed in a circle of integer radius?** Notice that we may not even be done with triangles and rectangles, so an easier version of the problem is to address that issue. At the other extreme, we may wish to remove the restriction of convexity.

This problem is a natural extension of Problem 557 in the February 1981 issue of the now defunct journal *Mathematics Student*, whose Competition Corner I edited for three years. During that time an average of 69 solutions were submitted to the 102 problems posed—and problem 557 got its fair share. The problem asked for polygons with three sides of length a and three of length b, inscribed in a circle of radius r, with a, b, and r integers. In solving the problem, the students were led to the Diophantine equation $a^2 + b^2 + ab = 3r^2$, whose solutions yield all of the hexagons with the desired properties. Some participants of the Competition Corner also studied hexagons with four sides of length a and two of length b (inscribed in a circle of radius r) and found that they can be obtained by solving yet another Diophantine equation, $a^2 + br = 2r^2$. Both of these equations yield infinitely many solutions, which can be found by standard methods. **Are these the only hexagons with integer sides inscribable in a circle of integer radius?** This question never arose. As a minor puzzle, I leave it to you to decipher what must be the lengths of the sides of the polygons in figures 1, 2, and 3, given that they can be inscribed in circles of radii 5, 7, and 9, respectively.

Please send your solutions to these problems to *Quantum*, 1742 Connecticut Avenue NW, Washington, DC 20009. The best results will be acknowledged, and their authors will receive free subscriptions to *Quantum* for one year and/or book prizes.

At sixes and sevens

In the May 1990 issue of *Quantum* I asked whether the Roseberry Conjecture, "All positive integers that are not multiples of 5 have an integer multiple consisting of 6's and 7's only," is true. Solutions were submitted by

CONTINUED ON PAGE 45

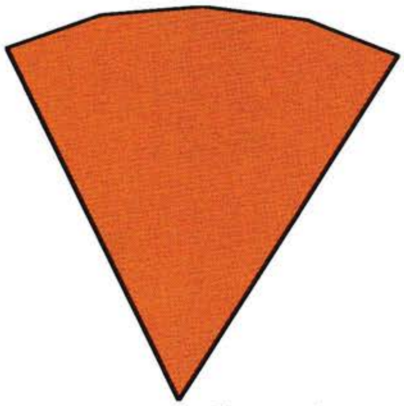


Figure 1

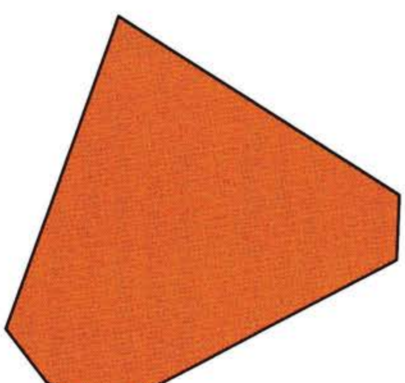


Figure 2

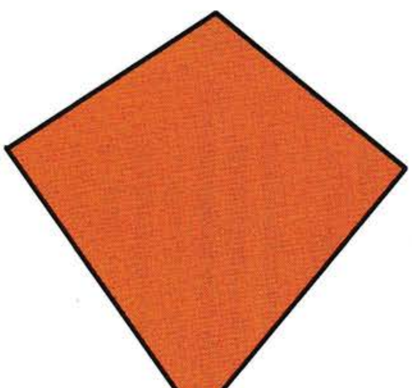


Figure 3

Neutrinos and supernovas

"When shall the stars be blown about the sky,
Like the sparks blown out of a smithy, and die?"
—William Butler Yeats, "The Secret Rose"

by Arthur Eisenkraft and Larry D. Kirkpatrick

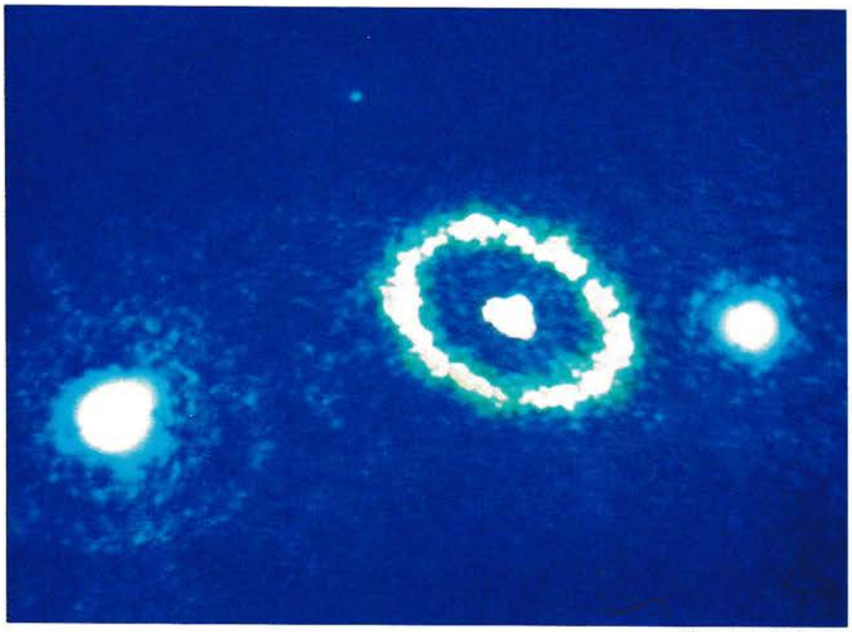


Photo courtesy of NASA

THE SUPERNOVA 1987A PROvided us with a personal view of a dying star and kindled new interest in the infant field of neutrino astronomy. The neutrino was originally proposed to "save" the laws of conservation of energy and momentum in beta decay. If a neutron decayed into a proton and an electron, the conservation laws required that the electron have a well-defined kinetic energy in the center of mass system. Experiments showed, however, that the electrons exhibited a spectrum of kinetic energies ranging from zero to the predicted value.

In 1930 Wolfgang Pauli proposed that a third particle was involved in beta decay. To agree with the conservation laws, the neutrino had to be neutral and have a very small rest mass, possibly zero. It took 26 years for the neutrino to be discovered because it interacts so weakly with matter—on average only one in a trillion neutrinos would be stopped in passing through the Earth. In spite of this extremely weak interaction, it's now known that there are three different types of neutrino: one paired with the electron, one with the muon, and one with the tau.

Although the mass of these neutrinos may be zero, this has not been confirmed. Since measuring devices can never be perfect, the best we can do is set an upper limit on the masses of the neutrinos. At present the mass of the electron neutrino is known to be less than 18 electron volts (eV), where the mass is expressed in its energy equivalent. The experimental limits on the other neutrino masses are not as low. The masses of the muon and tau neutrinos may be as high as 250 keV and 35 MeV, respectively. So any experiment that could place more restrictive limits would be welcome. Such an opportunity was provided by supernova 1987A, which occurred relatively close to Earth at a distance of 170,000 light years. After the observation of the supernova, experimentalists examined the data taken by several experiments that were running at the time and discovered a number of neutrino events.

In order to see how the observation of these neutrinos can help us determine their mass, let's consider the following simplified situation. **Assume that the supernova emits an extremely short burst of electron neutrinos and that neutrinos with an energy of 15 MeV arrive at the detectors 15 seconds after the arrival of 7.5-MeV neutrinos. What mass must the neutrinos have to account for the time delay in their arrival?**

Please send your solutions to *Quantum*, 1742 Connecticut Avenue NW, Washington, DC 20009. The best solutions will be published in *Quantum* and their creators will receive free subscriptions for one year.

CONTINUED ON PAGE 45

Art by Sergey Ivanov

Genealogical threes

A method of generating Pythagorean triples rooted in Euclid's algorithm for the greatest common divisor

by A.A. Panov

OUR STORY IS ABOUT MATHEMATICAL classics—Euclid's algorithm and Pythagorean triples. Euclid's algorithm is described in his *Elements* (about 300 B.C.) but was surely known long before that date. The history of Pythagorean triples can be traced even further back. A remarkable monument of human culture is a Babylonian clay cuneiform tablet that lists fifteen Pythagorean triples. The tablet dates from about 1500 B.C.[1]

Shaking the dust off these ancient notions, we'll talk about them using the "language of trees." This language is convenient for solving a number of equations and clarifies the relation between Euclid's algorithm and a method of constructing Pythagorean triples proposed recently by a British mathematician.

Euclid's algorithm

Euclid's algorithm finds the greatest common divisor (GCD) of two natural numbers.

> Let (m, n) be a pair of positive integers:
> (1) if $m = n$, then $d = m = n$ is the greatest common divisor of m and n; if $m \neq n$, go to step 2);
> (2) replace the larger of the numbers m and n with the difference after subtraction by the smaller and go back to step (1).

Maybe you're more familiar with another version of step 2:

> (2′) replace the larger number by the remainder after division by the smaller and go back to step (1).

It's a matter of taste, really.

Problem 0. *Prove that algorithms (1), (2) and (1), (2′) both yield the same result. (Hint: division is equivalent to repeated subtraction.)*

Euclid himself referred to his algorithm as "constant subtraction of the smaller from the larger" (*Elements*, Book 7, Proposition 2). This "repetitive subtraction" algorithm is the subject of our story.

Let's look at an example of how the algorithm works. Let $(m, n) = (20, 12)$. Writing out the consecutive pairs of numbers from right to left, we get the following chain:

$$(4,\ 4) \leftarrow (8,\ 4) \leftarrow (8,\ 12) \leftarrow (20,\ 12).$$

This means that the GCD of $(20, 12) = 4$. Applying the same procedure to the pair $(5, 3)$ we get

$$(1,\ 1) \leftarrow (2,\ 1) \leftarrow (2,\ 3) \leftarrow (5,\ 3).$$

You can see that each number in the second chain equals the corresponding number in the first chain divided by 4.

Now try to answer the following question.

Problem 1. *Let* d = *the GCD of* (M, N). *We'll say* m = M/d *and* n = N/d. *What's the GCD of* (m, n)? *How is the action of Euclid's algorithm on the pair* (M, N) *similar to its action on the pair* (m, n)?

From now on we'll limit ourselves to pairs (m, n) for which the GCD = 1. We'll call such pairs "simple pairs."

The genealogy of simple pairs

Let's look at another example. Applying Euclid's algorithm to the simple pair (3, 4) we get

$$(1,\ 1) \leftarrow (2,\ 1) \leftarrow (3,\ 1) \leftarrow (3,\ 4).$$

A portion of this chain coincides with a portion of the chain for the pair (5, 3), so we can join them together:

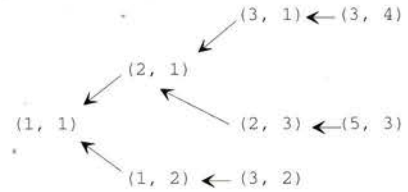
$$\begin{tikzcd}[row sep=small] & & (3,\ 1) \arrow[dl] & (3,\ 4) \arrow[l] \\ (1,\ 1) & (2,\ 1) \arrow[l] & & \\ & & (2,\ 3) \arrow[ul] & (5,\ 3). \arrow[l] \end{tikzcd}$$

We can add another simple pair (3, 2), and the picture gets more complicated:

---

[1]Now a part of the Plimpton Collection in the Butler Library at Columbia University.—*Ed.*

Art by Sergey Ivanov



This suggests that there may be a general pattern uniting all simple pairs. How do we find it? We could just add more simple pairs. But sooner or later we'd realize that the right question to ask is this: for any simple pair (m, n), what are the other pairs whose arrows are aimed at this one?

Problem 2. *Prove that if Euclid's algorithm produces an arrow from the pair* (M, N) *to the pair* (m, n), *then either* M = m + n, N = n, *or* M = m, N = m + n.

This problem suggests that we have to introduce two transformations t_1 and t_2 that turn the pair (m, n) into

$$t_1(m, n) = (m + n, n),$$

$$t_2(m, n) = (m, m + n). \qquad (1)$$

We can now proceed in reverse order. Starting from the pair (1, 1) and applying the transformations t_1 and t_2 (shown by upward and downward arrows, respectively), we get two new pairs (2, 1) and (1, 2). We apply the transformations to each of them, and so on.

Each pair now gives rise to two new pairs, and this process can be continued to infinity. As expected, figure 1 contains all the preceding chains as pieces of itself, but with all the arrows reversed.

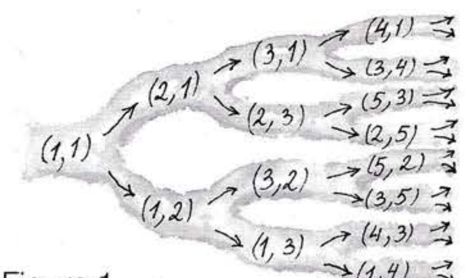


Figure 1

Problem 3. *Prove that any pair* (m, n) *in figure 1 is a simple one.*

Problem 4. *Prove that every simple pair shows up in figure 1 and that it occurs only once.*

After these problems it's quite natural to call the pattern shown in figure 1 the genealogical tree of simple pairs.

Problem 5. *Let the pair* (m, n) *lie on the genealogical tree in figure 1. There is a unique path connecting it to the first pair (1, 1). Show that moving along this path in the direction of the pair (1,1) is equivalent to applying Euclid's algorithm to the pair* (m, n).

So the genealogical tree contains all simple pairs. And Euclid's algorithm is applied by moving from the pairs (m, n) against the arrows.

Equation $XY = Z^2$

Now let's change the subject and try to find all integer solutions (X, Y, Z) of the equation

$$XY = Z^2. \qquad (2)$$

We'll be interested only in positive X, Y, Z. (Example: $X = 3$, $Y = 12$, $Z = 6$.)

At first glance it seems there's nothing to talk about. For a given Z we just have to break down the number Z into two factors. So for each Z the number of solutions can be computed quite easily.

Problem 6. *Fix* Z *and let* $Z = p_1^{a_1}, p_2^{a_2}, \dots, p_l^{a_l}$ *be the expansion of* Z *into prime factors. Prove that the number of solutions of equation (2) is equal to* $(2a_1 + 1)(2a_2 + 1) \dots (2a_l + 1)$.

I'd advocate another approach, though.

Problem 7. *Let* (X, Y, Z) *be a solution of equation (2); prove that for any* d > 0 *the triple* (dX, dY, dZ) *is also a solution. Let* (X, Y, Z) *be a solution of equation (2) and let the GCD of* (X, Y) *equal* d; *prove that* (X/d, Y/d, Z/d) *is also a solution of equation (2) and that the GCD of* (X/d, Y/d) *is 1.*

So in order to find all solutions of equation (2) we can consider only those triples (X, Y, Z) for which the GCD of (X, Y) equals 1. We'll call them "primitive solutions." All other solutions are obtained from primitive solutions by simple multiplication.

Problem 8. *Let* (X, Y, Z) *be a primitive solution of equation (2). Prove that there is a simple pair* (m, n) *such that* $X = m^2$, $Y = n^2$, $Z = mn$.

It's now clear that relations

$$X = m^2, Y = n^2, Z = mn \qquad (3)$$

define a one-to-one correspondence between simple pairs (m, n) and primitive solutions (X, Y, Z) of equation (2).

This means we can make full use of our preceding results about simple pairs. For example, by using relations (3) we can replace each simple pair (m, n) in figure 1 with the corresponding primitive solution (X, Y, Z). The resulting tree might naturally be called the genealogical tree of primitive solutions of equation (2). It contains all primitive solutions without exception.

There is, however, a more direct and convenient way to build up such a tree.

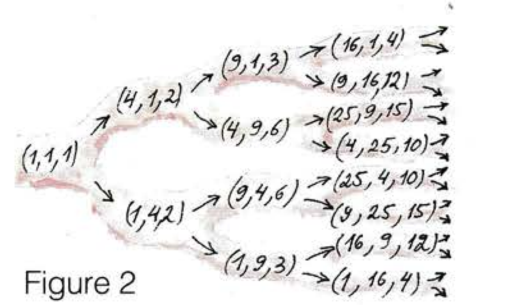


Figure 2

Problem 9. *Let a simple pair* (m, n) *correspond via (3) to the solution* (X, Y, Z). *Denote the solution corresponding to the pair* $t_1(m, n) = (m + n, n)$ *by* $T_1(X, Y, Z)$ *and the solution corresponding to the pair* $t_2(m, n) = (m, m + n)$ *by* $T_2(X, Y, Z)$. *Prove that*

$$T_1(X, Y, Z) = (X + Y + 2Z, Y, Y + Z),$$

$$T_2(X, Y, Z) = (X, X + Y + 2Z, X + Z).$$

The upshot is that the genealogical tree shown in figure 2 is directly generated by the transformations T_1 and T_2. We start with the obvious solution (1, 1, 1) and apply the transformations. The upward arrow corresponds to T_1 while the downward arrow denotes T_2.

Figure 2 is quite impressive and gives us a clear idea of the structure of the set of all primitive solutions of equation (2). But as I mentioned earlier, equation $XY = Z^2$ can be solved by a simpler approach. So let's move on to a more interesting example.

Pythagorean triples

Consider the equation

$$X^2 + Y^2 = Z^2.$$

Its positive integer solutions (X, Y, Z) are called Pythagorean triples. The first such triple is, of course, (3, 4, 5). Our goal is to construct a genealogical tree of Pythagorean triples similar to the tree in figure 2. How do we do that? Following the same approach as for the equation $XY = Z^2$, we have to (a) single out primitive triples from among all Pythagorean triples; (b) write out the relations $X = X(m, n)$, $Y = Y(m, n)$, $Z = Z(m, n)$ for primitive Pythagorean triples similar to relations (3); and (c) make the genealogical tree for pairs (m, n) and replace (m, n) with the corresponding triple $(X(m, n), Y(m, n), Z(m, n))$.

Steps (a) and (b) have been known for a long time. I'll give the necessary facts here without any comments or proofs.

A Pythagorean triple (X, Y, Z) is called primitive if the GCD of (X, Y) equals 1, X is odd, and Y is even. The triple (3, 4, 5), for example, is primitive.

It's known that for any Pythagorean triple (x, y, z) there exists a unique primitive Pythagorean triple (X, Y, Z) and a unique natural number d such that either $(x, y, z) = (dX, dY, dZ)$ or $(x, y, z) = (dY, dX, dZ)$. So having a list of all primitive Pythagorean triples makes it possible to list all the other Pythagorean triples as well.

A pair of integers (m, n) is called primitive if $m > n$, $n > 0$, the GCD of (m, n) is 1, and the numbers m, n have different parities (that is, one of them is even and the other odd). The pair (2, 1), for example, is primitive.

It's also known that the relations

$$X = m^2 - n^2,\ Y = 2mn,\ Z = m^2 + n^2 \qquad (4)$$

define a one-to-one correspondence between the set of all primitive pairs and the set of all primitive Pythagorean triples. For example, the pair (2, 1) generates the triple (3, 4, 5).

As for step (c) of our program, it has recently been carried out by the British mathematician A. Hall.

The genealogy of Pythagorean triples

In a brief note published in 1970 in the *Mathematical Gazette*, Hall proposed the following technique for constructing a genealogical tree for primitive pairs and primitive Pythagorean triples. He introduced three transformations t_1, t_2, t_3:

$$t_1(m, n) = (2m - n, m),$$

$$t_2(m, n) = (2m + n, m),$$

$$t_3(m, n) = (m + 2n, n).$$

By means of these transformations, starting from the pair (2, 1), the genealogical tree is built. Here the upward direction corresponds to transformation t_1, the horizontal direction to transformation t_2, and the downward direction to transformation t_3.

Problem 10. *Let a pair* (m, n) *create the Pythagorean triple* (X, Y, Z) *by means of relations (4). Designate the Pythagorean triple generated by the pair* $t_1(m, n)$ *as* $T_1(X, Y, Z)$, *the triple generated by* $t_2(m, n)$ *as* $T_2(X, Y, Z)$, *and the triple generated by the pair* $t_3(m, n)$ *as* $T_3(X, Y, Z)$. *Prove that*

$$T_1(X, Y, Z) = (X - 2Y + 2Z, 2X - Y + 2Z, 2X - 2Y + 3Z),$$

$$T_2(X, Y, Z) = (X + 2Y + 2Z, 2X + Y + 2Z, 2X + 2Y + 3Z),$$

$$T_3(X, Y, Z) = (-X + 2Y + 2Z, -2X + Y + 2Z, -2X + 2Y + 3Z)$$

Now using the transformations T_1, T_2, T_3 let's plot the genealogical tree starting with the triple (3, 4, 5).

Hall's remarkable result is that the tree in figure 3 contains all primitive pairs without exception; so the genealogical tree in figure 4 contains all primitive Pythagorean triples without exception.

The next series of problems proves this fact.

Problem 11. *Prove that all the pairs shown in figure 3 are primitive. This will show that all the triples in figure 4 are primitive Pythagorean triples.*

The transformations t_1, t_2, t_3 make it possible to move along the tree in figure 3 in the direction of the arrows.

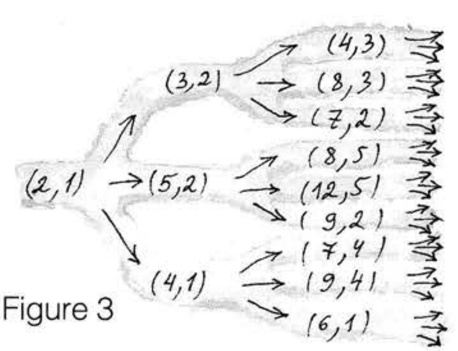


Figure 3

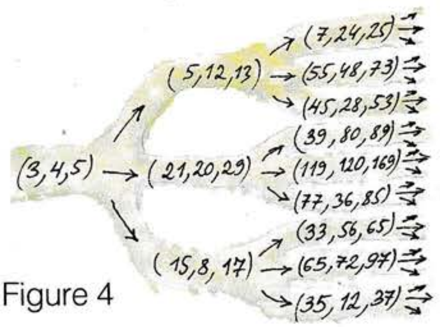


Figure 4

Now we'll find out how to move along the tree in the opposite direction.

Problem 12. *Let* (M, N) = t_i(m, n), *where* t_i *is one of the transformations* t_1, t_2, t_3. *Prove that* (m, n) = u_i(M, N), *where transformations* u_i *are defined by*

$$u_1(M, N) = (N, -M + 2N),$$
$$u_2(M, N) = (N, M - 2N),$$
$$u_3(M, N) = (M - 2N, N).$$

The transformations u_1, u_2, u_3 make it possible to move along the tree shown in figure 3 against the arrows. They carry out a peculiar Euclidian algorithm for primitive pairs, allowing a descent from an arbitrary primitive pair (m, n) to the initial pair $(2, 1)$.

Each pair (m, n) in figure 3 is approached by exactly one arrow. For the transformations u_1, u_2, u_3 this corresponds to the following fact.

Problem 13. *Let a pair* (M, N) *be primitive and* (M, N) ≠ (2, 1). *Prove that only one of the three pairs* (m, n) = u_i(M, N), i = 1, 2, 3, *is primitive. In addition,* m + n < M + N.

And, finally, the concluding problem.

Problem 14. *Prove that each primitive pair is contained in the genealogical tree shown in figure 3 only once and that the same holds for the primitive Pythagorean triples shown in figure 4.*

Other genealogies

In 1978 the Scandinavian mathematical journal *Normat* published an article by E. Selmer. In this paper he showed that there are two other genealogical trees containing all Pythagorean triples without either exception or repetition (fig. 5).

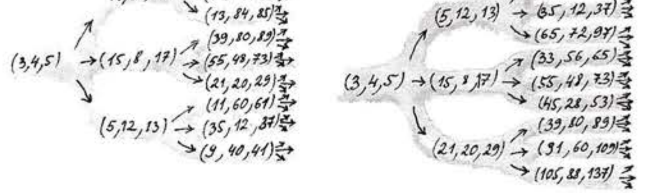


Figure 5

The first tree is built by using the transformations

$$T_1(X, Y, Z) = (2X - Y + Z, 2X + 2Y + 2Z, 2X + Y + 3Z),$$
$$T_2(X, Y, Z) = (2X + Y + Z, 2X - 2Y + 2Z, 2X - Y + 3Z),$$
$$T_3(X, Y, Z) = (2X + Y - Z, -2X + 2Y + 2Z, -2X + Y + 3Z).$$

The second tree is obtained from

$$T_1(X, Y, Z) = (X - 2Y + 2Z, 2X - Y + 2Z, 2X - 2Y + 3Z),$$
$$T_2(X, Y, Z) = (2X + Y + Z, 2X - 2Y + 2Z, 2X - Y + 3Z),$$
$$T_3(X, Y, Z) = (-2X + 3Y + 3Z, -6X + 2Y + 6Z, -6X + 3Y + 7Z).$$

Pythagorean triples have attracted the attention of mathematicians for thousands of years. But we can see that the subject certainly hasn't been exhausted, and interesting new facts continue to be discovered.

Summing up

Now a number of questions should at least be asked, if not answered. For instance, why does the genealogical tree fork into two branches for $XY = Z^2$ and into three branches for $X^2 + Y^2 = Z^2$? Next question: we've given three genealogical trees for the equation $X^2 + Y^2 = Z^2$, but only one for $XY = Z^2$; are there any other trees for these equations? Finally, how were the transformations T_1, T_2, T_3 generating Pythagorean triples found?

The genealogical tree for $XY = Z^2$ was constructed by directly applying Euclid's algorithm and looks sufficiently well motivated, which apparently isn't the case for the equation $X^2 + Y^2 = Z^2$. These two equations are, however, related. Indeed, writing equation $X^2 + Y^2 = Z^2$ in the form $Z^2 - X^2 = Y^2$, we can break down the left side into two factors: $(Z - X)(Z + X) = Y^2$. Substituting $U = Z - X$, $V = Z + X$, $W = Y$, we arrive at the equation $UV = W^2$. So there is a substitution that reduces the equation $X^2 + Y^2 = Z^2$ to the equation $XY = Z^2$.

A number of equations can be dealt with in the same way—that is, by finding a substitution that reduces them to the form $XY = Z^2$. Examples are the equations $X^2 + Y^2 = 2Z^2$ and $X^2 + 3Y^2 = Z^2$. You might try to construct genealogical trees for these equations as well.

Another remarkable equation should also be mentioned—Markov's equation:

$$X^2 + Y^2 + Z^2 = 3XYZ.$$

It has the property we're already used to: all its solutions, except the two obvious ones $(1, 1, 1)$ and $(2, 1, 1)$, are organized into a genealogical tree. (We make use of the fact that if a triple (X, Y, Z) solves Markov's equation, then the triple $(3YZ - X, Y, Z)$ is also a solution.) This tree is quite similar to the genealogical tree for the equation $XY = Z^2$. Is there anything connecting the two equations? What other equations have similar properties?

There's a lot here to think about. ◘

An incident on the train

Nothing out of Agatha Christie, but a mystery of sorts

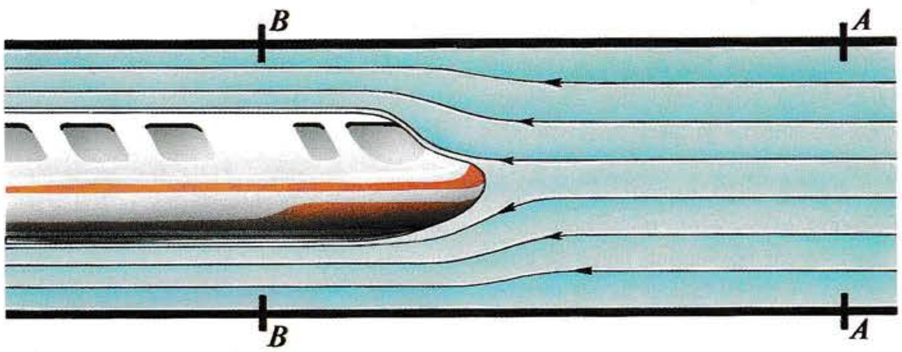
by Carlo Camerlingo (Italy) and Andrey Varlamov

NOT SO LONG AGO THE authors of these lines had to return from Venice to Naples on an express train. The train moved very fast (its velocity was approximately 150 km/h) and landscapes that looked like paintings by the Renaissance masters flitted by as we looked out the window. In exact agreement with its canvas-bound versions, the terrain was hilly, and we sometimes flew over a bridge or dove into a tunnel. In one of the especially long tunnels between Bologna and Florence, we suddenly felt a dull pain in our ears, as happens with passengers in airplanes taking off or landing. It was clear from external signs that the same sensation came over all of our fellow travellers: they all turned their heads, trying to get rid of the unpleasant feeling. But when the train finally burst from the narrow tunnel the unpleasantness passed, and only one of us, who wasn't used to such surprises on the railways, was interested in the origin of this phenomenon. Since it was evidently connected with the pressure difference, we began a lively discussion of the possible physical causes. At first glance it seemed to us that the air pressure in the gap between the tunnel walls and the train had increased in comparison with the atmospheric pressure, but there were qualitative reasons to expect the opposite effect as well. In such matters mathematics is the best judge, so we attempted to find some numerical answer to the problem. Soon the explanation was ready and it came down to this.

Let's consider a train with a cross-sectional area S_t that moves at velocity v_t in a long tunnel with a cross-sectional area S_0. First of all, let's switch to the inertial coordinate system associated with the train. We'll take the air flow as stationary and laminar, and we'll ignore its viscosity. The movement of the tunnel walls relative to the train need not be taken into account in this case—because of the absence of viscosity, it doesn't influence the air flow. We'll also consider the train sufficiently long so that we can ignore turbulence at the front and rear cars, and the air pressure in the tunnel will be taken as steady and constant along the entire surface of the train.

So by gradually eliminating minor details, we've moved from the actual movement of the train to a simplified physical model that we can try to describe mathematically. Here goes.

We have a long tube (formerly the tunnel) and a cylinder with streamlined ends (formerly the train) nestled in it coaxially. Air passes through this tube—away from the train (the cross section A–A in the figure) the air pressure p_0 equals the atmospheric pressure and the velocity of the air flow v_t is equal to the velocity of the train before it entered our system of calculation (but with the opposite sign). Let's examine a certain cross section B–B (just in case, we place B–B far from the ends of the train so our assumptions will actually bear out). We'll denote the air pressure in this cross section as p_B and the air velocity as v_B. These values can be linked with v_t and p_0 by means of the Bernoulli equation:



$$p_B + \frac{\rho v_B^2}{2} = p_0 + \frac{\rho v_t^2}{2}, \qquad (1)$$

where ρ is the density of the air.

Equation (1) has two unknowns, p_B and v_B, so to determine p_B we need another relation. This is provided by the condition of the conservation air mass that flows through any cross section of the tube in a unit of time:

$$\rho v_t S_0 = \rho v_B (S_0 - S_t). \qquad (2)$$

This equation expresses the fact that the air mass can neither appear nor disappear while it flows through the tube. It's usually called the condition of flow continuity.

As you probably noticed, we took the air density in equations (1) and (2) to be constant. This assumption is valid as long as the air velocities in different cross sections of the tube are much less than the mean square velocity of chaotic molecular motion; it's just this velocity that determines the characteristic time required to establish mean gas density on the macroscopic scale.

Getting rid of velocity v_B in equation (1) by means of equation (2), we get

$$p_B = p_0 - \frac{\rho v_t}{2}\left[\left(\frac{S_0}{S_0 - S_t}\right)^2 - 1\right]. \qquad (3)$$

The air density ρ can be expressed in terms of p_0 by the Mendeleyev–Clapeyron equation: $\rho = p_0\mu/RT$. After this substitution, we have

$$p_B = p_0\left\{1 - \frac{\mu v_t^2}{2RT}\left[\left(\frac{S_0}{S_0 - S_t}\right)^2 - 1\right]\right\}. \qquad (4)$$

In this expression there is a combination of parameters, $\mu v_t^2/RT$, that is evidently dimensionless. So the value $(RT/\mu)^{1/2}$ has the physical dimensionality of velocity. It's easy to recognize in it the mean square velocity of chaotic molecular motion (with an accuracy to one power). But in our aerodynamical problem another physical characteristic of gas is important: the velocity of sound propagation v_s in it. This value is determined by the same combination of temperature and molecular mass as the mean square velocity of molecular motion, but the numerical value of v_s depends additionally on the so-called adiabatic index γ, a characteristic number for every gas of the order of 1 (for air, $\gamma = 1.41$):

$$v_s = \sqrt{\gamma \frac{Rt}{\mu}}. \qquad (5)$$

Under normal conditions, $v_s \cong 1{,}200$ km/h. Using equation (5) we can rewrite the pressure expression (4) in final form, one that will be convenient for the discussion to follow (substituting $\mu/RT = \gamma/v_s^2$):

$$p_B = p_0\left\{1 - \frac{\gamma}{2}\left(\frac{v_t}{v_s}\right)^2\left[\left(\frac{S_0}{S_0 - S_t}\right)^2 - 1\right]\right\}. \qquad (6)$$

Now it's time to stop and think a little about this. We calculated the pressure near the train inside the tunnel. But our ears ached not because of the pressure itself but because of its change in comparison with the pressure p_B' when the train is in the open air.[1] We can easily determine this outside pressure directly from equation (6), noticing that the open air can be considered a tunnel with a cross-sectional area $S_0 \longrightarrow \infty$. So we have

$$p_B' = \lim_{S \to \infty} p_B(S_0)$$

$$= p_0\left\{1 - \frac{\gamma}{2}\left(\frac{v_t}{v_s}\right)^2\left[\left(\frac{1}{1 - \lim\limits_{S_0 \to \infty}\frac{S_t}{S_0}}\right)^2 - 1\right]\right\}.$$

This result was sufficiently evident without any calculation. It's interesting to observe that the relative pressure difference is

$$\frac{\Delta p}{p_0} = \frac{p_B - p_0}{p_0} = -\frac{\gamma}{2}\left(\frac{v_t}{v_s}\right)^2\left[\left(\frac{S_0}{S_0 - S_t}\right)^2 - 1\right]. \qquad (7)$$

From this expression we can see that when the train is entering the tunnel the pressure near it *decreases*, contrary to what we may have thought at first.

Now let's estimate the magnitude of this effect. As we mentioned earlier, $v_t = 150$ km/h, $v_s = 1{,}200$ km/h, and for narrow railroad tunnels the ratio S_t/S_0 can be estimated as 1/4 (in our tunnel there were two sets of rails). So

$$\frac{\Delta p}{p_0} \cong -\left(\frac{1.41}{2}\right)\left(\frac{1}{8}\right)^2\left(\frac{16}{9} - 1\right) \cong -1\%.$$

This value seems pretty small, but if we take into account that $p_0 = 10^5$ N/m² and take the area of the eardrum to be $\sigma = 1$ cm², we get an excess force $\Delta F = \Delta p\sigma \cong 0.1$ N, which may turn out to be quite noticeable.

So it seems the effect is explained, and we can call it quits. But something worried us about this last equation. Namely, from expression (7) it follows that even in the case of a normal velocity for an ordinary train $v_t \ll v_s$ (this combination of velocities is constantly encountered in aerodynamics and is called the Mach number), in sufficiently narrow tunnels the value $|\Delta p|$ may reach and even exceed the normal pressure p_0! Clearly, within the framework of our assumptions we're getting the absurd result that the pressure between the walls of the narrow tunnel and the train becomes negative!

But wait a minute! Maybe there's a breaking point in our result beyond which it ceases to be valid . . . Let's look at our findings a bit more closely.

If $|\Delta p| \cong p_0$, then

$$\frac{v_t}{v_s}\left(\frac{S_0}{S_0 - S_t}\right) \cong 1,$$

and so

$$v_t S_0 \cong v_s (S_0 - S_t).$$

Comparing the last equation with equation (2), we begin to understand the situation. If Δp reaches p_0, the velocity of air flowing in the gap between the train and the walls of such a narrow tunnel turns out to be of the order of the speed of sound, and we can't speak of laminar air flow here.[2] So the correct condition for applying equation (7) is not merely $v_t \ll v_s$ but the more rigid requirement

$$v_t \ll v_s\left(\frac{S_0 - S_t}{S_0}\right).$$

It's evident that for real trains and tunnels this condition is always met. Nevertheless, our investigation into the limited applicability of equation (7) isn't just an empty mathematical exercise. A physicist must always

[1] We should point out two circumstances here. First, there is the so-called Weber–Hefner law in biophysics, which says that any change in an external effect is apprehended by the body only when the relative change exceeds a certain threshold value. Second, if it so happens the tunnel is long enough, the body is able to adapt to the new conditions and the unpleasant sensation disappears. At the exit from the tunnel, however, it bothers you again.

[2] That is, the smooth flow becomes turbulent.

recognize the limits of the validity of any result obtained. But another reason for taking it seriously in our case is a quite practical one. In the last few decades fundamentally new forms of transportation, including high-speed trains, have been discussed more and more. One type of train moves on a magnetic cushion produced by a powerful superconducting magnet. (Such vehicles already exist. At last report, a prototype maglev (magnetic levitation) train in Japan can carry 20 passengers along 7 km of test track at a maximum speed of 516 km/h—that's almost half the speed of sound!) Since the vehicle hovers above the metal rails, resistance to its movement is determined solely by its aerodynamic properties.

The next step in developing this means of transportation was the idea of—believe it or not—enclosing the train in a hermetically sealed tube and reducing the pressure by pumping air out! You see how close this problem is to the one that captivated us. But here the physicists and engineers encounter the much more complex case in which $v_t \cong v_s$ and $S_0 - S_t << S_0$. The air flow here is far from laminar, and the air temperature changes considerably as the train moves. Modern science doesn't have the answers to all the questions generated in the pursuit of solutions to these problems. But even our simple estimate allows us, in principle, to estimate the threshold where such effects become important.

We'd like to leave you with a few questions about physics that might pop up on a train ride.

1. Why do the windows rattle when you're racing along at a nice clip and another fast train passes you going in the opposite direction? Is the force responsible for shaking the windows directed inward or outward?

2. Why does the noise from a moving train increase considerably when it enters a tunnel?

3. Which of the two rails of a rail line built along a meridian is worn down faster in the Northern Hemisphere? Southern Hemisphere?

CONTINUED FROM PAGE 35

When days are months

We received a number of correct solutions to the contest problem in the May 1990 issue asking how long a day would be when the length of a day is equal to the length of the month. The solution we present here is very similar to one by Earle Wallingford of Bozeman, MT. Similar solutions were submitted by Steve Fung (TX), Jason Jacobs (NY), and Mark Roseberry (KY).

Each will receive a free subscription to *Quantum* for one year.

In our solution, we first compare the angular momenta of the Moon revolving around the Earth, now and at the time when the length of the day is equal to the length of the month. We then turn our attention to the comparison of the Earth's angular momenta at these times as it spins on its axis. Finally, we apply the law of conservation of angular momentum to solve the problem.

As Jason Jacobs pointed out, the assumption that the Moon's orbit is in the plane of the Earth's equator means that all of the angular momentum vectors point in the same direction and we need only work with their magnitudes. The angular momentum L_m of the Moon due to its orbit about the Earth is given by

$$L_{mi} = M_m R^2_{mi} \frac{2\pi}{T_{mi}},$$

$$L_{mf} = M_m R^2_{mf} \frac{2\pi}{T_f},$$

where M_m is the mass of the Moon, R_m is the radius of the Moon's orbit, and T_{mi} and T_f are the Moon's initial and final orbital periods. Taking the ratio of the two expressions, we get

$$\frac{L_{mf}}{L_{mi}} = \frac{R^2_{mf}}{R^2_{mi}} \frac{T_{mi}}{T_f}.$$

It's important to realize that since the Moon is in orbit, it must obey Kepler's third law, which tells us that the square of the period is proportional to the cube of the radius. This can be derived by recognizing that the gravitational force of the Earth provides the centripetal force on the Moon. This gives us an expression for the ratio of the initial and final radii and allows us to write the ratio of the angular momenta in terms of a ratio of the periods:

$$\frac{L_{mf}}{L_{mi}} = \left(\frac{T_f}{T_{mi}}\right)^{\frac{4}{3}} \frac{T_{mi}}{T_f} = \left(\frac{T_f}{T_{mi}}\right)^{\frac{1}{3}}.$$

We now obtain the ratio of the initial L_{ei} and final L_{ef} angular momenta of the Earth spinning on its axis in terms of the mass M_e of the Earth, the radius R_e of the Earth, and the Earth's initial T_{ei} and final T_f rotational periods:

$$L_{ei} = \frac{2\pi I}{T_{ei}},$$

$$L_{ef} = \frac{2\pi I}{T_f}$$

(where $I = (2/5)M_e R_e^2$), which yields

$$\frac{L_{ef}}{L_{ei}} = \frac{T_{ei}}{T_f}.$$

The conservation of angular momentum can now be written as

$$L_{mi} + L_{ei} = L_{mf} + L_{ef}$$

$$= L_{mi}\left(\frac{T_f}{T_{mi}}\right)^{\frac{1}{3}} + L_{ei}\frac{T_{ei}}{T_f}.$$

Solving for the ratio of the initial angular momenta of the Moon and Earth, we get

$$\frac{L_{mi}}{L_{ei}} = \frac{1 - \frac{T_{ei}}{T_f}}{\left(\frac{T_f}{T_{mi}}\right)^{\frac{1}{3}} - 1}$$

$$= \frac{5 M_m R_m^2 T_{ei}}{2 M_e R_e^2 T_{mi}}$$

$$= 4.08\,.$$

Notice that we only need to know the ratios of the masses, periods, and radii to get the numerical value. If all times are expressed in terms of current Earth days, we get

$$5.08 T_f - 1 = 1.35 T_f^{\frac{4}{3}},$$

which can be solved with graphical or numerical techniques to obtain a value of $T_f = 53$ days. To solve the problem graphically, you plot each side of the equation for various assumed values of T_f and find where the two curves intersect.

This problem was inspired by a statement in *Exploration of the Universe* (5th ed., 1987) by Abell, Morrison, and Wolff that the period would be 47 days.

Arthur Eisenkraft *is the chair of the science department and physics teacher at Fox Lane High School in Bedford, NY.* **Larry D. Kirkpatrick** *is a professor of physics at Montana State University in Bozeman. Drs. Eisenkraft and Kirkpatrick serve as academic directors for the US Physics Team that competes in the International Physics Olympiad.*

CONTINUED FROM PAGE 34

David Watson (NY), Tim Kokesh (OK), Tim Hollebeek (PA), Kiran Kedlaya (MD), Brian Platt (UT), John Stafford (NC), Sergey Levin (RI), Andrew Dittmer (VA), Peter Kramer (NJ), John Clemens (IL), and Mark Roseberry (KY), after whom the conjecture was named and who is presently a freshman at Rose-Hulman. Each of them approached the problem somewhat differently; unfortunately, space limitations don't allow a complete reproduction of their results. Most of them treated the case of $n = 2^k$ separately by constructing (via mathematical induction) a k-digit multiple of n consisting entirely of 6's and 7's. (The same procedure can be applied to other pairs of digits if they're of different parity). Then, to resolve the case of $n = m(2^k)$, where m is odd and not a multiple of 5, they concatenated the multiple of 2^k obtained earlier with itself an appropriate number of times. This "appropriate number" can be shown to be less than or equal to m by yet another clever application of the pigeonhole principle (see *Quantum* Jan. 1990 and Sept./Oct. 1990). Congratulations to all of the successful students named above.

George Berzsenyi *is the chair of the Department of Mathematics at Rose-Hulman Institute of Technology in Terre Haute, IN.*

Why are the cheese holes round?

Maybe you've forgotten you ever wondered . . .

by Sergey Krotov

. . . in the middle of this place was a large oak-tree, and, from the top of the tree, there came a loud buzzing-noise.

Winnie-the-Pooh sat down at the foot of the tree, put his head between his paws and began to think.

*First of all he said to himself: "That buzzing-noise means something. You don't get a buzzing-noise like that, just buzzing and buzzing, without its meaning something. If there's a buzzing-noise, somebody's making a buzzing-noise, and the only reason for making a buzzing-noise that **I** know of is because you're a bee."*

Then he thought another long time, and said: "And the only reason for being a bee that I know of is making honey."

*And then he got up, and said: "And the only reason for making honey is so as **I** can eat it." So he began to climb the tree.*

A.A. Milne, "Winnie-the-Pooh"

HAVE YOU EVER THOUGHT why Winnie-the-Pooh is so lovable? Maybe because he reminds us of ourselves when we were little and asked so many silly questions (silly to grownups, anyway) and wanted the answers right away. But it's good to ask questions at any age. And it's especially useful when you're learning physics. Let's try it together and maybe you'll see it the same way I do.

Have you ever come across the fairy tale "Two Greedy Little Bears"? I'll never forget the colorful drawings of a cheese wheel vanishing before your eyes. The cheese was covered with a bright-red coating and was awfully "holey" inside. The holes were perfectly round and practically identical in size. Years have passed since then, but only recently did I figure out that this hole-ridden structure of cheese is due to one of the most fundamental laws of nature—Pascal's law. I'll remind you what it says: "Pressure applied to a liquid or gas is transmitted equally to all its parts." The leading role here is played by pressure. So let's discuss this notion first.

Do you remember the sad fairy tale "Gray Neck Swan," in which a crafty fox crawls onto a frozen pond where Gray Neck is swimming? Aware of the danger of breaking the thin ice, the fox sprawls on the surface, stretching out as much as it can. The force acting on the ice doesn't depend on the body's position, right? The fox isn't any lighter when it lies down than when it stands up, is it? Isn't there a contradiction here? Not at all. As it turns out, what matters is the surface area affected by the force of pressure. If the area of contact between the fox and the ice is increased, the force bending the ice is reduced and the fox moves on it safely. (The fox was crafty and knew all this.) To describe this and many other phenomena it's not enough to know only the overall force of pressure (the force with which bodies in contact affect each other); we have to know the force applied to each unit area of the contacting surface. It's this force that's called the "pressure."

Can you think of another tale in which everything (from the physical point of view) depends on pressure? It's Hans Christian Andersen's "Princess and the Pea." Why did a dried pea in her bed make the princess so uncomfortable? Again, it's all a matter of pressure. Obviously, both with and without the pea the overall force holding the princess on her bed is the same. But if a protruding object appears on the bed, the pressure at this point increases sharply, which immediately spoils the princess's mood. She could even develop insomnia. Surely you don't need to be a princess to detect a hard pea in your bed. Even a shepherd can do that. But to feel a pea through several layers of down mattresses (there were twelve of them in the story) requires a genuine royal sensibility.

So the pressure is defined as the ratio of the force acting perpendicularly to a surface to the total area of the surface. But Pascal's law apparently involves another kind of pressure—the pressure inside a liquid or a gas. All the points inside a liquid somehow "know" that it is being compressed from outside. In other words, pressure applied to the outer surface of the liquid is transmitted from point to point equally in all directions. And this is, in fact, an essential property of a liquid. That's how it's "constructed."

Let's discuss this fact in a bit more detail. Take a soft spring—for instance, a spring from an air gun. If you lay it on a table, the distance between adjacent coils is the same along the entire length of the spring. But if you stand it upright, the coils start to "fall

Art by Dmitry Krymov


Крымов

down" (because of the force of gravity), moving closer together. Eventually, different sections of the spring will be compressed to varying degrees: the lower the coils, the smaller the distance between them. What's going on here? The mutual displacements of the coils produce elastic forces in the spring. The lower the coils, the greater the portion of the spring's weight they carry and, consequently, the greater the compression they receive. So the pressure in various sections of the spring is different. If you want to visualize the pressure pattern inside a body, squeeze a foam sponge in your hand. Some parts of the sponge get compressed, while the others get stretched. The greater the compression at a specific point, the smaller the pores. So we can estimate the internal pressure in a spring by the distance between adjacent coils, and in a sponge by the pore size.

Unlike solid bodies, both liquid and gas are usually subjected to compression only. If an impermeable casing is filled with a liquid and then compressed, the liquid is compressed equally throughout the entire volume (we ignore gravity), and we can't distinguish one point inside the container from another. It's important that regardless of the shape of the outer surface, the pressure is transmitted equally from any point to all adjacent points.

In order to make this idea more obvious, I'm afraid I'm going to have to dredge up some memories that are probably not among your happiest. I'm talking about injections. Yes, "shots." No doubt you remember that before making an injection the doctor presses the syringe piston and healing liquid squirts out of the needle. Imagine now that someone has punched small holes all over the surface of the syringe and stuck needles in them. The resulting object would resemble a porcupine. If we now press the piston of the syringe-porcupine, the jets spurting out of needles positioned at the same height will be identical. This is because the liquid's behavior is governed by Pascal's law. The liquid is pushed out of holes positioned at the same height with the same force. For holes positioned at different heights, we have to take into account the force of hydrostatic pressure.

To compare the elastic properties of a liquid with those of a solid body, let's take another example. Mentally put a spring inside a narrow container (so that the diameter of the spring coincides with the inner diameter of the container) and fill another container of the same size with water. Now imagine that the walls of both containers suddenly disappear. What happens to the spring and the water? The spring stays where it was as if nothing had happened. The water flies off in all directions like a popped soap bubble. Why? Because liquids and solid bodies have different ways of transmitting pressure. A spring transmits the pressure along its length only, practically speaking, while the water transmits it equally in all directions: up, down, and sideways, in accordance with Pascal's law.

It just so happens that a similar scenario was observed by Pascal himself when he discovered the law. His classic experiment was similar to our mental experiment with the syringe. True, in Pascal's experiment the walls of the container (a barrel) didn't disappear, they were broken. The shape of the resulting "fountains" depends on the pressure in various parts of the liquid. Now we can easily explain the "action" of a down mattress. It's like a heap of little springs oriented randomly relative to one another. Each spring transmits the pressure along its length but, because of the chaotic positioning, the pressure exerted by the pea is transmitted to . . . But I don't want to deprive you of the fun of finding the right answer. I'll just tell that in spite of all your exertions and attempts, royal intuition enabled the princess to unerringly discover any dirty trick, even if it was perpetrated by some big shot who knows a little physics.

And now the time has come to answer the main question of this article. (You haven't forgotten it, have you?) Let's briefly review how cheese is made—or, to be more precise, how holes in cheese are made. First, the cheese "dough" is prepared. Then it's compressed at high pressure and put in special molds. The wheels of cheese are taken out of the molds and left in a warm place for ripening. This is when the process of "fermentation" begins. Carbon dioxide gas is created inside the compressed dough. This results in the formation of bubbles. The more carbon dioxide, the larger the bubbles. (Don't forget that at this stage the inside of the future cheese is a soft, homogeneous mass.)

When the cheese gets harder, the pattern of the internal "breathing" of the fermenting cheese is recorded by the carbon dioxide bubbles. As for the shape of the cavities, because of Pascal's law the pressure inside the bubbles is transmitted equally in all directions since the dough resembles a liquid in its elastic properties. So the bubbles acquire a strictly spherical shape. Violation of this rule would have meant that there were areas of greater rigidity or, conversely, cavities inside the cheese. The harder the cheese, the less the bubbles inside blow up, so the holes are smaller. Some varieties of cheese are made without compression at the beginning of the process; carbon dioxide is released into cavities already present in the dough. As a result, you get an irregular pattern of frozen bubbles whose harmony can be understood only by a cheese expert.

So you see how many small questions we had to ask ourselves to answer a single big one: "Why does cheese have round holes?"

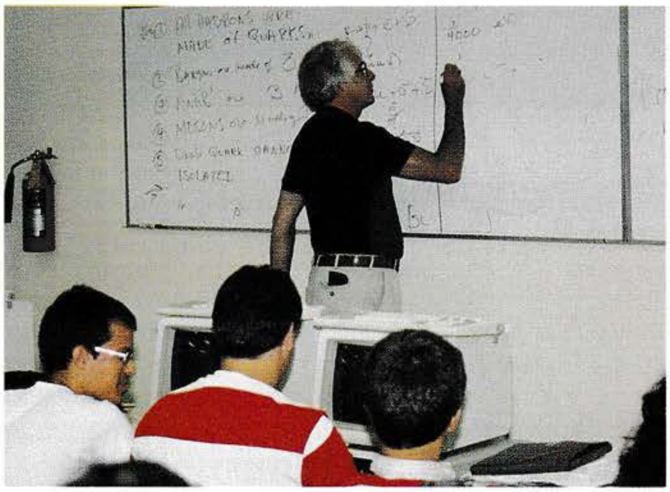
"Hallo, Pooh," said Rabbit.
"Hallo, Rabbit," said Pooh dreamily.
"Did you make that song up?"
"Well, I sort of made it up," said Pooh. "It isn't Brain," he went on humbly, "because You Know Why, Rabbit; but it comes to me sometimes."
A.A. Milne, "The House at Pooh Corner"

The Sky's *Not* the Limit!

The year 1992 has been declared the International Space Year (ISY) by the United Nations. Scientists from many countries will meet at various conferences, seminars, and symposiums to discuss the future of international cooperation in space. We hope there will be many new agreements on joint projects, including perhaps one about a joint Mars mission. All these projects will need many new researchers. Many of them will be among those who are presently going to high school. For this reason work with youth has been an important part of the ISY. One of the projects under development is the 1992 International Space Olympiad in Washington, DC.

Summer study in the USSR and US

To prepare for this olympiad, several American and Soviet organizations, including the magazines *Kvant* and *Quantum*, the US International Space Year Association, the Soviet Aerospace Society "Union," the National Science Teachers Association, and the International Educational Network, have decided to organize an International Summer Institute in the summer of 1991 in the United States and the Soviet Union. The program will feature **advanced classes** in mathematics, physics, biology, and other space-related subjects; lectures by **prominent scientists**; trips to major **scientific laboratories**; **sports and recreation**; and many **cultural activities**.



Nobel Laureate Sheldon Glashow of Harvard University instructs participants in a previous International Educational Network summer camp.

Three-stage competition

Sixty students from the US and 60 from the USSR will be selected, and we expect that students from other countries will also be interested in participating. The selection process will be based on the results of a three-stage competition. The questions for the first round are printed below. The second round will also be by correspondence and will include two math and two physics problems related to space. A total of 300 students will be invited to participate in the third round, which will be given at local universities or schools in the presence of the organizers' representatives.

Three-week program

The winners will participate in either the American or the Soviet part of the program, which will each last three weeks. The American session will take place July 1–21, 1991, while the Soviet session will take place August 1–21, 1991. Each session will feature two weeks of study and one week of travel in the host country. The winners of the competition, depending on their total score, will receive **scholarship prizes and awards** that will cover all or part of the program costs.

To enter the competition, please fill out the form and mail it, along with your answers to the questions printed below, postmarked no later than December 31, 1990, to:

Dr. Edward Lozansky, President
International Educational Network
3001 Veazey Terrace, NW
Washington, DC 20008
(Telephone: 202 362-7855)

Yes, I am interested in the 1991 International Summer Institute!

Last name ____________________ First name ____________________

Home address ____

City ____________________ State ________ Zip ________ Birthdate ________ Sex ______

Home phone (______) ____________________ Parent's office phone (______) ____________________

School name ____________________

School address ____________________

____________________ Phone (______) ____________________

Name of math or science teacher who can recommend you ____________________ (print first and last name)

Please answer the following questions:

1. *When was the first manned space ship launched? Who piloted this ship?*
2. *Who was the first man on the moon?*
3. *Name all American and Soviet women who have been in space.*
4. *Write a short essay explaining why you would like to participate in this program.*
5. *Could you write this essay with a ball point pen while orbiting the Earth? Explain.*

Teachers are encouraged to copy this page and distribute it to potential participants.

11th Tournament of Towns

The latest problems from this friendly intercity rivalry

THE TOURNAMENT OF Towns, an international competition in mathematical problem solving, continues to grow in popularity. You may have read about it in the first issue of *Quantum* (January 1990). Well, here are the problems from the last tournament, held in spring of this year. We hope you find them attractive and instructive. If you do, join the Tournament of Towns! Write to N. Konstantinov, USSR, 103006, Moscow K-6, Gorkogo 32/1, *Kvant* magazine. (Our phone number is 095 250-4111, and our fax number is 095 251-5557.)

Junior grades (ages 13 to 15)

O-level (beginners)

1. For every natural n prove the equality

$$\left(1+\frac{1}{2}+\ldots+\frac{1}{n}\right)^2+\left(\frac{1}{2}+\ldots+\frac{1}{n}\right)^2+\ldots$$
$$+\left(\frac{1}{n-1}+\frac{1}{n}\right)^2+\left(\frac{1}{n}\right)^2=2n-\left(1+\frac{1}{2}+\ldots+\frac{1}{n}\right).$$

2. Two circles c and d are plotted on the plane, one outside the other. Points C and D are the most distant points of these circles. Two smaller circles are constructed inside c and d: the first circle touches c and the two tangents drawn from C to d; the second touches d and the two tangents from D to c. Prove that the smaller circles are equal.

3. Is it possible to compose a 3x3x3 cube out of twenty-seven 1x1x1 cubes, 9 of which are red, 9 blue, and 9 white, so that the little cubes in each row (parallel to an edge of the big cube) are of two different colors?

4. In a set of 61 coins that look alike, 2 coins are counterfeit and the rest are genuine. The counterfeit coins weigh the same but their weight differs from that of a genuine coin. How can one tell whether a counterfeit coin is heavier or lighter than a genuine one by three weighings on a pan balance? (It's not necessary to identify the counterfeit coins.)

A-level (main variant)

5. Find the maximum number of parts into which the plane Oxy can be divided by 100 graphs of different quadratic functions of the form $y = ax^2 + bx + c$.

6. A square is rotated 45° about its center. The sides of the rotated square divide each side of the initial one in the ratio $a{:}b{:}a$ (which is easy to calculate). Take an arbitrary convex quadrilateral, divide its sides in the same ratio $a{:}b{:}a$, and construct a new quadrilateral whose sides pass through the corresponding pairs of division points like the sides of the rotated square described above. Prove that two such quadrilaterals have equal areas.

7. Fifteen elephants stand in a row. Their weights are expressed by integer numbers of kilograms. The sum of the weight of each elephant (except the last one) and the doubled weight of the elephant to its right is exactly 15 metric tons. Find the weight of each elephant.

8. Let $ABCD$ be a rhombus, P a point on its side BC. The circle passing through A, B, P meets line BD again at point Q, and the circle passing through C, P, Q meets BD again at point R. Prove that A, R, and P lie on one straight line.

9. Find the number of pairs (m, n) of positive integers, both not greater than 1,000, such that

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

(recall that $2^{1/2} = 1.414213\ldots$).

10. Let's call a collection of natural numbers "basic" if their sum is 200, and every positive integer not greater than 200 can be represented as a sum of some numbers from the collection, the representation being unique up to the order of summands. (A trivial basic collection consists of 200 units.)

(a) Find a nontrivial example.

(b) How many different basic collections are there?

Senior grades (ages 15 and older)

O-level (beginners)

11. Construct a triangle given its two sides if it's known that the median drawn from their common vertex divides the angle between them in the ratio 1:2.

12. Prove that (a) for any $n = 4k + 1$ ($k = 0, 1, 2, \ldots$) there exist n odd natural numbers whose sum is equal to their product; (b) for any other natural n such a set of odd numbers does not exist.

13. (a) Some vertices of a dodecahedron must be marked so that each face contains a marked vertex. What is the smallest number of marked vertices for which this is possible?

(b) The same question for an icosahedron.

(Recall that a dodecahedron has 12 pentangular faces meeting three at each vertex; an icosahedron has 20 triangular faces meeting five at each vertex.)

14. Substitute 103 for 61 in problem 4.

A-level (main variant)

15. Prove that for all natural n there exists a polynomial $P(x)$ divisible by $(x-1)^n$ such that its degree is less than 2^n and all of its coefficients are equal to $1, 0$, or -1.

16. Substitute 500 for 200 in problem 10.

17. Either p or q guests are expected to visit a birthday party; p and q are coprimes. What is the smallest number of slices (not necessarily equal) into which a birthday cake must be cut in advance so that in both cases every guest gets an equal share of the cake?

18. Let $ABCD$ be a trapezoid, H the midpoint of its base AB, and $AC = BC$. Let a line l passing through H cut line AD at P and line BD at Q. Prove that the angles ACP and QCB are equal or their sum equals $180°$.

19. Does there exist a convex polyhedron having a triangular section (by a plane not passing through the vertices), each vertex of which is a meeting of (a) no less than 5 faces? (b) exactly 5 faces?

20. A square sheet of paper with side a is covered with blots, each of area less than 1, so that any straight line parallel to the edges of the sheet crosses one blot at most. Prove that the total area of the blots is less than a.

SOLUTIONS ON PAGE 61

. . . HAPPENINGS

AHSME—AIME—USAMO—IMO

Is this just alphabet soup to you?

IF THE ACRONYMS IN THE TITLE aren't familiar to you, your school may not be as progressive as you would like to believe, and it needs your help! More specifically, you should find out who's in charge of mathematical competitions at your high school, call that person's attention to this article, and make absolutely certain she/he follows up on it. Our country is in dire need of future scientists, mathematicians, and engineers—they're sitting in our classrooms, waiting for encouragement to develop their talents toward such careers. But first, they need to be recognized. The competitions listed in the title will help you in that task, so please take advantage of the opportunities they offer.

After identifying these competitions, I'll briefly describe them. For more details, you should contact Dr. Walter E. Mientka, the Executive Director of the American Mathematics Competitions, at the Department of Mathematics and Statistics, University of Nebraska, Lincoln, NE 68508-0322. His telephone number is 402 472-2257. Walter is a staunch supporter of mathematics education at all levels, and he is one of the nicest gentlemen in mathematical circles, whom I strongly recommend to all of my readers. It should also be noted that all these competitions are sponsored not only by the Mathematical Association of America but also by the following organizations: Society of Actuaries, Mu Alpha Theta, National Council of Teachers of Mathematics, Casualty Actuarial Society, American Statistical Association, American Mathematical Association of Two-Year Colleges, American Mathematical Society.

AHSME = American High School Mathematics Examination

This is a multiple-choice examination; the students are given 90 minutes to solve 30 problems. The 42nd annual AHSME will be administered at the high schools on Tuesday, February 26, 1991; the deadline for registration is December 7, 1990, but late registrations (within reason) are usually accepted. Last year over 394,000 students from 6,411 schools participated in the AHSME. These are impressive numbers, but there is much room for well-deserved growth. The main purpose of the AHSME is to discover talented students, so it should be administered at every high school in the US. The $15 registration fee entitles each school making a report on three or more students to one copy of the Solutions Pamphlet, an Intramural Award (pin or medal), and a Summary of Rewards and Awards. The Examinations are sold in bundles of 10 for $7.50 per bundle.

AIME = American Invitational Mathematics Examination

Students who score at least 100 on the AHSME are automatically invited to the AIME, which consists of 15 answer-oriented problems, with each correct answer being an integer between 0 and 999. Unlike the AHSME, there's no penalty for wrong answers. It's also administered at the high schools. The number of participants varies from year to year, depending on the difficulty of the AHSME. The AIME is a three-hour examination, and there is no charge for participating in it.

USAMO = USA Mathematical Olympiad

Based on a weighted average, the top-scoring students of the AHSME/AIME are invited to the USAMO, which is also administered at the high schools. The time limit in the USAMO is 3½ hours; the students are expected to provide complete answers to five problems within that time. Generally, about 150 students take part in the USAMO, whose eight winners are properly recognized in splendid ceremonies in Washington, DC, each year.

IMO = International Mathematical Olympiad

The IMO was started in 1959; the US has been participating in it since 1975. At the 31st IMO, held in 1990 in Beijing, China, a total of 54 countries participated, most of them with complete teams of six members. The students usually have 4½ hours on each of two consecutive days to solve six problems, each worth 7 points. With 174 points, the US team finished in third place this year.

As outlined above, the first stage in this pyramid of mathematical competitions is the AHSME. Without entering this examination, nobody can advance to the higher levels. Most capable students can only benefit from the excellent problem-solving activities generated by these competitions. There are many more than 400,000 of them—that is to say, *you*—in this great country of ours. My own estimate is about 100 times that figure!

—George Berzsenyi

Bulletin Board

Computer tutor for calculus

Broderbund Software has released its tutoring program *Calculus* for IBM/Tandy computers with Microsoft Windows. Previously available only for the Macintosh, the program can serve as an extension of classwork, a refresher course, or a private tutor. *Calculus* brings abstract mathematical formulas to life via a special module which animates, demonstrates and explains the sequence of operations required to solve basic calculus problems. Since the program moves at the student's own pace, it's equally useful for those who need tutoring as for those who want to accelerate their learning. The program requires an IBM/Tandy (or compatible) computer with 640K of memory and a hard disk. A mouse is recommended. For information on ordering, write Broderbund Software, Inc., 17 Paul Drive, San Rafael, CA 94903-2101, or call 415 492-3200.

"Scientific American Frontiers" premieres

In October the Public Broadcasting System premiered "Scientific American Frontiers," a series provided to students through a coordinated school outreach program. Underwritten by GTE Corporation, Scientific American Frontiers will air one hour per month until February 1991, offering innovative, amusing, informative, and unusual science features. The season premier featured roller coaster technology, among other topics. Teachers may videotape the show and create their own science video library with the available SAF classroom materials. Scientific American Frontiers is produced in association with *Scientific American* magazine, and replaces the PBS series "Discover: The World of Science." For information on how to receive the free classroom materials, call toll free 800 523-5948, or write on school letterhead to Scientific American Frontiers School Program, 10 North Main Street, Yardley, PA 19067-9986.

Computer path to math

If you are working toward a career in any math-related science, you may be interested in *Mathematica*, Wolfram Research's general system for doing mathematics by computer. Designed for the Macintosh, *Mathematica* allows students to perform with ease the computational tasks required in mathematics, engineering, statistics, physics, chemistry, economics—any coursework that involves mathematical computation. The system will help students in algebra, integration, differentiation, matrices, and many other numerical computations, giving the student more time to delve into the conceptual issues of the problems.

Wolfram Research is now offering *Mathematica* to students at the reduced rate of $139 (72% off the retail price). Students who take advantage of this special offer will receive Stephen Wolfram's book *Mathematica: A System for Doing Mathematics by Computer*, as well as user manuals and an installation guide. Four megabytes of RAM are recommended. For more information, or to receive an order form, call toll free 800 441-MATH (6284).

Math

M16

The colony will perish. Let V_t and B_t be the numbers of viruses and bacteria t minutes after infection. Then $V_{t+1} = 2V_t$ and $B_{t+1} = 2(B_t - V_t)$. For the ratio of these quantities we have $B_{t+1}/V_{t+1} = B_t/V_t - 1$; therefore, $B_t/V_t = n - t$ ($B_0/V_0 = n$). So the last bacterium will be killed during the nth minute.

M17

The plainest solution to this very simple problem is to show that triangle AKD = triangle CDH ($AD = CH$, $AH = DC$, angle DAH = angle DCH). To be strict, however, a more detailed consideration of the equality of angles in different cases is needed (see figure 1).

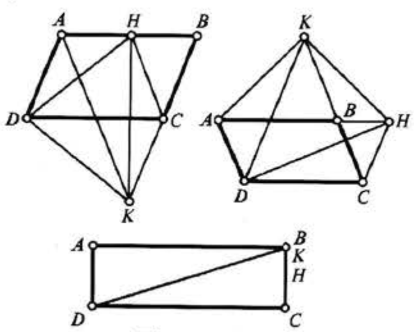


Figure 1

A solution that's valid for all cases at once and gives us some supplementary information involves vectors. Denote by R the rotation of an arbitrary vector through angle α = angle HCB (evidently, α = angle BAK, too). Then vector $\mathbf{DK} = \mathbf{DA} + \mathbf{AK} = \mathbf{BC} + R(\mathbf{AD}) = R(\mathbf{CH}) + R(\mathbf{DC}) = R(\mathbf{CH} + \mathbf{DC}) = R(\mathbf{DH})$ (fig. 2). Therefore, $DK = DH$ and, moreover, angle $HDK = \alpha$, which means that triangle DHK is similar to both triangle CHB and triangle ABK.

Finally, a third solution, using line reflections, should be mentioned. The diagonal AC is symmetrical to segments DH and DK with respect to midperpendiculars of the sides CD and DA, respectively. So $DH = AC = DK$.

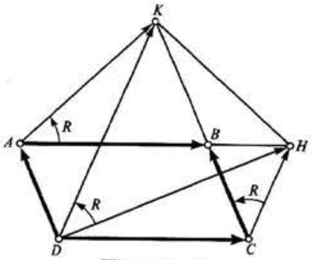


Figure 2

M18

If $c_1, c_2, ..., c_7$ are the counterfeit coins and $g_1, g_2, ..., g_7$ are the genuine ones, then the order of weighings can be as follows:

1st weighing: c_1 against g_1;

2nd weighing: g_1, c_2, c_3 against c_1, g_2, g_3;

3rd weighing: $g_1, g_2, g_3, c_4, c_5, c_6, c_7$ against $c_1, c_2, c_3, g_4, g_5, g_6, g_7$.

Each time the set of coins mentioned first turns out to be lighter, and so it contains more counterfeit coins than the second set. This leads successively to the conclusion that the coins (1) c_1, (2) c_2, c_3, (3) c_4, c_5, c_6, c_7 are counterfeit.

The method is easily generalized: to confirm that a given n coins are counterfeit and the other n are genuine, we need no more than $\log_2 n + 1$ weighings (the notation $\log_2$ stands for the greatest integer function). This mode of expertise is very economical (only 10 weighings are required for n = 1,000), though we can't prove it to be optimum. It would be interesting to prove that the minimal number of weighings grows unboundedly (or to refute it).

M19

A diameter crosses a chord if and only if one of its ends lies on the minor arc subtended by the chord. If it does, the other end lies on the arc symmetrical to the first with respect to the center of the circle (fig. 3). Consider such pairs of arcs for all the given chords. If the total length of the chords is not less than πk, the total length of all the arcs is greater than $2\pi k$, k times the length of the circumference. So there exists a point on the circumference covered with more than k arcs. The diameter drawn from this point will intersect at least $k + 1$ chords.

It's easy to construct a set of chords that satisfies the condition of total length arbitrarily close to πk: we can approximate half the circumference with a set of disjoint chords (fig. 4) and take each of them k times.

M20

(a) Notice first that $S(8 \cdot 125) = S(1{,}000) = 1 = S(125)/8$. We'll need the following properties of the function $S(A)$:

(1) $S(A + B) \leq S(A) + S(B)$,
(2) $S(A_1 + ... + A_n) \leq S(A_1) + ... + S(A_n)$,
(3) $S(nA) \leq nS(A)$,
(4) $S(AB) \leq S(A)S(B)$.

To verify (1) it suffices to inspect the very process of addition of A and B digit by digit. Property (2) follows from (1) by induction, (3) is a particular case of (2). Finally, if $A = a_n 10^n + a_{n-1}10^{n-1} + ... + a_0$, then, by (2) and (3),

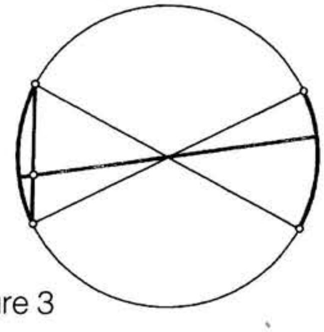


Figure 3

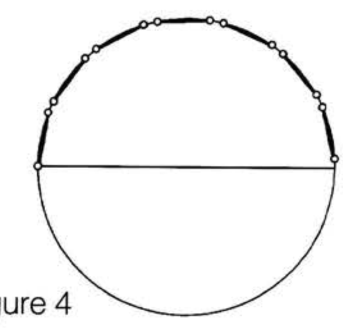


Figure 4

$S(AB) \le S(a_nB) + \ldots + S(a_0B) \le a_nS(B) + \ldots + a_0S(B) = S(A)S(B)$.

Now the required inequality is quite easy to prove:

$$S(N) = S(1{,}000N) = S(125 \cdot 8N) \le S(125)S(8N) = 8S(8N).$$

(b) k must be of the form 2^r5^q; then $c_k = 1/S(2^q5^r)$. The estimation for $k = 2^r5^q$ almost reproduces that of case (a):

$$S(N) = S(10^{r+q}N) \le S(2^q5^r)S(kN) = S(kN)/c_k.$$

Since, for $N = 2^q5^r$,

$$S(kN) = S(10^{r+q}) = 1 = c_kS(N),$$

the value $1/S(N)$ for c_k can't be increased.

It remains to show that for $k = 2^r \cdot 5^q \cdot Q$, where Q is coprime with 10 (and $Q > 1$), the ratio $S(kN)/S(N)$ can be made arbitrarily small. We can consider $k = Q$ because $S(kN) \le S(QN) \cdot S(2^r5^q)$.

First let's find a number m such that $10^m - 1$ is a multiple of Q. (Evidently, there exist two numbers in this form, $10^s - 1$ and $10^t - 1$, $s > t$, having equal remainders modulo Q; their difference $10^t(10^{s-t} - 1)$ is divisible by Q; we can take $m = s - t$.) Denote $(10^m - 1)/Q$ by R; then for any natural n

$$R_n = (10^{mn} - 1)/Q = R(10^{m(n-1)} + 10^{m(n-2)} + \ldots + 10^m + 1).$$

Now let $N_n = R_n + 1$. Then $S(N_n) \ge (n - 1)R$, since $R < 10^m - 1$, and $S(QN_n) = S(QR_n + Q) = S(10^{nm} + Q - 1) = 1 + S(Q - 1) = S(Q)$. Finally, $S(QN_n)/S(N_n) \le S(Q)/(n - 1)R \longrightarrow 0$ when $n \longrightarrow \infty$.

Physics

P16

Since the collision of the ring with the step is absolutely inelastic, the ring's momentum changes after the collision. The step acts on the ring along its radius R (fig. 5). During the collision the projection of the ring's momentum on the axis OX (going along R) drops to zero. The projection of the momentum on the axis OY

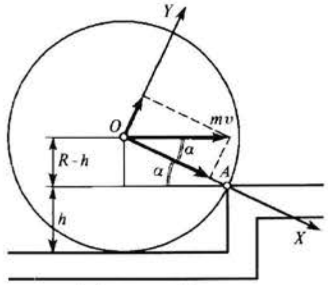


Figure 5

doesn't change. After the collision the ring's total momentum becomes equal in absolute value to $mv\sin\alpha$, and its velocity $v\sin\alpha = v[(R - h)/R]$ (see figure 5). Now let's make use of the energy conservation law. Immediately after the collision the kinetic energy of the ring is

$$2 \cdot \frac{mv^2}{2} \frac{(R-h)^2}{R}.$$

The factor 2 appears because of rotational movement.

After climbing the step, the ring acquires the potential energy mgh and kinetic energy $2\,(mv_h^2/2)$. So

$$\frac{mv^2(R-h)^2}{R} = mgh + mv_h^2.$$

The velocity of the ring after it "climbs" the step, therefore, is

$$v_h = \sqrt{v^2\frac{(R-H)^2}{R} - gh}.$$

The minimum velocity $v_{\min}$ at which the ring can still climb the step corresponds to $v_h = 0$; that is,

$$v_h = \sqrt{v_{\min}^2 \frac{(R-h)^2}{R} - gh} = 0,$$

from which we get

$$v_{\min} = \frac{R}{R-h}\sqrt{gh}.$$

P17

Denote the pressure of the liquid acting on the bigger piston (having the area S_1) by p and the atmospheric pressure by p_0. Then the total force acting on the piston upward is equal to $F_1 = S_1p$, while the force acting on the same piston downward is the sum of the string tension T and the force of the atmospheric pressure $F_1' = p_0S_1$. Since the piston is in equilibrium, we can write

$$S_1p = S_1p_0 + T.$$

A similar equation holds for the lower piston: the tension force T and the force of the atmospheric pressure $F_2' = p_0'S_2$ act upward, while the water pressure $F_1 = p'S_2$ acts downward. The water pressure p' on the lower piston is higher than that acting on the upper piston by a factor of ρgl. Since the lower piston is also in equilibrium, then

$$T + p_0S_2 = (p + \rho gl)S_2.$$

Solving both equations simultaneously we get

$$T = \rho gl\frac{S_1S_2}{S_1 - S_2}.$$

A good way to verify the answer is to substitute extreme values of the parameters. Let $S_2 \longrightarrow S_1$. In this case $T \longrightarrow \infty$. Indeed, the whole structure remains in equilibrium because of the pressure of the water on the ring with area $S_1 - S_2$ on the upper piston. When $S_1 \longrightarrow S_2$ the pressure on the ring tends to infinity so that the string tension also tends to infinity.

Thus, when $S_1 = S_2$ we get an infinite value for T. But such a limiting transition is impossible. We've assumed that the system remains in equilibrium. Actually, for $S_1 = S_2$ there is no equilibrium since the system of pistons falls with a constant acceleration g. The tension of the string is then equal to zero. This is a good example of how careful we should be with limiting transitions in physics. We should always make sure that such a transition doesn't alter the phenomenon.

P18

Warmed by the hand pressing it against the frosty window, the coin warms and melts the ice under it. Since the edge of the coin is slightly thicker than its body, the area of contact at first is primarily along its circumference. The rest of the coin is separated from the window by a thin layer of air. The thermal conductivity of the air is much less than that of the metal. So the ice along the circumference is the first to melt. After the ice under the edge melts enough, the rest of the coin comes into contact with the ice, which then starts to melt under the entire area of the coin.

P19

Consider the case of the connection in series first. The current in this circuit is

$$I = \frac{U}{R_1 + R_2},$$

where U is the voltage difference in the circuit, $R_1 = \rho(l/\pi r_1^2)$ is the resistance of the thin wire (of radius r_1), $R_2 = \rho(l/\pi r_2^2)$ is the resistance of the thick wire (of radius r_2).

The power released by the current on each of the resistances is equal to $N = I^2R$—that is,

$$N_1 = \frac{U^2}{(R_1+R_2)^2}R_1,$$

$$N_2 = \frac{U^2}{(R_1+R_2)^2}R_2.$$

In the stationary regime—that is, when the temperature of the wires no longer changes—each of the wires releases power equal to $N' = ks(T - T_0)$ into the surroundings, where k is the proportionality factor, $s = 2\pi rl$ is the surface area of the wire, T is the wire temperature, and T_0 is the temperature of the surroundings. In the stationary regime $N' = N$—that is,

$$\frac{U^2}{(R_1+R_2)^2}R_1 = k2\pi r_1 l(T_1 - T_0),$$

$$\frac{U^2}{(R_1+R_2)^2}R_2 = k2\pi r_2 l(T_2 - T_0),$$

where T_1 and T_2 are the temperatures of the thin and thick wires, respectively. Dividing the first equality by the second, we get

$$\frac{T_1 - T_0}{T_2 - T_0} = \frac{R_1}{R_2} \cdot \frac{r_2}{r_1}.$$

Since $r_2 > r_1$ and $R_1 > R_2$, we get $T_1 - T_0 > T_2 - T_0$, or $T_1 > T_2$.

Consider now the case of connection in parallel. The voltage drop on the resistance R_1 and R_2 is then the same, and the power released on each of them is

$$N_1 = \frac{U^2}{R_1},$$

$$N_2 = \frac{U^2}{R_2}.$$

Proceeding as we did in the first case, in the stationary regime we get

$$\frac{U^2}{R_1} = k2\pi r_1 l(T_1 - T_0),$$

$$\frac{U^2}{R_2} = k2\pi r_2 l(T_2 - T_0),$$

from which we obtain

$$\frac{T_1 - T_0}{T_2 - T_0} = \frac{R_2}{R_1} \cdot \frac{r_2}{r_1}.$$

Substituting $R_1 = \rho(l/\pi r_1^2)$ and $R_2 = \rho(l/\pi r_2^2)$, we arrive at the relation

$$\frac{T_1 - T_0}{T_2 - T_0} = \frac{r_1}{r_2} < 1,$$

or $T_1 < T_2$.

So in the first case the thin wire heats up; in the second case, the thick one.

P20

Plot the track of a light beam from an infinitely distant source through the eye. The beam is subjected to two refractions on the two surfaces of the eye's lens (fig. 6). According to the law of refraction,

$$\frac{\sin\alpha}{\sin\beta} = \frac{n_2}{n_1},$$

where n_1 is the absolute refractive index of the first medium (water or air), n_2 the absolute refractive index of the lens.

This formula suggests that if n_1 decreases (that is, if water is replaced with air), angle β decreases as well. This means that after refraction on the outer surface of the lens, the beam will go lower when the eye is in contact with air than when it's in contact with water. If under water the image of a distant object is projected on the retina, the image of the same object in the air will fall in front of the retina. So it turns out the person is nearsighted.

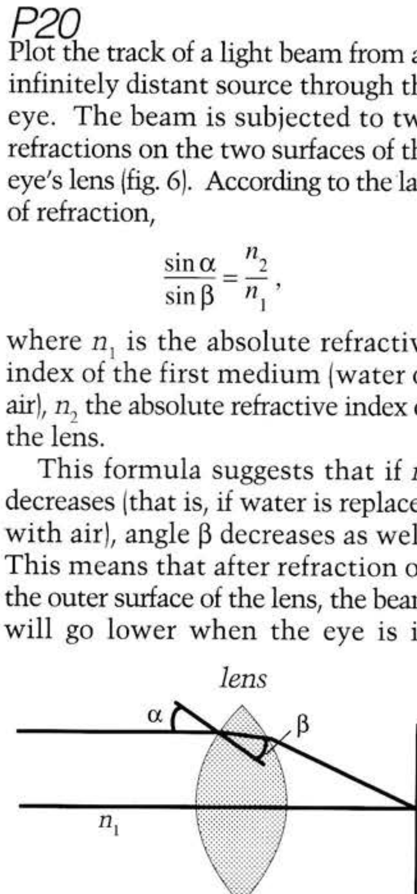


Figure 6

Brainteasers

B16

No, it isn't symmetrical. Notice that turning one of the stars upside down gives us a figure with central symmetry (fig. 7).


Figure 7

B17

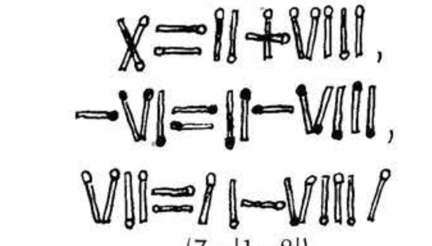
First, notice that each ball on the left pan is heavier than the ball of the same color on the right (otherwise there's a ball on the left that's lighter than the ball of the same color on the right, and we can then exchange them without tipping the balance). If there are no less than three balls on each side, we could exchange the pair of balls with minimum mass differences without affecting the balance. So there are at most two balls on each side. Obviously, there can be one ball in each pan. Two balls are also possible provided that the mass difference for each pair of balls of the same color is the same.

B18

See figure 8.

B19

Four vertices are at the corners of the square. Each of the vertices inside the square is a common vertex of exactly two rectangles. Let n be the number



(7 = |1 − 8|)

Figure 8

of rectangles into which the square is cut, m the number of points of the square that are common vertices of exactly two rectangles. The total number of vertices for all the rectangles can now be calculated in two ways. On one side it's equal to $4n$, on the other $4 + 2m$. So $m = 2(n - 1)$, which is an even number. Adding the four corners of the square we get $m + 4$, which is also even.

B20

The thermal expansion of mercury is greater than that of steel, so the force pushing the ball upward decreases and the ball sinks lower.

Going to extremes

1. The solution to this problem is similar to that for problem 4 in this article.

2. (a) If the radius of the circumcircle of triangle ABD isn't smaller than the radius of the circumcircle of triangle ABC, then point H doesn't belong to the "double sector" (fig. 9) delimited by the smaller arc AB of the circle ABC and the arc symmetrical to it with respect to AB. We can easily see that the circles symmetrical to circle ABC with respect to the three sides of triangle ABC have a common point, the orthocenter H of ABC. This implies that the three "double sectors" built on each of the sides of ABC have a unique common point. It lies inside the circumscribed circle only if triangle ABC is acute, and then it coincides with point H (fig. 9). If the triangle isn't acute the common point coincides with the vertex of its nonacute (that is, either right or acute) angle (fig. 10).

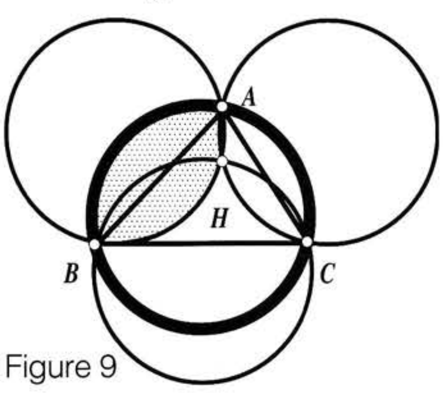


Figure 9

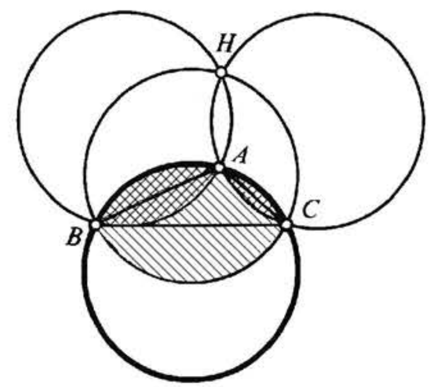


Figure 10

(b) If the smallest circle passing through points A, B, C contains some other given point D inside it, then the solution to problem (a) implies that D is the orthocenter of the acute triangle ABC. Therefore, any of the circles circumscribed around the (obtuse!) triangles ABD, BCD, or CAD provides a solution.

3. The solution to this problem is similar to that for problem 8 in the article.

4. Answer: $1 \cdot 2 \cdot 3 = 1 + 2 + 3$. If x is the largest of the numbers sought, then $xyz = x + y + z \le 3x$—that is, $yz \le 3$. All that's left is to work out all the possibilities: $yz = 1 \cdot 1, 1 \cdot 2, 1 \cdot 3$.

5. Take the longest edge.

6. Consider the position of the checkers one move before the first checker returns to its starting place.

7. Answer: $x_1 = x_2 = x_3 = x_4 = x_5 = 0$ or 2. First let's show that all x_i are equal. Under the assumption that this *isn't* the case, choose the largest x_i and, if it isn't unique, take the largest x_i for which $x_{i+1} < x_i$ (we assume that $x_6 = x_1$). Because of the symmetry of the system we can assume that this is the number x_1. Subtracting the fifth equation from the fourth, we arrive at a contradiction: $0 \ge x_4 - x_1 = x_1^2 - x_2^2 > 0$. So $x_1 = x_2 = x_3 = x_4 = x_5 = x$, where x satisfies the equation $2x = x^2$.

8. Let Q_1 be the smallest cube from the breakdown that touches the surface of the original cube, Q_2 the smallest cube adjacent to the face of Q_1 parallel to its outer face F, Q_3 the smallest cube adjacent to the face of the cube Q_2 parallel to F, and so on. We get a sequence of cubes that get smaller and smaller, the last of which is adjacent to the face of the original cube parallel to F. This contradicts the choice of the cube Q_1.

9. (a) Let's assume that airplanes flew from 6 airports to airport O. Take the two (A and B) for which the angle AOB is the smallest. Then angle $AOB \le 60°$, which implies that one of the distances AO or BO is greater than AB, which is impossible.

(b) Use the "ordering rule." Let's assume that at least one plane landed at each airport. Then for any airport A_1 there is a chain of airports $A_1, A_2, \ldots,$ where each A_{i+1} stands for one of the airports from which the plane flew to A_i. It's easy to see that the chain has to close up: a plane from A_1 will fly to a certain airport A_n—that is, $A_n = A_1$. If n is greater than 2, we get a contradiction: $A_1A_2 < A_2A_3 < \ldots < A_{n-1}A_n < A_nA_1 < A_1A_2$. So $n = 2$ and the set of airports breaks down into pairs, which is impossible with an odd number of airports.

Kaleidoscope

1. At midnight the velocity of the Earth's rotation is added to its orbital velocity, whereas at midday it's subtracted.

2. It moves faster in the winter in the Northern Hemisphere, since in this season the Earth passes its perihelion.

3. Saturn's ring isn't a solid body.

4. Acceleration caused by the Sun is approximately the same for both the Earth and the Moon. The two of them form a single system revolving around its center of mass, which, in turn, revolves around the Sun.

5. No, because (unlike the case of a circular orbit) the force of gravity al-

ternately performs positive and negative work, so the planet or satellite keeps speeding up and slowing down.

6. No satellite can hover over such a region because its orbital plane has to go through the center of the Earth.

7. No—the Earth and the Moon revolve around different centers of attraction.

8. No additional speed is needed since bodies in the equatorial zone are already in orbit.

9. The force is equal to that with which the table acts on the Earth—that is, with a force equivalent to the table's weight.

10. In spite of air resistance the velocity of the satellite increases. Although friction reduces the mechanical energy of the satellite, only some of its potential energy is transformed into heat; the rest is transformed into kinetic energy.

The mental microexperiment. Neither weight nor weightlessness has anything to do with the collision. The principal role is played by mass and velocity. So when you're working in outer space, be careful not to bump into your spacecraft.

Tournament of Towns

1. The identity can be proved either by directly opening the brackets (note that each fraction $1/k$, $k = 1, 2, ..., n$, appears in exactly k brackets on the left side), or by induction on n.

2. Let r_c and r_d be the radius of circles c and d and $CD = l$ (fig. 11).

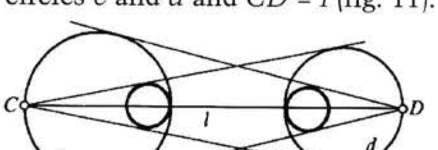


Figure 11

Notice that the dilation with center C and scale factor $2r_c/l$ takes the circle d into the given smaller circle inside c; and so the radius of the latter equals $2r_cr_d/l$.

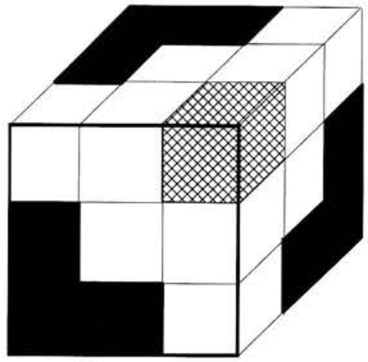


Figure 12

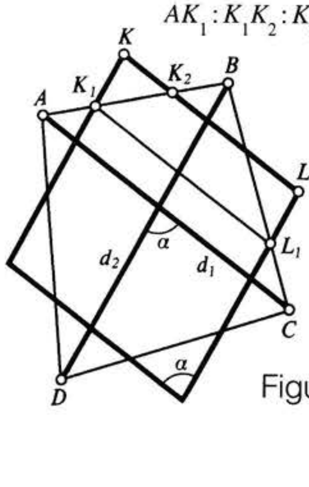
3. Yes, it's possible. An example is shown in figure 12 (the eight cubes that are out of view are red).

4. Put aside one coin and divide the rest into three equal piles of coins. Two of them, say A and B, will necessarily weigh the same, and two weighings are sufficient to identify them and determine whether the third pile C is lighter or heavier. Then pile A (or B) is divided in two and the halves are weighed against each other. If they balance, the coins in A are all genuine and pile C contains one or both counterfeit coins; if they don't, A and B each contain one counterfeit coin and C is entirely genuine.

5. The maximum number of parts is $10{,}001 = 100^2 + 1$. Consider the nth graph G_n added to $n-1$ graphs $G_1, ..., G_{n-1}$ already drawn. The number of parts of the plane that are split in two by G_n equals the number of arcs intercepted on G_n by $G_1, ..., G_{n-1}$ (including two infinite arcs). So it doesn't exceed $2n-1$ (there are at most $2(n-1)$ points where G_n intersects $G_1, ..., G_{n-1}$). Finally, recall that $1 + 3 + ... + (2n-1) = n^2$.

6. The second quadrilateral is a parallelogram whose sides are parallel to the diagonals d_1 and d_2 of the first one; the lengths of the sides are $(a + b)/(2a + b)$ times those of the corresponding diagonals (fig. 13). For the squares, $a/b = 1/2^{1/2} = (a + b)/(2a + b)$. So the areas of both quadrilaterals are equal to $(1/2)d_1d_2\sin\alpha$, α being the angle between d_1 and d_2.

$$AK_1 : K_1K_2 : K_2B = a : b : a$$



$$\frac{KL}{AC} = \frac{K_1L_1}{AC} = \frac{K_1B}{AB} = \frac{a+b}{2a+b}$$

Figure 13

7. The elephants weigh 5 metric tons each. Let W_k be the weight in kilograms of the kth elephant from the right and $d_k = W_k - 5{,}000 > -5{,}000$. Then $2d_k + d_{k+1} = 0$, which yields $d_k = (-2)^{k-1}d_1$. If $d_1 > 0$, then $d_1 \geq 1$ (d_1 is an integer) and $d_{14} = -2^{13}d_1 < -5{,}000$; if $d_1 < 0$, then similarly $d_{15} < 5{,}000$. It follows that $d_1 = 0$, $d_k = 0$, and $W_k = 5{,}000$ for all k.

8. Use the equality of angles inscribed in the same arc and symmetry of the rhombus with respect to its diagonal BD to prove successively that angles BAP, BQP, RCB, and RAB are equal. A nice point of the proof is to show that R lies between B and Q; this can be derived from the equalities $QP = QA = QC$, which means that Q is the point of circle CPQ most distant from PC.

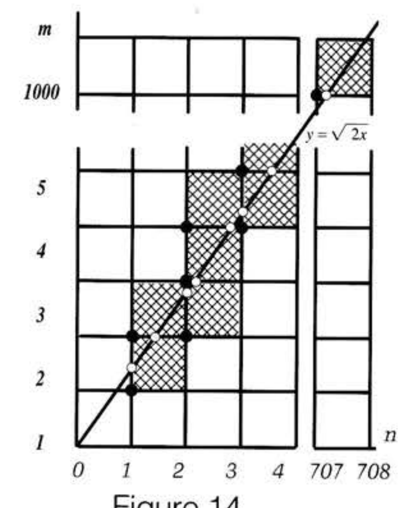


Figure 14

9. The number of pairs is 1,706. The idea is to interpret the given inequalities in terms of the coordinate plane. A pair (n, m) satisfies them if and only if the line $y = 2^{1/2}x$ lies between the lines $y = mx/(n + 1)$ and $y = (m + 1)x/n$—that is, intersects the square cell $\{(x, y): n \leq x \leq n + 1, m \leq y \leq m + 1\}$ (fig. 14). To count up the number of such squares for $0 < n \leq 1{,}000$, $0 < m \leq 1{,}000$, notice that it's equal to the number of intersections of the line $y = 2^{1/2}x$ with the lines $x = 1$,

2, ..., 1,000/$2^{1/2}$ (1,000/$2^{1/2}$ = 500 · 1.414 = 707) and y = 2, 3, ..., 1,000, which yields 707 + 999 = 1,706. (Compare this with the solution to problem 5.)

10. Any basic collection arranged in increasing order has the following form:

$$\{\underbrace{1, 1, ..., 1}_{p \text{ times}}, \underbrace{p+1, p+1, ..., p+1}_{q \text{ times}}, \underbrace{(p+1)(q+1), ..., (p+1)(q+1)}_{r \text{ times}}, ...\},$$

where p, q, r, ... are arbitrary natural numbers. The total of such a collection is $N = (p + 1)(q + 1)(r + 1) ... - 1$. For $N = 200$ the number $N + 1 = 201 = 3 \cdot 67$ has only two prime factors. That's why in this case there are only three basic collections: the trivial one (p = 200), the collection {1, 1, 3, 3, ..., 3} (p = 2, q = 66), and the collection {1, 1, ..., 1, 67, 67} (p = 66, q = 2).

11. The median of a triangle divides it into parts of equal area. Expressing the equality of areas in terms of the sides a and b, the median m between them, the angle γ between a and m, and the angle 2γ between m and b, we obtain the equation $am\sin\gamma = bm\sin 2\gamma$, which gives us $\cos\gamma = a/2b$. Given a and b we can thus construct γ.

12. (a) We may take, for example, the numbers $2k + 1$, 3, and all the rest equal to 1.

(b) The sum of n required odd integers must be equal to their product. So n is odd, according to the condition $n = 4k + 3$. Let m be the number of these integers having the remainder 3 modulo 4 (the rest have the reminder 1). Then the sum and the product of all n integers have the same remainders modulo 4 as $2m - 1$ and $(-1)^m$, respectively. But this is impossible: for an even m the first remainder is 3, the second is 1; for an odd m they are 1 and 3, respectively.

13. (a) The smallest number possible is 4.

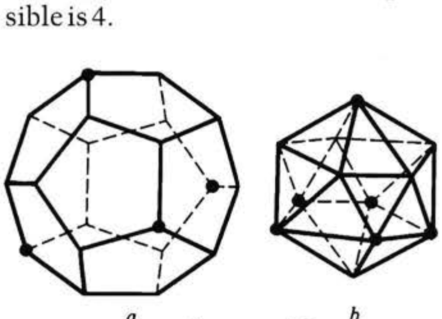


Figure 15

(b) The smallest number possible is 6.

For examples, see figure 15. Since any vertex of a dodecahedron belongs to 3 of its 12 faces, the number 4 = 12/3 in (a) is minimal. In case (b) the reasoning is somewhat more sophisticated (a simple estimate 20/5 = 4 won't suffice).

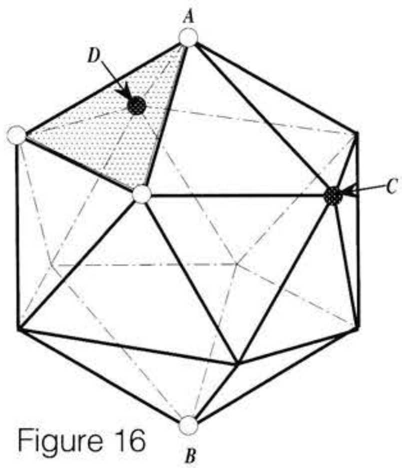


Figure 16

Consider 6 pairs of opposite vertices (like A and B in figure 16) of an icosahedron. If the number of marked vertices is less than 6, then one of the pairs (A, B) would be unmarked, and one of the vertices of the pair (say, A) would be joined by an edge to no more than two marked vertices, C and D. So among the 5 faces adjacent to A there could be at most 2 · 2 = 4 faces having a marked vertex, C or D. This leaves at least one face unmarked.

14. The solution to problem 4 works with $6k + 1$ coins for any natural k; 103 = 6 · 17 + 1.

15. This is the polynomial

$$(x-1)(x^2-1)\dots(x^{2^{n-1}}-1).$$

16. The answer to (b) is 3. Substitute 500 for 200 in the solution to problem 10 and notice that 501 also has only 2 prime factors: 501 = 3 · 167.

17. The smallest number of slices is $p + q - 1$. Dividing the cake into p equal parts by $p - 1$ parallel cuts and into $q - 1$ cuts parallel to the first ones, we get the required partition into $(p - 1) + (q - 1) + 1 = p + q - 1$ slices. To prove that this number can't be diminished, let's represent any partition in question by a graph.

Plot $p + q$ points on a plane, p of them for p equal parts of cake, q for q equal parts. Join the pairs of points corresponding to the parts that have a common slice of the given partition in them. The points and junctures are called the vertices and edges of the graph, respectively. Each edge is related to its own slice of cake, so the number of slices isn't less than the number of edges. On the other hand, every two vertices are connected by a chain of edges with each other. (If they aren't, there is a connected subset of the vertices, which aren't joined to any vertex outside the subset. This corresponds to a part of cake that can be distributed either among $k < p$ guests in the first case, or among $l < q$ guests in the second case. But $k/p \neq l/q$ for coprime p and q.) It remains to notice that a connected graph with $p + q$ vertices has at least $p + q - 1$ edges.

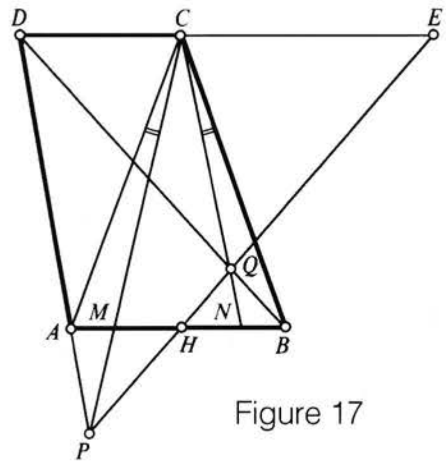


Figure 17

18. Let the line AB meet CP at M and CQ at N, and the line PQ meet CD at E (fig.17). It suffices to show that $MH = HN$. By the obvious similarities of triangles APH and DPE and triangles BQH and DQE, the following equalities hold:

$$MH/AH = CE/DE = NH/BH.$$

Since $AH = BH$, we have $MH = NH$.

To obtain a solution with no calculations at all, consider a central projection of the figure onto a plane p passing through AB from a center O such that the plane OCD is parallel to p.

19. Yes, in both cases—such polygons exist.

(a) A triangular prism with two

icosahedrons (fig. 15b) erected on its bases satisfies all the conditions except convexity. But the icosahedral bulbs can easily be "flattened" by stretching and contracting their edges.

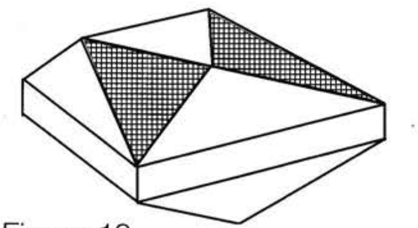


Figure 18

(b) Change the prism in the example above to the polyhedron shown in figure 18 and "flatten" the icosahedral bulbs so that the faces adjacent to the shaded triangles become the extensions of the latter to form 6 quadrangular faces of a new polyhedron. Its top view is shown in figure 19.

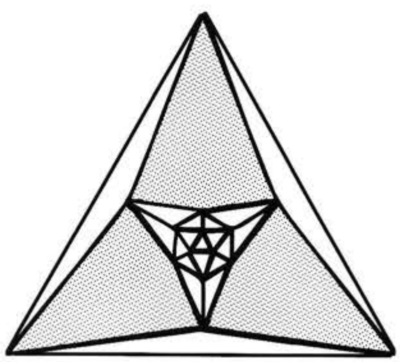


Figure 19

20. Let $S < 1$ be the area of one of the blots, x and y the lengths of its projections onto the perpendicular sides of the square; then $S < S^{1/2} \le xy \le (x + y)/2$. Adding up these inequalities for all the blots and taking into account the fact that projections of different blots are disjoint, we determine that the total area of the blot is less than $(a + a)/2 = a$.

Natural logarithm

1. We have $t = (1/k)\ln(N/N_0)$. The generation time, 20 minutes, is how long it takes to double the population, or $N/N_0 = 2$. So $t_g = (1/k)\ln 2$. Since t_g = 20 min, 20 min = $(1/k)\ln 2$, k = ln 2/20 min. Using a calculator, we get ln 2 = 0.6931, so that k = 0.6931/20 min = 0.0347 (to three significant figures).

2. Five hours and thirty-two minutes.

3. The rational number 271,801/99,990 has the value 2.7182818281828..., where 1828 repeats infinitely many times. We need only show that the digit in the tenth decimal place isn't 1. This requires use of the binomial expansion for $kt = 1$ and a calculator or computer that can carry 11 or 12 significant digits. But you'll need to be very clever to evaluate the $m = 12$ case. An ordinary calculator lacks the necessary precision.

Math surprises

1089. If a, b, c are the decimal digits of the larger of the two numbers that you take the difference of, then the digits of the difference are $a - c - 1$, 9, $10 + c - a$. But the sum of $100(a - c - 1) + 90 + (10 + c - a)$ and $100(10 + c - a) + 90 + (a - c - 1)$ is 1089.

6174. If *abcd* is the largest number you can form with the four digits a, b, c, d, then *dcba* is the smallest, and the difference between these two numbers is $999(a - d) + 90(b - c)$. Since both $a - d$ and $b - c$ are single-digit numbers, this leaves at most 100 cases to check, and we'll leave the rest of the problem to you. (In fact, there are other ways to reduce the work still further.)

153. Any five-digit number is at least 10000, but the sum of the cubes of its digits is less than five times 1000 = 5000, so that our operation decreases it. In the same way, you can see that the operation decreases longer numbers even more. So you need only check numbers with at most four digits. Once again, there are ways to reduce the work a bit further, but this time there almost inevitably remains quite a lot of sheer checking to be done.

0, 0, 0, 0. Here there's a very nice argument. Just look at whether the numbers you started with are even (E) or odd (O). For example,

either	O, E, E, E or E, O, O, O
yields	O, E, E, O;
then	O, E, O, E,
and then	O, O, O, O,
and finally	E, E, E, E.

Every pattern of odds and evens appears here (up to cyclic rearrangement), and so we see that after at most four turns all the numbers will be even. Then after four more turns they'll be multiples of 4, and four turns later they'll be multiples of 8, and so on. But since the numbers aren't getting any bigger, the only way they can end up being divisible by a very large power of 2 is by being identically zero.

It turns out that the same thing happens whenever the number of starting numbers is any power of 2, but in all other cases there are starting patterns that don't ultimately end in zeros.

1, 11, 21, 1211, etc. What's the rule here? You just read each sequence aloud in a suitable way, and you'll get the next one. For example, the first sequence consists of one "one," so the second sequence *is* "one one." This consists of two "one," so the next sequence is "two one," which in turn may be described as one "two," one "one," and so leads to "one two one one," and so on.

The problem about the rate of growth is much harder. I proved some time ago that each of the later sequences is about 1.303577269034296391257 099112152551890730702504659 40 ... times as long as the one before it, where the approximation gets better and better for later and later sequences. This mysterious number is the largest solution of the equation

$$f = x^{71} - x^{69} - 2x^{68} - x^{67} + 2x^{66} + 2x^{65} + x^{64} - x^{63} - x^{62} - x^{61} - x^{60} - x^{59} + 2x^{58} + 5x^{57} + 3x^{56} - 2x^{55} - 10x^{54} - 3x^{53} - 2x^{52} + 6x^{51} + 6x^{50} + x^{49} + 9x^{48} - 3x^{47} - 7x^{46} - 8x^{45} - 8x^{44} + 10x^{43} + 6x^{42} + 8x^{41} - 5x^{40} - 12x^{39} + 7x^{38} - 7x^{37} + 7x^{36} + x^{35} - 3x^{34} + 10x^{33} + x^{32} - 6x^{31} - 2x^{30} - 10x^{29} - 3x^{28} + 2x^{27} + 9x^{26} - 3x^{25} + 14x^{24} - 8x^{23} - 7x^{21} + 9x^{20} + 3x^{19} - 4x^{18} - 10x^{17} - 7x^{16} + 12x^{15} + 7x^{14} + 2x^{13} - 12x^{12} - 4x^{11} - 2x^{10} + 5x^{9} + x^{7} - 7x^{6} + 7x^{5} - 4x^{4} + 12x^{3} - 6x^{2} + 3x - 6.$$

(Thanks to Ilan Vardi for his accurate recomputation of this number and its defining equation.)

CHECKMATE!

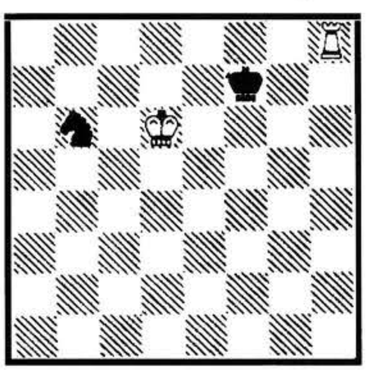
Rook versus knight

Kingdoms lost because of a horse

by Yevgeny Gik

THIS CORRELATION OF FORCES, rook *vs.* knight, is theoretically a wash. But if the horse strays a bit too far from its king, its fate hangs by a thread. It's interesting that computers have had a great deal of success in this sort of endgame. In particular, a machine has found a record-setting position in which the rook, given the best possible play by both sides, takes the knight in the twenty-seventh move—white: Kc1, Rf8; black: Ka3, Ne2.[1]

Let's look at an interesting étude.



A. Kopnin, 1987
To win.

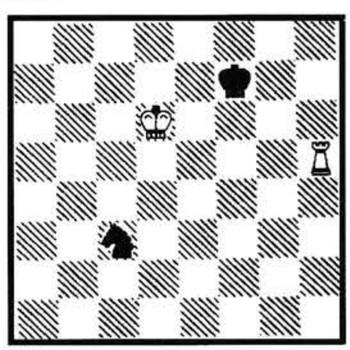
The main variation of the solution is this:

1.	Rh4	Nc8+
2.	Kd7	Nb6+
3.	Kc6	Nc8
4.	Rh7+	Kf6!
5.	Rh6+	Kg7
6.	Re6	Kf8!
7.	Kd7	Kf7
8.	Rh6!	

The author of the study supplements it with a number of additional variations. Here's the longest:

6.	...	Na7+
7.	Kd6!	Kf8
8.	Kd7	Nb5
9.	Re3	Nd4
10.	Rd3	Nc2
11.	Kd6	Kf7
12.	Rf3+	Kg6
13.	Kc5	Ne1
14.	Re3	Nc2
15.	Re2	

Now let's look at the following position.



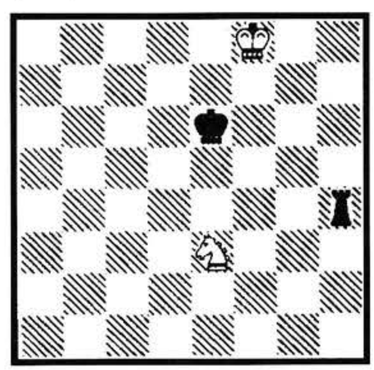
To win.

The computer that studied it indicated these initial moves:

1.	Ke5	Na4
2.	Rh7+	Ke8
3.	Kd6	Nb6
4.	Rh8+	Kf7

There's no need to go any further—we now have before us the étude of Kopnin we just looked at. So we can say that the machine constructed a more complicated study than the famous composer of chess problems!

If a win is possible in the battle between rook and knight, the computer will always find it. But a human being is capable of working through such endgames. Let's take a look at three interesting examples from the play of grandmasters.



Neiman—Steinitz
(Baden-Baden, 1870)

The first Chess King efficiently finishes the game:

1.	...	Re4!
2.	Nd1	

Or 2. Ng2 Kf5 3. Kf7 Kg4 4. Kf6 Re2; 2. Nc2 Kd5 3. Na3 Kc5 4. Kf7 Kb4 5. Nb1 Re2, and it's all over.

2.	...	Rf4+

---

[1] In this installment of Checkmate! the algebraic method of notation is used, in which only the piece and its destination are given.—*Ed.*

3.	Kg7	

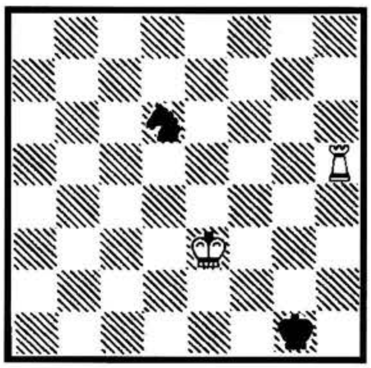
After 3. Ke8 Rf3!, white would be in a vise: 4. Nb2 Rb3.

3.	...	Rf3
4.	Kg6	

There's no saving the horse even with 4. Nb2 Kd5 5. Kg6 (5. Na4 Rb3) 5. ... Kd4 6. Kg5 Rf1 7. Kg4 Rb1 8. Na4 Rb4.

4.	...	Ke5
5.	Kg5	Kd4
6.	Kg4	Rf1
7.	Nb2	Rb1
8.	Na4	Rb4

The horse is a goner.



Bogolyubov—Rubenstein (San Remo, 1930)

1.	...	Nc4+

It's not hard to convince ourselves that 1. ... Kg2 won't save black.

2.	Kd3	Nb2+

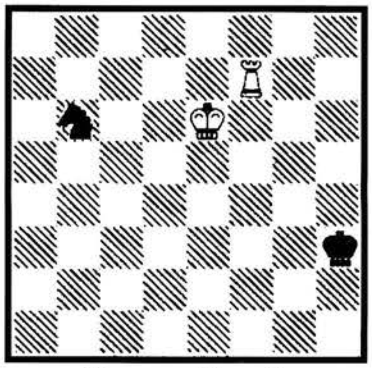
The knight isn't able to join up with his king, and 2. ... Nd6 won't help either: 3. Rd5 Ne8 4. Kd4 Nf6 5. Rf5 Ne8 6. Kc5.

3.	Ke2	Nc4

For 3. ... Kg2, the response 4. Rh4! is decisive; here's another variation: 3. ... Na4 4. Rg5+ Kh2 5. Kf2 Kh3 6. Kf3 Kh2 7. Rg2+ Kh3 8. Rc2 Nb6 9. Rc6.

4.	Rc5	Nd6
5.	Kf3	Kh2
6.	Rd5	Nc4
7.	Kf2	Kh3
8.	Rd3+	Kh2
9.	Rd4	

Black resigns.



Karpov—Ftacnik (Salonika, 1988)

This was the last game of the Chess Olympiad. Although the USSR team was assured of victory, it was important for the former world champion to win in order to take first place on his board. Almost up to the very end the Czech grandmaster defended himself precisely, but as soon as the rook-*vs.*-knight endgame appeared on the board, he made a fatal mistake.

83.	...	Nc4?

The correct move would have been 83. ... Na4!, and in a roundabout way—b2-d1-f2 (e3) or c5-d3—the knight would have been reunited with his king.

84.	Rf3+!	Kg4

Now, as in a real étude, two echo-variations arise. One was actually played out, and in the other the basic idea is realized in the form of a linkage: 84 ... Kg2 85. Rc3! Nd2 (85. ... Nb6 86. Rb3 Na4 87. Kd5) 86. Rc2.

85.	Rd3!	Kg5

These replies are no better: 85. ... Nb2 86. Rd2! Nc4 87. Rd4; 85. ... Nb6 86. Rb3.

86.	Kd5	Nb6+

Another trap for the knight: 86. ... Nb2 87. Rd2 Na4 88. Rc2 Nb6 89. Kd6 Kf5 90. Rc5+ Kf6 91. Rc6 Na4 92. Kd5 Kf5 93. Kd4 Nb2 94. Rc1 Ke6 95. Rb1 Na4 96. Rb4.

87.	Ke5	Nc4+
88.	Ke4!	Nb6
89.	Rd8!	Nc4
90.	Rd4	Kb6
91.	Ke5	Nc8
92.	Ke6	Na7
93.	Kd7	

Black resigns. ◘

Index of Advertisers