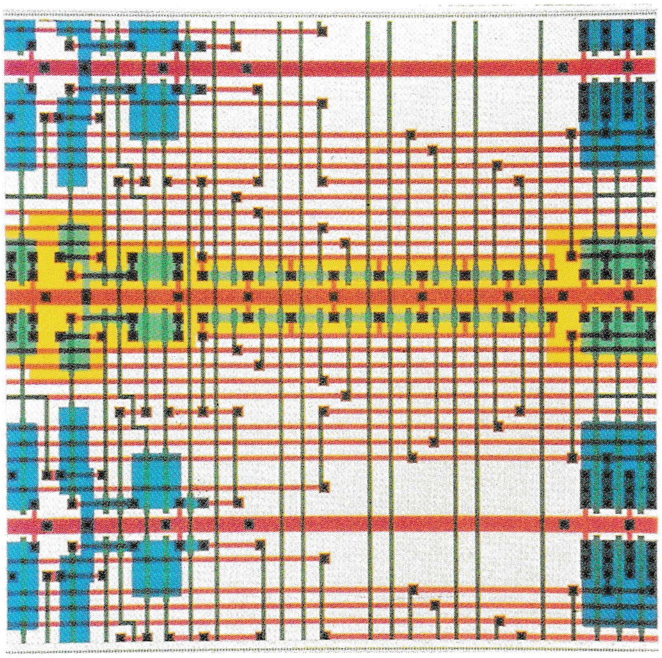
MARCH/APRIL 1993

$5.00

QUANTUM

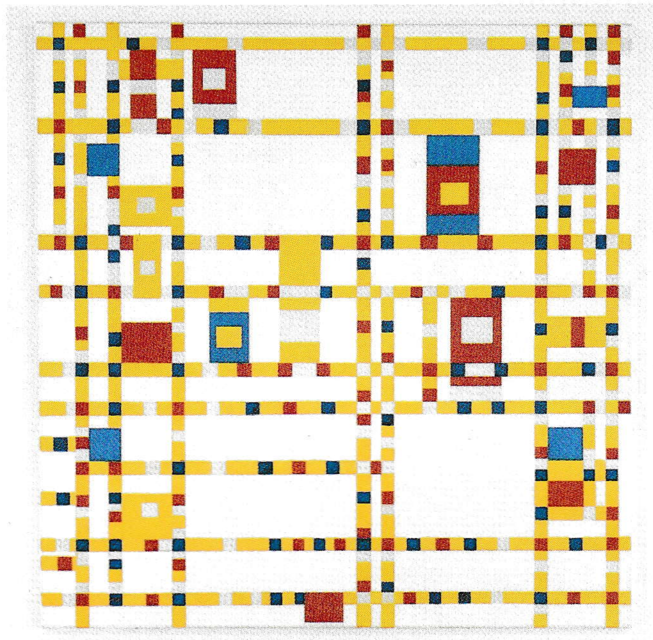
THE STUDENT MAGAZINE OF MATH AND SCIENCE

SPRINGER INTERNATIONAL



Courtesy of MoMA and Intel Corporation

Diagram of "CRISP" microprocessor by AT&T Bell Laboratories



Collection, The Museum of Modern Art, New York © MoMA

"Broadway Boogie Woogie" (1942–43) by Piet Mondrian

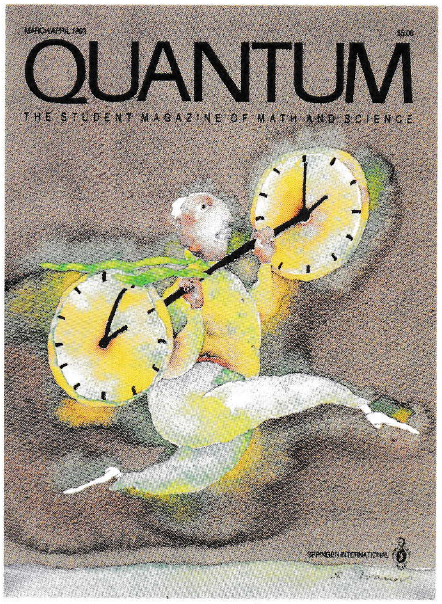
THE DESIGNERS OF AT&T'S "CRISP" MICROCHIP would no doubt be amused to see their "artwork" displayed next to a masterpiece by the modern Dutch master Piet Mondrian (1872–1944). But the design has made the rounds of museums throughout the US, along with twenty-nine other microprocessor diagrams of extraordinary intricacy and vibrant color, as part of the traveling exhibition "Information Art: Diagramming Microchips." The drawings have delighted and confounded many a viewer. Undisputably the work of human hands, harmonious and pleasing to the eye, they are but the first step in producing truly "fine" art: silicon chips no larger than a thumbnail. Packed within that space are as many as several million electronic components capable of performing millions of calculations per second.

So the diagrams are obviously not "art for art's sake." Color-coded by layer like multilevel roadmaps, they are used by engineers for guidance in correcting or verifying a circuit design. But then, many artists reject that slogan as well. According to the art historian H. L. C. Jaffe, Mondrian believed that "art can be a guide to humanity, that it can work toward eliminating casual facts of appearance and the arbitrary outlook of the individual, and thereby substitute a new harmonious view of life for the conventional tragic conception of existence. To this purpose Mondrian chose the strict and rigid language of geometry to produce first of all an extreme purity, and on another level, a Utopia of superb clarity and force." Mondrian himself wrote: "One serves mankind by enlightening it." The architects of the microcomputer revolution would surely agree.

"Information Art: Diagramming Microchips" has entered the second half of its national tour. Sponsored by Intel Corporation Foundation and organized by New York's Museum of Modern Art (MoMA), it is scheduled to appear at the Chicago Atheneum, April 1 to May 21; the Laguna Gloria Art Museum in Austin, Texas, June 12 to July 25; the Elvehjem Museum of Art in Madison, Wisconsin, September 1 to December 4; and the Georgia State University Art Gallery in Atlanta, January 4–31, 1994. The exhibition includes designs by thirteen manufacturers and universities. For those unable to attend, a 48-page booklet with 32 color and 10 black-and-white illustrations is available from MoMA.

QUANTUM

MARCH/APRIL 1993 VOLUME 3, NUMBER 4



Cover art by Sergey Ivanov

Funny things start happening when you approach the speed of light. The person sprinting across our cover is doing just that. (We suspect he's a mathematician who, right in the middle of his workout, heard about the article "In the Curved Space of Relativistic Velocities" in the latest issue of *Quantum*.)

He probably doesn't feel his *relativistic mass* increasing, in accordance with the formula

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}},$$

where m_0 is his *rest mass*, v is his speed, and c is the speed of light. But we're certain he feels the added mass of the barbells he's carrying. Why the weights are in the form of clocks is anybody's guess—until you read the article that begins on page 4. Afterwards, you may notice some artistic license in the rendering, but it's not the discrepancy in the times shown . . .

FEATURES

4 High-Speed Photography
What's that you see?
by B. M. Bolotovsky

10 Linguistic Math
Errorproof coding
by Alexey Tolpygo

26 Far-Flung Phenomena
The inevitability of black holes
by William A. Hiscock

34 Common Ground
In the curved space of relativistic velocities
by Vladimir Dubrovsky

DEPARTMENTS

2 Publisher's Page

16 Math Investigations
The Worm Problem of Leo Moser (part II)

17 Brainteasers

18 At the Blackboard I
Combinatorics–polynomials–probability

24 How Do You Figure?

30 Physics Contest
How about a date?

32 Kaleidoscope
Miss or hit

42 Looking Back
The problem book of Anania of Shirak

47 In Your Head
Landau's license plate game

49 Crisscross Science

50 At the Blackboard II
The satellite paradox

53 Happenings
Behind the scenes at the IMO ... Bulletin Board

56 Answers, Hints & Solutions

64 Toy Store
Wacky pyramids

PUBLISHER'S PAGE

Do as we say . . .

. . . not as we do (yet): On diversity in Quantum

THIS IS A STORY OF GOOD intentions, and how the road to—no, not hell, *heaven*—is paved with them.

It is the official position of the National Science Teachers Association that "science teachers must consciously strive to overcome the barriers created by society which discourage women from pursuing science for its career opportunities and for the enjoyment it brings to involved students." One of the ways to "eliminate sex role stereotyping" is to include "appropriate role models" in textbooks and other student reading material.

NSTA is also working to bring students of diverse cultural backgrounds into the scientific fold. As stated in another NSTA position statement, "we appreciate the strength and beauty of cultural pluralism." We believe (as I've noted in this space) that "all children can learn and be successful in science." To this end, science teachers have the responsibility to "expose culturally diverse children to career opportunities in science, technology, and engineering."

Those are our intentions. How well do we—in particular, *Quantum*—live up to them?

Well, leafing through the previous fifteen issues of *Quantum*, I'm hard-pressed to find evidence of women and persons of varied ethnic backgrounds. Perhaps, as a white male, I tend not to notice the overwhelming homogeneity of images and authors in our magazine. But I can imagine how a young woman, say, or African-American, would feel left out. I have to admit that *Quantum* to date may not look particularly inviting to nonwhites and nonmales. Readers initially attracted by the content may end up being repulsed by an unintended, unspoken subtext: "You don't belong here."

Quantum is *not* a whites-only boys club. Yet that is the perception among some of our readers. So—what are we going to *do* about it?

We are redoubling our efforts to find authors who reflect the diversity of the American people and the world at large. And if such variety is lacking in some branches of math and science, we'll try to project an image of

WOMEN SCIENCE MAJORS ATTENDING A CONFERENCE on women in science, math, and engineering in Washington, D.C., in November 1992 were surveyed to find out what they thought the country should do to improve math and science education. The predominant responses:

K–8:

- Make math and science more fun, exciting, and interactive;
- Let girls know that it's okay to like science and math.

High school:

- Get more role models and mentors who can counsel students;
- Have girls take more math/science courses, spend more time in math/science.

College:

- Keep up the encouragement for individual students;
- Provide more research, internship, and lab opportunities.

The students were relatively optimistic about careers in science and math, saying they saw no barriers other than a lack of money for their education. College faculty who attended, however, when asked to describe the continuing challenges to the full participation of women in science, cited penetrating the inner circle and being taken seriously by colleagues and superiors; dealing with child care and other family responsibilities; and overcoming the negative perceptions of women in math and science that girls develop at an early age.

When asked to name their heroes/heroines in math and science, more than half of the young women were unable to come up with a single name. Albert Einstein placed a distant second to "Nobody."

"What Works: Women in Science, Math, and Engineering" was sponsored by the Women's College Coalition, 1090 Vermont Ave. NW, Third Floor, Washington, DC 20005. This national conference brought together—for the first time—women science majors, their faculty, and many of the country's leading women scientists.

the future. By faithfully presenting the current state of affairs, we simply reinforce it. It's a vicious circle that must be broken: to create a better future, we must present that future as if it already exists!

This won't be as easy as it might seem. Currently more than half of our material comes from Russia, which has not yet entered the crucible of "consciousness raising" in the area of gender. The articles we receive—which I hope you'll agree are instructive and charming—are almost invariably by men. The Russian artwork, beautiful and unique to *Kvant* and *Quantum*, is generally by males, of males—and for males? I think the artists would say no, but I can understand how some readers might feel excluded and lose heart.

Our efforts must therefore be concentrated on the American side. As more teachers and professors become familiar with *Quantum*, I hope we can persuade more of them to write for us—to jot down that curious stray thought or nagging question, work it through in a fresh and engaging way, and share it with our young readers. In my view, the best articles are those having to do with a person's own research, that take some interesting aspect and develop it in a way that will appeal to a young person who aspires to do similar work. Students need to see the excitement and interest of those who make science and math their life's work.

Quantum's American field editors will continue to search for authors with the special talents needed to write for us. But if you know of any potential authors, especially those who could serve as role models for young women and ethnically diverse students, please let us know about them. We'll get in touch with them in a flash. With your help, and persistence on our part, I believe *Quantum* can come to reflect the extraordinary talent found among *all* of the best scientists and mathematicians in this heterogeneous world. The road to a more open and diverse scientific community of the future is paved with this intention.

—Bill G. Aldridge

QUANTUM

THE STUDENT MAGAZINE OF MATH AND SCIENCE

A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the Russian 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) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York Inc. Volume 3 (6 issues) will be published in 1992–1993. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by the Russian Academy of Sciences and the Russian Academy of Pedagogical Sciences, as well as original material in English. **Editorial offices:** NSTA, 3140 North Washington Boulevard, 2nd Floor, Arlington, VA 22201. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. Production Editor: Madeline R. Kraner.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (202) 328-5800; (New York) Thomas Heitzman (212) 460-1675; and G. Probst, Springer-Verlag, Heidelberger Platz 3, 1000 Berlin 33, FRG, telephone (0) 30-82 07-1, fax 30-82 07-1.

Second class postage paid at New York, NY, and additional mailing ofices. **Postmaster:** send address changes to: *Quantum* Springer-Verlag New York, Inc., Journal Fulfillment Services Department, 44 Hartz Way, Secaucus, NJ 07096-2491.

Subscription Information: North America

Student rate: $14; Personal rate (nonstudent): $18; Institutional rate: $30; Single Issue Price: $5. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus, NJ 07096; or call 1-800-SPRINGER (in New York, call [201] 348-4033).

All Countries Outside North America

Subscription rates in U.S. currency as above (all rates calculated in DM at the exchange rate current at the time of purchase) plus postage and handling. SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Air mail delivery to all other countries is available upon request. Orders may be placed through your bookseller or directly through Springer-Verlag, Heidelberger Platz 3, 1000 Berlin 33, FRG.


S. Ivanov

What's that you see?

On the perceived shape of rapidly moving objects

by B. M. Bolotovsky

MAYBE YOU'VE HEARD THE saying: "Better to see once than hear twice." The meaning is quite clear: someone can talk to you till they're blue in the face, but if you want to convince yourself of something, you want to see it with your own eyes. Seeing is believing.

But can we always believe what our eyes are telling us? No. Sometimes we see things that just aren't there. I'm not talking about hallucinations or bad vision here. An observer, of sound mind and possessed of excellent eyesight, can under certain conditions see a very distorted picture of a phenomenon, something not at all like what is happening in actual fact. Why? Because the senses can deceive you. Well, let's put a movie camera in place of the human observer. You might think that its "readings" are objective. And yet when we look at the resulting film, we see the same thing as the observer whom we doubted and replaced with the movie camera.

And what we see on the screen, and what the observer saw, are not at all what is actually happening. How is that possible?

I'll give you some examples that will show how it's possible. You won't find anything mysterious in the examples. They simply show that we must really think about every physical observation. We have to try and understand what instrument readings mean and use them to recreate the true picture of a phenomenon.

Photographing light on the wing

About 20 years ago the American physicist Michel Duge designed a photographic shutter that operated at extraordinarily fast speeds. Why are such shutters needed?

If we photograph a stationary object or scene, the sharpness of the picture isn't affected by the length of time the shutter is open—the "exposure time." The object is stationary, and the image of the object on the film is also stationary. So increasing the exposure time doesn't decrease the sharpness of the image. But if the object moves, that's another story. Then the image of the object on the film moves as well. In this case, the shorter the exposure time, the sharper the photograph will be. Conversely, the longer the exposure time, the more the image will blur.

On some photographs (usually photos of running people or speeding cars), the blurring is appropriate. It emphasizes the speed of the object. But as a rule, the faster the object, the shorter the exposure time—that is, the time the shutter is open, allowing light to enter the camera. (In this article we won't get sidetracked by the fact that photographic film has a determined sensitivity—the "film speed"—and that you need "faster" film for shorter exposures. Let's just assume that the film speed can be chosen to correspond to the necessary exposure times.)

Let's get back to the fast shutter I mentioned. Its exposure time was fantastically short—around 10^{-11} s. It was "fantastic," of course, only until the shutter was invented. After that it was no longer fantasy but an actual, honest-to-goodness achievement. But what on earth was it good for?

This kind of lightning-quick shutter could be used to photograph objects that move at velocities approaching the speed of light. The image on the film would certainly be sharp enough. But where can we find such objects? They have to be big enough to show up on a photograph—say, on the order of 1 cm. But we can't accelerate bodies that big to velocities near the speed of light, at least not at present. That requires an enormous expenditure of energy. (To accelerate a body with a mass of 1 g to half the speed of light, an energy expenditure of the order of 10^{11} J is needed. That's the energy generated every second by ten nuclear power plants.) The maximum velocities of macroscopic bodies under earthly conditions are of the order of tens of kilometers per second. We're able to accelerate electrons to velocities near the speed of light. But you can't directly photograph an electron—it's too small. We need an ob-

Art by Sergey Ivanov

ject big enough to see that moves at almost the speed of light. And Duge, the inventor of the high-speed shutter, found such an object: light itself!

A laser was used as the light source. It produced radiation in the form of extremely brief pulses lasting approximately 10^{-11} s. The light pulse emitted by the laser can be visualized as some volume filled with light waves and moving through space at the speed of light. This volume is called a wave (or light) packet. Obviously the packet's length is equal to $l = c\tau$, where c is the speed of light and τ is the time interval during which the pulse was emitted—that is, the pulse width. Substituting $c = 3 \cdot 10^8$ m/s and $\tau \cong 10^{-11}$ s, we get $l \cong 3$ mm. It's possible to photograph a "body" of that size.

The procedure is seemingly quite simple: we place a camera off to the side of the path of the "body" of light, aim at some point along the path, and shoot at the right moment. This setup is shown in figure 1. If we use such a setup, however, we'll see nothing on the film. Why? Because the image appears on the film only if light from the moving body lands on it. This can be "its own" light, if the body radiates (for instance, when you take a picture of a light bulb). If the body doesn't glow, it must be illuminated, and then light from the external source is reflected or scattered from the body and strikes the film. From figure 1 you can see that the light waves in the packet propagate along a path that doesn't allow them to enter the camera's lens. It's also pointless to shine light from an external source on such a packet—it neither reflects nor scatters light. The light waves from an external source pass through the wave packet as if it were an incorporeal specter, changing neither itself nor the packet.

There was a way around this difficulty. A glass vessel filled with water is placed along the path of the clump of light. A few drops of milk are added to the water, which makes it slightly cloudy, or turbid. This turbid medium scatters light passing through it. When the clump of light enters such a vessel with turbid water, the light waves in the clump begin to scatter and the clump becomes visible. Now we can photograph it. Of course, the entire setup must be placed in a completely dark room so that extraneous light sources don't interfere with the procedure.

This kind of setup has produced some beautiful photographs of light "on the wing," as it were. One of them is reproduced in figure 2. This is the first photograph in the history of science that was obtained under household conditions and represented an object with a velocity close to the speed of light (in water the speed of light is slower than that in a vacuum by a factor of $1/n = 3/4$, where n is the refractive index of water). During the time of exposure the light clump in the vessel moved only 2.2 mm. But the image on the photograph is "stretched" to more than 2.2 mm. The reason for this will become clear as you read on.

After this, Duge performed another exquisite experiment. He created a "dumbbell" made of light and photographed it.

The light dumbbell

Before I describe the results of this experiment with the dumbbell of light, let me say a few words about why such an experiment was needed.

According to the theory of relativity a body's length, measured in some frame of reference, depends on the body's velocity relative to this frame of reference. In particular, if the body is at rest and its length is equal to l_0, then when the body moves with a velocity v its length will be $l = l_0\sqrt{1 - v^2/c^2}$, where c is the speed of light. The factor $\sqrt{1 - v^2/c^2}$ is always less than unity, so when the body moves, its size in the direction of movement (this is what I meant when I said "length") is reduced by the factor $1/\sqrt{1 - v^2/c^2}$.[1] For velocities reached at present under earthly conditions, the factor $\sqrt{1 - v^2/c^2}$ differs insignificantly from unity. If a body has a velocity of 10 km/s (the order of magnitude of the velocity of space ve-

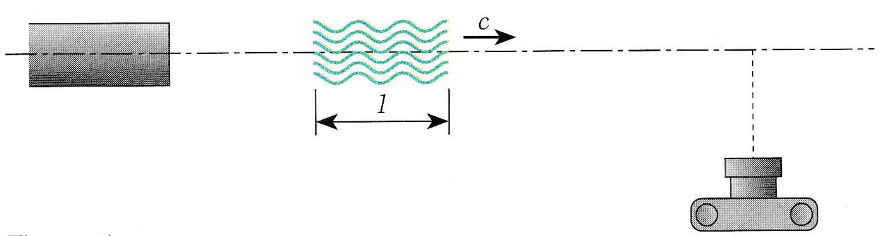


Figure 1
With this setup, it is impossible to photograph the light pulse—the light waves in the packet do not land on the lens.

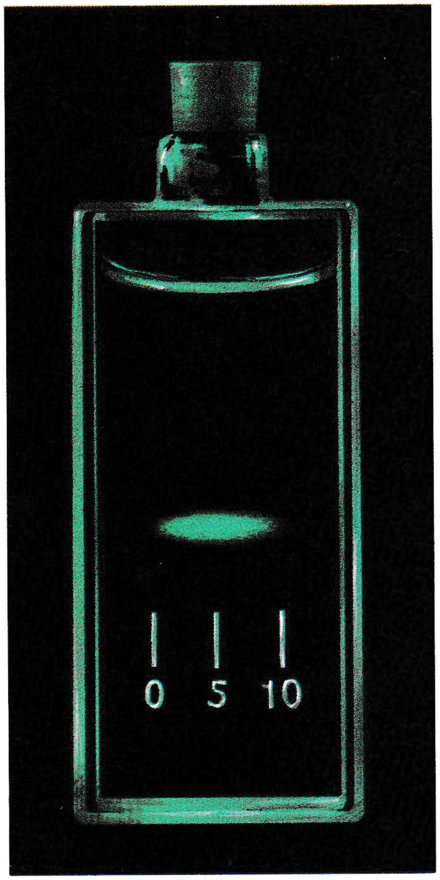


Figure 2
Photograph of a light pulse moving in a turbid (scattering) medium. The exposure time is 10^{-11} s.

[1]This shortening is often called the Lorentz transformation after the Dutch physicist and theoretician H. A. Lorentz, who was one of the first to predict this effect (in 1892).

hicles), then this factor is equal to 0.99999999944—that is, it differs from unity in the tenth place after the decimal point. So it's difficult to find the shortening during movement on the basis of velocities that are currently achievable. But this assertion of the theory of relativity is confirmed by a great number of consequences that arise out of it.

The change in a body's size when it is in motion leads to a change in its shape. A moving ball, if its velocity is close to the speed of light, must flatten out and turn into a pancake whose plane is perpendicular to the direction of travel; a cube must change into a parallelepiped; and so on.

Since we don't have bodies at our disposal that move at sufficiently high speeds, it wasn't possible to directly verify the shape change. But this phenomenon was subjected to a rather detailed theoretical analysis. This analysis showed that the question isn't as simple as it seems. The size of the ball really does decrease in the direction of movement. But looking at a moving ball, an observer won't notice this shortening. What gets in the way is the method of observation itself.

To understand why this happens, let's look at some examples that have a bearing on the question. Imagine a fine summer day, and we're lounging, admiring the landscape. Sun shining, trees rustling, birds singing . . . we perceive everything simultaneously.

But at the moment we hear a bird chirp, the bird may actually be silent. The sound propagates in the air with a velocity of approximately 340 m/s. So there is a measurable delay between the moment the bird produces its warble and the time we perceive it. We see the Sun at some point in the sky. But the light from the Sun actually takes about 8 minutes to reach the Earth. So we can't assert that now, at this very moment, the Sun is right where we see it.

Here's another example. We're photographing the starry sky. What does the photo we make tell us? Do we think that it reflects the position of the stars at the very moment we took the picture? Of course not. The light from the distant stars takes tens, hundreds, thousands of years to reach the Earth. One star may have gone out long ago, but we won't find out about it for quite a while. Its light continues to arrive, and we see this star on the photograph. Another star was born, but its light hasn't reached us yet, and so it's absent from the photograph.

The same phenomenon is responsible for all these discrepancies: the propagation velocity of the signal that brings us information about the object. While the signal is traveling toward us, the object's position changes. If we photograph some objects that are at different distances from the camera, the information about the positions of these bodies is recorded by the light that arrives at the film simultaneously. But the images of the bodies on this photograph reflect the positions of these bodies at different times. These times are shifted into the past in relation to the moment the picture was taken.

Now let's get back to the question posed earlier: why can't an observer see the Lorentz transformation?

Everything I've said about photographing several objects applies as well to one object that has finite dimensions. If the object is stationary and the exposure time is long enough so that the signal (light) has time to reach the lens from the object's most distant point, we can say with confidence that the photograph correctly represents both the body's position and its dimensions. It's another matter if the object is in motion.

Photographing a moving body presents a complex scenario. The different points of the body are depicted in the positions they occupied at different times. For instance, let the body being photographed approach the camera. Then those points of the body that are farther from the lens when the shutter was opened will seem to lag in comparison with the closer points, so that the images of these points on the photo correspond to an earlier moment in time. Thus, on the photo the object appears stretched in the direction of travel. It's clear that this apparent stretching will be particularly noticeable if the body's velocity is close to the speed of light.

Calculations showed that the apparent stretching will compensate for the Lorentz transformation. So an observer can't see the change in the shape of a rapidly moving body—a ball will still look like a ball.

How can we check the correctness of our reasoning? We need to photograph a moving body whose velocity is close to the speed of light. In this case both the shortening in the direction of travel and the time lag of the signals from different points of the body will be more perceptible. And

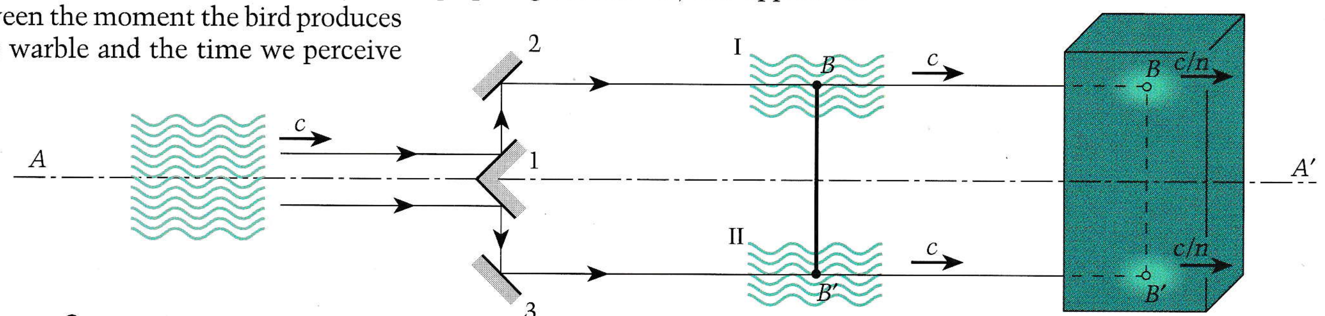


Figure 3
The angled mirror 1 divides the incident light packet into two identical parts: half of the packet is reflected toward mirror 2, the other half toward mirror 3. The mirrors are positioned such that the two light packets travel side by side along parallel paths. These packets are connected by the imaginary line BB′, thus forming a "dumbbell" made of light. In a medium with a refractive index n, the dumbbell moves with the velocity c/n.

that's why Duge and his colleagues created a dumbbell out of light.

To make the dumbbell, you just split one light packet into two halves—two packets. Then you separate these halves to get two wave packets in parallel motion. And there you have your light dumbbell (you have to imagine the rod stretching between the packets).

You can make the dumbbell with the system of mirrors shown in figure 3, for example. The dumbbell's axis BB' is perpendicular to the line AA' and is cut in half by that line. Rotating the entire system of mirrors about the axis AA', we can change the orientation of the dumbbell in the vertical plane perpendicular to the axis AA'. In figure 3 the dumbbell is shown in the vertical position; if we turn the system of mirrors 90° (in either direction relative to the axis AA') we get a horizontal dumbbell, and so on.

A light dumbbell was sent into a vessel filled with turbid water. The packets in the dumbbell entered the vessel simultaneously and were photographed.

Figure 4 presents photographs of dumbbells turned in different directions relative to the axis AA'. The top photo was taken of a vertically oriented dumbbell, and in the photo the dumbbell is also vertical. The second photo was taken of a dumbbell that was rotated approximately 30° from the vertical position about the axis AA'. The third photo shows the image of a dumbbell rotated approximately 60° from the vertical position. The bottom photo depicts a horizontal dumbbell. In all the photos the light packets in the dumbbell are moving from left to right, as in figure 3.

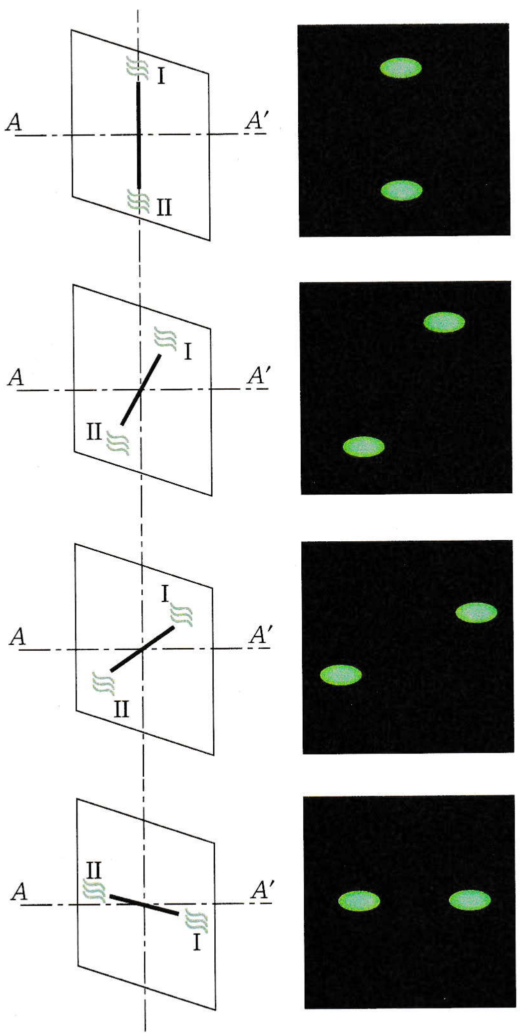


Figure 4
Photographs of light dumbbells oriented differently relative to the camera's line of sight. The exposure time is 10^{-11} s.

These photos are surprising. Only the top photo perhaps turned out the way we'd expect. The two packets in the dumbbell are depicted one above the other—just the way it should be if we're photographing a vertically orientated dumbbell. The other photos offer a strange picture of the phenomenon. The bottom photo is perhaps the strangest, so let's look at that one first.

The bottom photo was taken of a dumbbell lying in the horizontal plane perpendicular to the line AA' (that is, parallel to our camera's line of sight). According to all the laws of nature known to us, this photo should show one spot of light, not two—at the moment the picture was taken, the light packet closer to the camera blocks the one farther away. Yet there they are: two packets flying in formation, one after the other. The camera "sees" a dumbbell orientated parallel to the velocity, while in fact the dumbbell is orientated perpendicular to the velocity.

The images in the middle two photos are also unexpected. It's clear that the closer the dumbbell's axis is to the horizontal, the more one of them lags behind the other. Yet we know from the way we set up the experiment that both packets are traveling at one and the same velocity.

This is all easily explained if we recall our reasoning about the kind of information we can extract from a photo. If we have a photo containing two objects, one of which is closer to the camera when the picture was taken, then the image of the farther one corresponds to an earlier moment in time. And this is precisely what we're seeing in the four photographs in figure 4: the packet of the dumbbell that is farther from the lens is depicted at an earlier moment than is the closer packet—that is, the farther packet hadn't yet reached the point where it would be when the shutter was tripped. So the image of the farther packet is shifted backward in the photograph in relation to the image of the nearer packet.

Now it's clear why the image in the photo of the light packet (figure 2) was so blurred. The additional "smearing" arises because the packet has length in the direction of the camera's line of sight.

Conclusion

So you can't always believe your own eyes. The eye receives visual information by way of light waves. A camera essentially mimics the eye. From time immemorial people saw only slowly moving objects. And when we say "slow," we mean slow compared to the speed of light. In this sense the fastest earthly motions are slow, and our vision gives us a correct representation of an object's shape. But if an object is moving with a velocity close to the speed of light, our eye (and a camera as well) gives a distorted picture. Aware of this peculiarity of our vision, we can always make the necessary adjustments and restore the truth. We just need to exercise due caution when interpreting the data supplied by our senses.

Errorproof coding

How can we be sure the message is getting through?

by Alexey Tolpygo

THIS ARTICLE IS DEVOTED to codes. But I'm not going to talk about deciphering secret writing—though it's an interesting mathematical problem to make it comparatively easy for the addressee to decipher the code and practically impossible for everyone else. I'll consider another question: how to transmit information over a distance through inevitable noise so that the recipient can understand its meaning (decode it) in spite of "typos." As a matter of course, it's desirable to transfer the maximum amount of information at minimum cost. And the origin of this problem can be traced back to ancient times. Carving letters on a rock is far from easy work. One would like to carve less and say more. But how? Well, one can denote whole syllables or words, rather than letters, by a single symbol. This is the approach used in hieroglyphic systems of writing. But then we'd have to memorize too many symbols . . .

I think you have a pretty clear idea of the problem before us. To tackle it mathematically, we have to render it in the more formal language of mathematics.

Code space and code distance

First of all, we must establish an alphabet. It can be the 33-letter Russian alphabet or the 26-letter English alphabet, but from the mathematical point of view it's more natural to use fewer characters. It will suffice, for instance, to take the binary Morse code—a dot and a dash (or zero and one). So we adopt the following definition: an *alphabet* is a fixed set of p arbitrary characters.

It's desirable that p be a relatively small number (also, for certain algebraic reasons, in code theory p is usually a prime number or the power of a prime). The characters can be chosen any way we like, and we'll take them to be the numbers $0, 1, 2, \ldots, p-1$.

Further, an arbitrary string of characters of the alphabet will be called a *word*. This doesn't conform to the laws of natural languages: neither in English, nor in any other real languages, do *all* strings of letters make intelligible words. However, we have to reckon with distortions of words when the information is transmitted—for instance, typographic errors introduced as a text is prepared for publication. So it's only reasonable to regard any string of characters as a word—perhaps a senseless one.

So we don't care about the meaning, but we'll impose another, formal restriction on our words that makes the theory simpler: we'll assume that all words are the same length.

The set of all words—both intelligible and senseless—will be called the *code space*. The set of words that make sense (in a real language, its "vocabulary") in mathematical code theory is simply called the *code*. However, we'll use the words "vocabulary" and "code" interchangeably.

Exercise 1. How many words are there in a code space if the number of letters in the alphabet is p and the length of a word is n?

The fundamental notion in code theory is the so-called *Hamming distance* (or just *distance*) between two words,[1] defined as the number of places in which these words differ from each other.

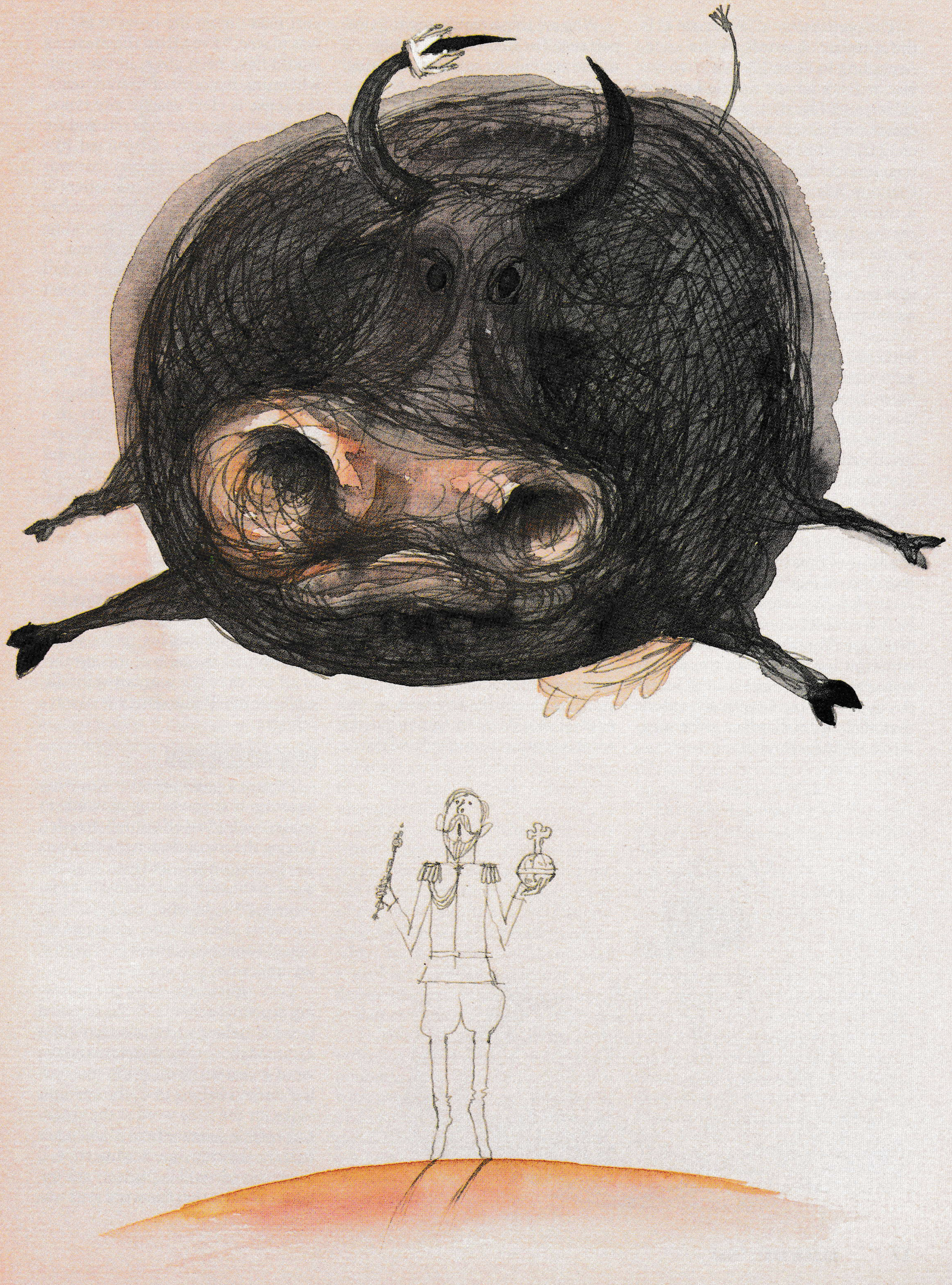
For example, the words "absorption" and "adsorption" are quite close to each other—not in meaning, but in writing; whereas the word "aberration" is rather far from them, though it's a little bit closer to the first of them. In terms of code distance this is expressed numerically: the distance between the first two words is 1; between the first and third, the distance is 3 (the symbols in the third, fourth, and sixth places don't coincide); between the second and third, the distance is 4. We'll denote the code distance between the words a and b as $d(a, b)$.

Exercise 2. Find the code distance between the words "drawer" and "lawyer," "room" and "moor," "01121" and "21221."

The vocabularies of natural languages are out of our control: they are created by peoples and history. But the vocabulary, or code, that we're going to study is to be composed by us. So what are the properties we'd like to require of this code? To formulate them, we need to introduce two more notions.

[1]In mathematics, the terms "space" and "distance" can be used in a very broad sense. Another example of a very unusual interpretation of these words can be found in the article "In the Curved Space of Relativistic Velocities."

Art by Pavel Chernusky



The *cardinality* $|C|$ of a code C is the number of words in the code. According to Ilf and Petrov, two Soviet writers whose novels became classics of Russian satire, Shakespeare used a code with a cardinality of 12,000, while one of their characters, Ella the Cannibal—a very pretentious lady of very limited (to put it mildly) intellectual power—managed with a code of cardinality 30. Of course, one can hardly get by with such a small code in reality. So our first requirement of the code is that its cardinality be as great as possible. (We can certainly enlarge its cardinality at the expense of increasing n and p. So to get a mathematically interesting problem, we must consider these two numbers fixed.)

The *code distance* $d(C)$ is the minimum distance between code words. The second demand on a code is that $d(C)$ be as large as possible.

I'll explain a little later why it's good to have a large code distance. Right now I just want to point out that the two requirements contradict each other. And this is how it has to be if we want to pose a meaty math problem. The problem consists of constructing a code with maximal code distance for a given cardinality; or, vice versa, constructing a code of maximal cardinality for a given magnitude of code distance.

Why are these requirement so important? The first is clear enough. As to the second, let me tell you a historical anecdote.

Reporting on the accession to the throne of the last tsar, Nicholas II, one of the provincial Russian newspapers, through the oversight of a typesetter, informed its readers that "upon the head of His Imperial Majesty was placed the crow." The next day the paper explained with deepest apologies that, of course, one should have read the report otherwise—as ". . . was placed the cow." Bad luck! The error corrupting the meaning crept into the correction, too! In Russian the words "crown," "crow," and "cow" are "корона," "ворона," and "корова," so you see that the cause of the provincial newspaper's troubles was that the distance from the correct word to both of the misprinted words is 1: only one mistaken letter completely changes the meaning.

I must note, however, that the newspaper's troubles were of a political rather than semantic nature: although the readers laughed at the amusing errors, they understood exactly what was being placed on the royal noggin. And this is due to the redundancy of newspaper text in Russian, or generally any text in any natural language. Imagine that the word was simply omitted and the report read "upon the brow of H.I.M. was placed." Even then the meaning would have been clear. In the overwhelming majority of literary texts, the sense is clear despite any typos. Exceptions are extremely rare and are usually deliberately constructed, like the notorious Latin phrase that was sent, as legend has it, to the jailers of King Edward II: "*Edwardum occidere nolite timere bonum est.*" Depending on where you place a comma, it means either "Don't be afraid to kill Edward, it's a good deed" or "Don't kill Edward, it's good for you to refrain from this deed."

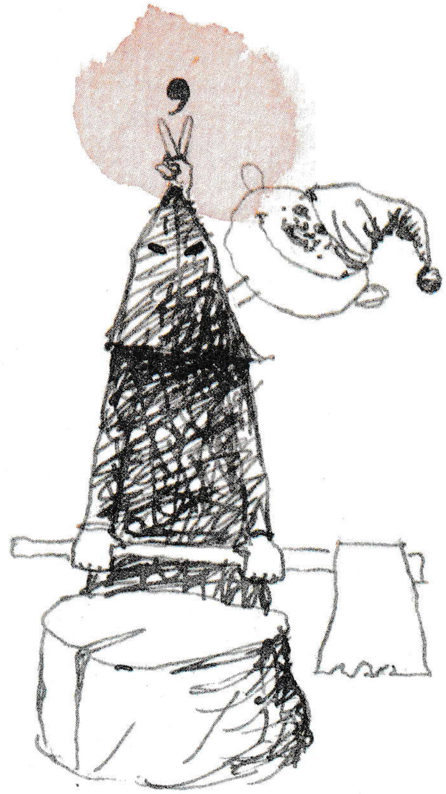
The redundancy of natural languages is much greater than that of mathematical text. Indeed, imagine seeing the following equality in a book: "$6 \div 3 = 3$." Clearly, the typesetter has made an error. But where? Should you read 9 instead of 6, or 2 instead of one of the 3's? Or, perhaps, the sign was misprinted and one should read "$6 - 3 = 3$"? Other readings are also possible.

In code theory the intuitive concept of redundancy has an exact formal definition: the *redundancy* of a code is the number $\log_p(p^n/|C|)$—the logarithm of the ratio of the number of all words to the number of intelligible words. Redundancy is always positive.

Error-detecting codes

Picture a radio operator receiving a certain text consisting of words in some code ("intelligible" words). During the transmission the text was corrupted by noise, and instead of the word a, the distorted word a' was received. Will the operator be able to correctly decode the message received—that is, ascertain that word a' was indeed a?

Let's offer our operator the following decoding scheme. If the received word a' belongs to the code C, it's assumed there were no errors and the word is written down as the true one. If a' doesn't belong to C, the operator looks for the code word at the smallest possible distance from a' and takes that for the true one. For instance, if the received word is "wrisc," the operator decides that the word sent was



"wrist." It's not inconceivable, of course, that the original word was "brisk," but that would mean two errors at a time (the distance between "wrisc" and "brisk" is 2). And it's only natural to assume there was only one mistake.

Does this scheme always yield the correct result? If there are no stipulations restricting the code and the conditions of transmission, we certainly can't guarantee anything. It might be that all the letters in a word were transmitted incorrectly. Also, it's not clear what to do if there are several words equally distant from a': if the word comes through as "buttor," does it mean "butter" or "button"? And so forth.

Under certain conditions, however, this scheme always yields the correct result from decoding.

THEOREM. *Let $d(C) \geq 3$. Then the result from decoding according to the rule described above is always correct provided that there is not more than one error in every transmitted word.*

Indeed, if a is the transmitted word and a' the received one, then $d(a, a') \leq 1$. On the other hand, for any word $b \in C$, $b \neq a$, we have $d(a, b) \geq 3$. Although our distance isn't a "real" one, it's easy to verify that it satisfies the Triangle Inequality $d(a, b) + d(b, c) \geq d(a, c)$. Applying it to our situation, we quickly see that a' is closer to a than to any other word b:

$$d(a', b) \geq d(a, b) - d(a', a) \geq 2 > d(a', a).$$

Therefore, a' will be decoded as a, which is just what we need.

According to this theorem, we can say that a code C with a code distance $d(C) \geq 3$ *corrects one error.*

Exercise 3. Prove a more general theorem: if $d(C) \geq 2r + 1$, then the code corrects r errors.

Now we see why too low a redundancy is undesirable: such a text is more informative (a good thing, seemingly), but for that very reason it's extremely difficult if not impossible to correct errors in the text. Mathematical texts offer a dramatic example of this: the typeset versions sent to the authors for proofing usually abound in misprints that the typesetters fail to notice.

On the other hand, too high a redundancy is also undesirable. It's easy to construct a code that corrects one or several mistakes: one can simply repeat each letter of a message a certain number of times (for instance, instead of "fox" write "fffoooxxx"); to correct, say, three errors, one should make seven repetitions of each letter. But this method is obviously uneconomical.

Now that our task has been outlined more clearly, I want to describe one approach to its solution. We'll set a certain code distance—say, 2 or 3—and try to construct a code with as big a cardinality as possible. To begin with, we'll consider a problem involving a code that doesn't correct errors but enables us to detect them.

Telephones in a certain city have 6-digit numbers. How many telephones can be installed in this city such that any two numbers differ in at least two digits? (This would mean that a connection doesn't occur when a number is dialed with a single error, so this error can thereby be detected.)

In our terms, a "word" here is any set of six digits, and "intelligible word" is the number of an operational telephone. Obviously, there are 1,000,000 words in all. What's the greatest number of intelligible words?

It's easy to see that this number doesn't exceed 100,000—otherwise, by the so-called pigeonhole principle[2] one could find two numbers having the same first five digits (there are only 100,000 different 5-digit numbers), and these two numbers would differ only in the sixth digit. But can we find 100,000 numbers that satisfy the condition?

This problem was posed at the 31st Moscow Math Olympiad and was solved there by only one participant. In a sense, it's a "classic" olympiad problem—very difficult to get a handle on, but once it's solved you need only a couple of lines to write it down. Here, we can simply take all the numbers whose *sum of digits is divisible by 10.*

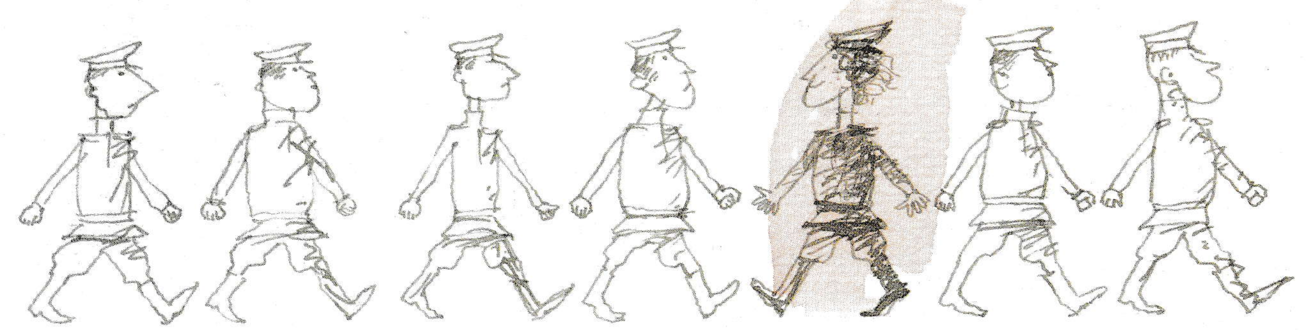
Indeed, any two such numbers can't differ in a single place, because the difference between the sums of digits of $abcdef$ and a_1bcdef is $a - a_1$, which is of course less than 10. On the other hand, we can put arbitrary digits in the first five places and choose the sixth digit so as to make the total sum divisible by 10. Therefore, there are as many numbers divisible by 10 as there are different 5-digit numbers—that is, 100,000.

Exercise 4. Consider all the telephone numbers $x_1x_2x_3x_4x_5x_6$ such that the sum $a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6$ is divisible by 10. What numbers $a_1, a_2, \ldots, a_6$ are needed to make the code single-error-detecting?

Linear codes

I want to emphasize one very important feature of the code above: the letters of our alphabet are not just some symbols—they're numbers that can be put through arithmetic operations (if only just addition). It's the algebraic structure of a code that enables us to construct efficient codes,

[2]See, for instance, "Pigeons in Every Pigeonhole" in the January 1990 issue of *Quantum*.—*Ed.*



which is why the whole theory is often called algebraic code theory.

Let's turn back to the alphabet with the symbols 0, 1, ..., $p-1$. These symbols can be viewed as remainders after division by p. Then, to find the sum of any two of them, we must add them as ordinary numbers and take the remainder of the sum when divided by p. Subtraction and multiplication are defined similarly. These operations are called *addition, subtraction, and multiplication modulo p*. For example, for $p = 5$ we write: $3 + 4 \equiv 2 \pmod 5$, $3 \cdot 4 \equiv 2 \pmod 5$ (because both $3 + 4 = 7$ and $3 \cdot 4 = 12$ have a remainder of 2 when divided by 5).[3] In particular, it follows from what's been said in this article that addition and multiplication of remainders obey all the laws that normal arithmetic operations obey except that division—the operation inverse to multiplication—is always possible only for a prime number p. So from now on we assume that p is a prime (that is, the alphabet consists of a prime number of characters), and mostly p will equal 2.

As in the above telephone problem, one can easily construct a code detecting a single error for two possible characters 0 and 1 ($p = 2$): a word of zeros and ones $x_1, x_2, \ldots, x_n$ belongs to C if the number of ones is even. Clearly, any two such sets can differ in not less than two places (more exactly, in an even number of places)—that is, $d(C) \geq 2$. A more formal definition of this code is given by the equation $x_1 + x_2 + \ldots + x_n \equiv 0 \pmod 2$, which distinguishes the words from the code. So we can say that this code is given by a linear equation.

Let's extend this idea a little bit. A code can be defined by a system of several linear equations:

$$\begin{cases} a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0, \\ b_1x_1 + b_2x_2 + \cdots + b_nx_n = 0, \\ \vdots \\ c_1x_1 + c_2x_2 + \cdots + c_nx_n = 0, \end{cases}$$

where the unknowns x_i and the coefficients a_i, b_i, ..., c_i are all regarded as remainders modulo p. That is, the coefficients take the values 0, 1, ..., $p-1$, and the equal signs mean that the left sides of all these equations, calculated in the conventional way, are divisible by p—that is, yield zero remainders when divided by p. We assume that the number of unknowns is much greater than the number of equations, so the system has a large number of solutions (although a finite number, of course, because there are only p^n strings $(x_1, \ldots, x_n)$). The set of all solutions of such a system is called a *linear code.* The code we constructed above (with $d(C) = 2$) is an example of a linear code, since it's defined by one equation with n unknowns.

What is the cardinality of a linear code? Without delving into the theory of linear systems, I'll simply say that each equation enables us to express one unknown in terms of the others. So if there are n unknowns and m equations, then (in general!) the values of $n - m$ unknowns may be assigned arbitrarily—which can be done in p^{n-m} ways (compare exercise 1)—while the remaining m unknowns are uniquely expressed in terms of the first $n - m$ unknowns. Thus, the cardinality of the code is p^{n-m}, and its redundancy equals $\log_p (p^n/p^{n-m}) = m$.

What does this mean in practice? Suppose we can assign arbitrarily the first $n - m$ unknowns. Then you can write your message in the first $n - m$ places of each word—you can put whatever you want there. What are the other places for? They're called *control places*. When they're filled up so as to satisfy our system of equations, they ensure that any error violating the equations inevitably shows up. And not only that: later on we'll see that the proper choice of a system even allows the person receiving the message to correct such errors.

In order to understand how many equations will suffice to correct one or several errors, let's continue to explore the "geometry" of the code space.

A *sphere* with radius r and center a, where a is some word, is defined as the set of all words b such that $d(a, b) \leq r$.

Apparently, this is just the same as the usual definition of a sphere. The unusual thing about it is that a sphere consists of a finite number of points; this number is called the *volume* and is denoted by V_r (r is the radius of the sphere). It follows that a sphere of zero radius has a nonzero volume $V_0 = 1$. Then, as the radius increases from zero to one, the volume remains unchanged, because there are no points lying a fractional distance from the center a. But when r becomes equal to one, the volume makes a leap: the sphere captures the points a unit away from a. The next leap occurs at $r = 2$, and so on. So the sphere of radius r is a union of a finite number of "spheres" of radii 0, 1, 2, ..., ρ (ρ is the largest integer not exceeding r), and the volume of each sphere is positive.

For instance, if $p = 2$, then the

[3]More details can be found in "Go Mod with Your Equations" (*Quantum*, May/June 1992).—*Ed.*

sphere of radius 1 centered, say, at zero—the point $(0, 0, \ldots, 0)$—comprises all strings $(x_1, x_2, \ldots, x_n)$ that differ from the zero string in only one place. They have the form $(0, \ldots, 0, 1, 0, \ldots, 0)$, and, of course, their number equals the number of places—that is, n. A sphere of radius 2 consists of words having two ones and $n - 2$ zeros; its volume is equal to $C(n, 2) = n(n-1)/2$ (see, for instance, "Combinatorics–polynomials–probability" in this issue). So $V_1 = 1 + n$, $V_2 = 1 + n + n(n-1)/2$.

Exercise 5. Find the volumes of spheres of radii 1 and 2 for an arbitrary p.

Let's recall the single-error-correcting algorithm considered above: a received word a' is corrected to a if and only if a belongs to the code C and $d(a, a') \le 1$—that is, if a' is contained in the sphere with radius 1 centered at a. Clearly, a necessary and sufficient condition for this algorithm to work successfully is that unit spheres with centers at all words of the code are disjoint. If we want the code to correct r errors, the spheres of radius r should be disjoint. But then the inequality $|C| \cdot V_r \le p^n$ or $|C| \le p^n/V_r$ should hold. In particular, if $p = 2$ and the code corrects one error, $|C| \le 2^n/(n+1)$.

Take, for instance, $n = 7$. Then the cardinality of the code must satisfy $|C| \le 2^7/8 = 2^{7-3}$, so it should be defined by not less than three equations. If n is the length of a word in a code and m is the number of equations that define it, then the number $(n-m)/n$ is called the *transmission rate* or the *code rate*. It's desirable to make this number as large as possible. But as we've just seen, for, say, $n = 7$, the rate can't be greater than 4/7. Now I'll describe a code whose rate is exactly 4/7.

The Hamming Code

Consider an arbitrary system of three equations in seven unknowns—for example,

$$\begin{cases} x_1 + x_2 \qquad + x_4 \qquad\qquad\qquad = 0, \\ \qquad x_2 + x_3 + x_4 + x_5 + x_6 \qquad = 0, \\ x_1 \qquad\qquad\qquad\qquad + x_6 + x_7 = 0 \end{cases}$$

(x_i take two values, 0 and 1, and are added modulo 2). A word $(x_1, x_2, \ldots, x_n)$ is included in the code if it satisfies all three equations. What happens if it is transmitted with one error?

By way of illustration, consider a solution 1001101 of the system and make an error in the third place. The incorrect word 1011101 is no longer a solution to the given equations. However, it still satisfies the first and third equations, since they don't involve x_3 (in other words, their coefficients of x_3 are zero). Only the second equation is violated. In fact, let's plug the ordinates of the incorrect word into the three equations for a closer look at what goes awry. We find

$$\begin{cases} 1 + 0 \quad + 1 \qquad\quad = \mathbf{0} = 0 \\ \quad 0 + 1 + 1 + 1 + 0 = \mathbf{1} \ne 0 \\ 1 \qquad\qquad + 0 + 1 = \mathbf{0} = 0 \end{cases}$$

Look at the column of "wrong" answers highlighted above:

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

This turns out to be the column of coefficients of x_3 in the given equations. And this is no coincidence: after making an error in the ith place, a little thought will show that we always get a column of coefficients of the ith unknown in the left sides of our equations.

However, our radio operator, after receiving the word 1011101, will be unable to tell exactly where the error was made: the word 1011001, like the correct one 1001101, also belongs to the code and differs from the received word in only one place (the fifth). The reason for this ambiguity is clear: the columns of coefficients of x_3 and x_5 in our equations are the same.

Well, then all we have to do is create a system of equations whose columns of coefficients are all different!

The columns written side by side form a 3×7 array of zeros and ones called the matrix of a system. For the system considered above, it has the form

1101000
0111110
1000011.

The matrix we need is actually unique up to the order of its columns, because there are only $2^3 = 8$ different triples of zeros and ones, and one of them is three zeros (which is senseless to use). So we simply write down all nonzero triple columns in any order—say,

1110100
1101010
1011001,

and get the required system

$$\begin{cases} x_1 + x_2 + x_3 \qquad + x_5 \qquad\qquad = 0, \\ x_1 + x_2 \qquad + x_4 \qquad + x_6 \qquad = 0, \\ x_1 \qquad + x_3 + x_4 \qquad\qquad + x_7 = 0. \end{cases}$$

The code defined by these equations is called a *Hamming code*. If we start with $n = 15$, we find that $|C| \le 2^{15}/16 = 2^{15-4}$, so $m = 4$. We can thus use the method outlined above to construct a Hamming code with the transmission rate 11/15. If $n = 31$, we get a code with a rate of 16/31, and so on. Each of them corrects one error, and the transmission rate approaches 1 as n increases. One should be aware, however, that the longer the words of a code, the greater the danger that more than one mistake will be transmitted.

Exercise 6. In the Hamming code given above, the word 1010101 was received. What word was transmitted?

The code that corrects one error was constructed by R. W. Hamming in the late 1940s. A new problem naturally posed itself: how to construct a code that corrects two or more errors. That turned out to be much more difficult. It wasn't solved for another decade. The resulting codes are called BCH codes in honor of the mathematicians who invented them: R. Bose, D. Chaudhuri, and A. Hocquenghem. But they deserve a separate article—here we leave the topic for the time being.

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

The Worm Problem of Leo Moser

Part II: The story of the shrinking blanket

by George Berzsenyi

AS PROMISED IN PART I OF this account, in this article I'll describe the progress made on Leo Moser's Worm Problem to date. In each of the figures, the linear distance between A and B is 1 unit—the length of the worm. So they will automatically accommodate the worms, which assume the shape of a straight line segment. The corresponding areas are given to five-decimal-point accuracy.

Figure 1 is a sector of a circle of central angle $2\theta_0$ and radius $\frac{1}{2}\csc\theta_0$, where θ_0 is the smallest positive root of the equation $\tan\theta = 2\theta$. The area of this sector is $\frac{1}{2}\theta_0 \csc^2\theta_0$, whose value is approximately 0.34501. This discovery was made in 1972 by Jack Wetzel, who has done much to popularize the Worm Problem.

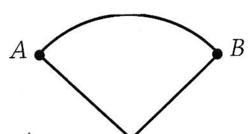


Figure 1

Figure 2 is the union of an isosceles triangle of altitude 1/4 and base 1 and a semi-ellipse of major axis 1 and minor axis 1/2. The area of this figure is approximately 0.32135. It was discovered by John Garriets shortly after Wetzel announced his discovery.

Figure 3 is a rhombus with major diagonal 1 and minor diagonal $\sqrt{3}/3$. The are of this figure is approximately 0.28867. It was discovered by John Garriets and George Poole around 1974.

Figure 4 is a truncation of the rhombus in figure 3. In it, CD is parallel to AB and is of length $1 - \sqrt{3}/2$. The area of this figure is approximately 0.28608. This result is also due to Garriets and Poole and appeared, along with figure 3, in their joint 1974 article in *The American Mathematical Monthly*.

Finally, figure 5 is a 60° sector of a circle of radius 1/2 with a 30°–60°–90° triangle joined to either side. Its area is approximately 0.27523. This is the latest discovery, which I referred to in part I. It was made by Rick Norwood and my friends, George Poole and Michael Laidacker. Their article appeared on pages 153–62 of *Discrete Computational Geometry*, Vol. 7 (1992). For exact references in the literature, the reader is referred to the references at the end of their article.

My first challenge to *Quantum* readers is a minor exercise: **verify that the areas of the shapes in figures 1 through 5 are as reported.** My second challenge is to **show that each of these shapes will cover the ⊔-worm**, consisting of three line segments of length 1/3 at right angles to one another. Next, try to **reconstruct the proofs that each of these shapes covers all worms of length 1.** Finally, and most importantly, attempt to **make successful improvements on them and propose conjectures of your own.**

In part III of this account, I'll tell you about some of the conjectures known to me, and I'll describe two other "special worms" that you'll need to keep in mind as you make your own conjectures. ◙

The purpose of this column is to direct the attention of *Quantum*'s readers to interesting problems in the literature that deserve to be generalized and could lead to independent research and/or science projects in mathematics. Students who succeed in unraveling the phenomena presented are encouraged to communicate their results to the author either directly or through *Quantum*, which will distribute among them valuable book prizes and/or free subscriptions.

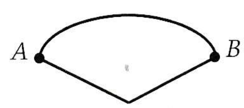


Figure 2

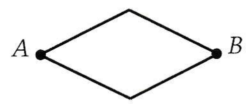


Figure 3

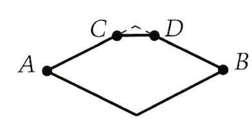


Figure 4

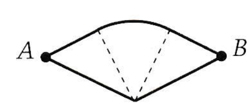


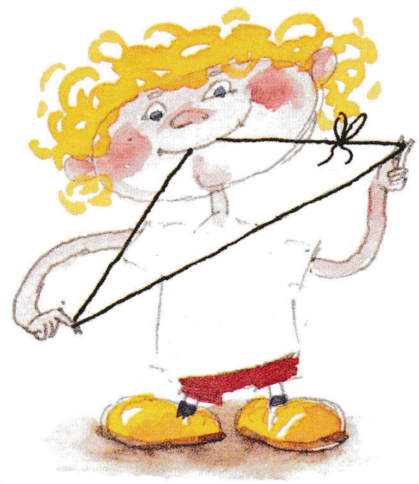
Figure 5

BRAINTEASERS

Just for the fun of it!

B76

Blond and blue-eyed. The proportion of blonds among blue-eyed persons is greater than among the population as a whole. Is it true that the proportion of blue-eyed people among blonds is greater than among the entire population? (A. Savin)

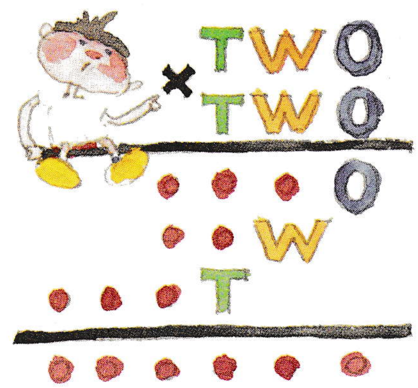


B77

Long heights. Does there exist a triangle, two of whose heights are not shorter than the sides on which they are dropped? If it does, what are its angles? (A. Savin)

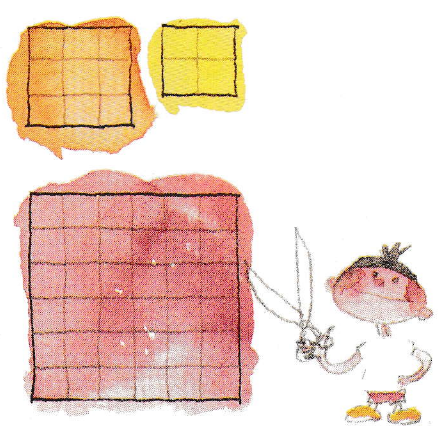
B78

Dripping hot and cold. Two identical laboratory pipettes are filled with water to the same level. The water is cold in one pipette and hot in the other. As the pipettes are emptied, the drops are counted. From which pipette will more drops fall? (A. Buzdin)



B79

Two times two. Each letter in the "alphametic rebus" shown in the figure stands for some digit—different letters denote different digits, dots denote arbitrary digits. What number is TWO? (A. Shvetsov)



B80

Neat shearing. You have to make one square out of the three squares—2×2, 3×3, and 6×6—shown in the figure. How can you do this, cutting the squares into the smallest possible number of pieces? (V. Proizvolov)

Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON PAGE 59

AT THE BLACKBOARD I

Combinatorics—polynomials—probability

Anagrams of words, numbers, sweet peppers . . . anything

by Nikolay Vasilyev and Victor Gutenmacher

THIS ARTICLE WILL BE UNderstandable to everybody familiar with the most elementary algebra. But we're sure that high school juniors and seniors will also find it interesting. We're going to demonstrate how the same combinatorial formula penetrates a number of branches of mathematics and its applications: combinatorial theory, polynomial algebra, and probability.

Factorials

There is a convenient formula for calculating the sum of the first n natural numbers (positive integers):

$$1+2+\ldots+n=\frac{n(n+1)}{2}.$$

The product of the first n natural numbers can't be expressed by a similar formula, but this value, which occurs often in combinatorics and other mathematical fields, has a special notation: $n!$ (read as "n factorial"). The choice of the exclamation mark perhaps has to do with the magnitude of this number, which is very large even for moderate values of n. To show how fast $n!$ grows with the growth of n, let's compute it for n from 1 to 10:

$$\begin{aligned}
1! &= 1,\\
2! &= 1\cdot 2 = 2,\\
3! &= 1\cdot 2\cdot 3 = 6,\\
4! &= 3!\cdot 4 = 24,\\
5! &= 4!\cdot 5 = 120,\\
6! &= 720,\\
7! &= 5{,}040,\\
8! &= 40{,}320,\\
9! &= 362{,}880,\\
10! &= 3{,}628{,}800.
\end{aligned}$$

The definition of $n!$ implies the following formula, relating the factorials of two subsequent natural numbers n and $n + 1$:

$$(n+1)! = n!\cdot(n+1). \qquad (1)$$

To find the product of numbers from 1 to $n + 1$, the product of numbers from 1 to n must be multiplied by one more factor, $n + 1$.

Notice that, plugging $n = 0$ into (1), we get $1! = 0!\cdot 1$; for this reason it's assumed that $0! = 1$. This agreement proves reasonable and useful in various general formulas.

Problems

1. Find $n!$ for $n = 11, 12$.
2. Is it possible for $n!$ to end in exactly five zeros? What is the smallest n such that the number $n!$ ends in six zeros?
3. Prove the formula $(n + 1)! - n! = n!\cdot n$.
4. Find the sum $1\cdot 1! + 2\cdot 2! + \ldots + n\cdot n!$. (Use the previous problem.)
5. Check the equality

$$\frac{n!}{k!(n-k!)}+\frac{n!}{(k-1)!(n-k+1)!}=\frac{(n+1)!}{k!(n-k+1)!}$$

for $n = 7$, $k = 3$, and prove it for all natural n and k, $0 < k \le n$.

6. Find four triples x, y, z such that $x!\cdot y! = z!$. (Substitute $n = k! - 1$ into formula (1).)

Permutations

Factorials appear in the most natural way when we count the number of permutations of different objects.

Let's take the four letters B, U, S, H and find how many ways they can be arranged in a row—that is, how many words can be compiled from these letters. It turns out that this number of ways equals $4! = 4\cdot 3\cdot 2\cdot 1 = 24$. Indeed, any one of the four letters can be put in the first place, any of the three remaining can take the second place, any of the two unused ones can take the third place, and finally the last letter will find itself in the fourth place. All these permutations are written out in figure 1. Permutations of the letters of some word are called its *anagrams*.

Here's one more example. Consider all the permutations of ten figures 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. They can be viewed as 10-digit numbers if they do not begin with zero, and as 9-digit numbers if they do. So in all there are 10! numbers of both kinds.

These examples illustrate a general statement:

> The number of permutations of n different things is equal to $n!$.

It's often necessary to select from all permutations only those possessing a certain property. For instance, from the anagrams of the word

Art by Pavel Chernusky


DEAR MUTUNAQ REVEY AYD

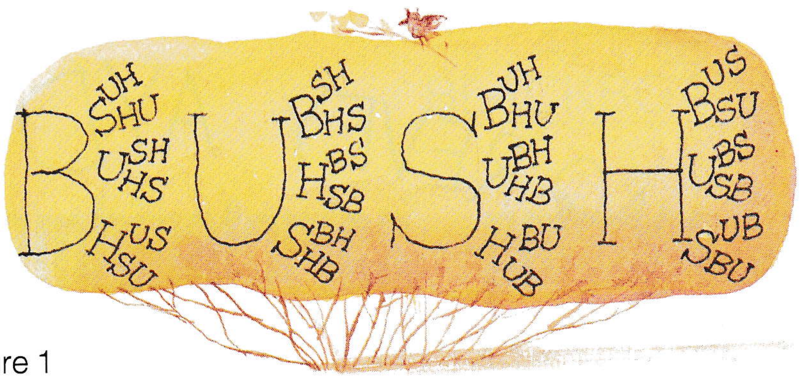


Figure 1

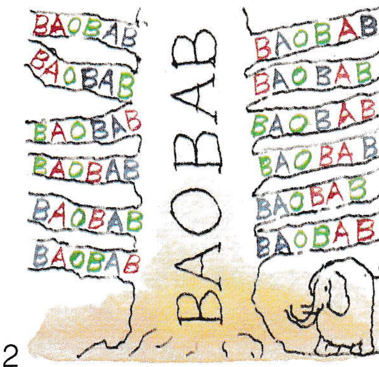


Figure 2

BUSH, written out in figure 1, only one (besides the word BUSH itself) makes sense in English: HUBS. Among the 10! numbers specified above, there are exactly four that are divisible by all the numbers from 2 to 18: 2,438,195,760; 3,785,942,160; 4,753,869,120; and 4,876,391,520. To find anagrams that make sense, or the above four numbers among all 10! numbers, one has to search through a lot of cases. Similar problems repeatedly arise in manufacturing and economics when optimal variants must be found.

Problems

7. (a) How many anagrams does the "word" NISELAP have? (b) Find among them the name of a breed of dog.

8. Find two 9-digit numbers comprising all the figures 1, 2, 3, 4, 5, 6, 7, 8, 9 such that one of them is 8 times the other one.

9. (a) The vertices of a regular hexagon drawn on the plane are to be labeled with the letters A, B, C, D, E, F. In how many ways can this be done? (b) How many of these ways yield a hexagon ABCDEF? (The letters can be read either clockwise or counterclockwise.)

10. In how many ways can eight rooks be placed on a chessboard so that none of them attacks another one?

Permutations with repetitions

If some letters of a word are the same, the number of anagrams—*permutations with repetitions*—is expressed by a ratio of factorials. For instance, the number of anagrams of the word BAOBAB is $6!/3!2!1! = 60$.

Why is that? In this word the letter B is repeated 3 times, A is repeated twice, and O occurs once. Imagine that all these letters are different: for instance, three B's and two A's may be colored different. Then we'd have 6 distinguishable symbols that can be ordered in 6! different ways. But then, each anagram of the word—OAABBB, BOAABB, AAOBBA, ...—will correspond to 3!2! permutations of these symbols (fig. 2), because the three B's in it can be rearranged in 3! ways, while the two A's can be rearranged in 2! ways.

In the general case, the number of permutations with repetitions is given by the following formula:

> If a word consists of n_1 letters A_1, n_2 letters A_2, ..., n_r letters A_r, the number of its anagrams equals
>
> $$\frac{(n_1 + n_2 + \ldots + n_r)!}{n_1!n_2!\ldots n_r!}.$$

Of course, this formula works as well with permutations of anything. For example, the number of permutations of the figures 0, 0, 0, 0, 1, 1, 3 is equal to

$$\frac{7!}{4!2!1!} = \frac{4!\cdot 5\cdot 6\cdot 7}{4!\cdot 2} = 5\cdot 3\cdot 7 = 105.$$

Problems

11. How many anagrams do the following words have: (a) ANAGRAM; (b) REGISTER?

12. If a mother has 3 bananas, 2 pears, and 2 oranges, in how many ways can she give the fruit to her daughter in one week, one piece of fruit per day?

13. Decode the phrase in the drawing on the previous page, in which every word is replaced by an anagram.

14. How many ways are there to make a necklace of one black, two white, three red, and five blue beads?

The power of a sum

Permutations of letters (anagrams) naturally arise when two or more polynomials are multiplied, and the above combinatorial coefficients (the numbers of anagrams) arise when like terms are collected.

You know very well how to square the sum of two numbers:

$$(a + b)^2 = a^2 + 2ab + b^2.$$

A similar expression can be obtained for the square of the sum of three or more terms. Let's square, for instance, the sum $a + b + c$:

$$\begin{aligned}(a + b + c)(a + b + c) &= aa + ab + ac + ba + bb + bc + \\ &\quad ca + cb + cc \\ &= a^2 + b^2 + c^2 + 2ab + 2bc + 2ac.\end{aligned}$$

A similar formula is valid for $(a + b + c + d)^2$ (in figure 3 each monomial expresses the area of the respective rectangle). Generally, to compute the square of the sum of n numbers, you have to add together the squares of all n numbers and the doubled products of all pairs of the numbers.

When $a + b + c$ is meticulously multiplied by itself twice, the following formula emerges:

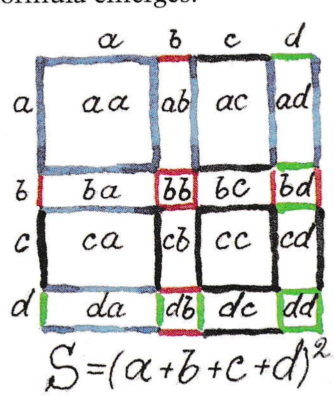


Figure 3

	a	b	c	d	e
a^2b	2	1	0	0	0
b^2c	0	2	1	0	0
b^2a	1	2	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮

	a	b	c	d	e
a^3	3	0	0	0	0
b^3	0	3	0	0	0
c^3	0	0	3	0	0
⋮	⋮	⋮	⋮	⋮	⋮

	a	b	c	d	e
abc	1	1	1	0	0
abd	1	1	0	1	0
abe	1	1	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮

$$(a + b + c)^3 = a^3 + b^3 + c^3 + 3(a^2b + a^2c + b^2a + b^2c + c^2a + c^2b) + 6abc.$$

The coefficients 1, 3, and 6 that arise here can be obtained without the tedious routine of multiplying and collecting like terms: they're just the numbers of anagrams, which we already know. The term a^3, say, can appear in a unique way—by choosing the letter a from each of three multiplied sums

$$(a + b + c)(a + b + c)(a + b + c).$$

To obtain the term a^2b, we must select b from one of the sums and a from the other two—that is, the anagrams corresponding to this term are *aab, aba, baa*; their number, which we know to be equal to $(2 + 1)!/(2!1!) = 3$, is just the coefficient of this term. Finally, the coefficient of abc is equal to $3!/(1!1!1!) = 6$—the number of anagrams of a three-letter word whose letters are all different.

A similar formula can be worked out for the cube of the sum of a greater number of terms. For example,

$$(a + b + c + d + e)^3 = (a^3 + \ldots) + 3(a^2b + \ldots) + 6(abc + \ldots). \qquad (2)$$

The ellipses in each set of parentheses denote the terms resulting from the first one (written out) after all possible letter substitutions.

Now, the general formula:

> The coefficient at $a_1^{n_1}a_2^{n_2}\ldots a_r^{n_r}$, obtained by raising the sum $a_1 + a_2 + \ldots + a_r$ of r terms to the nth power (here $n = n_1 + n_2 + \ldots + n_r$, $n_1 \geq 0$, $n_2 \geq 0$, ..., $n_r \geq 0$), is equal to the number of anagrams of a word comprising n_1 letters A_1, n_2 letters A_2, ..., n_r letters A_r—that is, to
>
> $$\frac{n!}{n_1!n_2!\ldots n_r!}.$$

(Naturally, if some number n_j equals 0, then $a_j^{n_j} = 1$, so the letter a_j lacks the corresponding monomial—we remind you that $0! = 1$.)

Let's look at formula (2) again. It's interesting that the question "How many terms are there within each pair of parentheses?" is also reduced to counting permutations with repetitions. Let's write out all our five letters, and under every letter we'll write the exponent with which this letter enters a certain monomial (if a letter doesn't occur in it at all, we write the exponent 0—see the tables above). Then the monomials in the parentheses $(a^2b + \ldots)$ will be related to "words" of five figures—one 2, one 1, and three 0s—taken in every possible order, so the total number of these monomials equals $5!/(3!1!1!) = 20$; the number of terms in the parentheses $(a^3 + \ldots)$ will be equal to the number of anagrams of the "word" 30000—that is, $5!/(4!1!) = 5$; and the number of terms of the form abc will equal $5!/(3!2!) = 10$.

We can verify that there was no mistake in our reasoning by counting the total number of all the monomials before collecting like terms—in other words, by substituting 1 for every letter in (2). Then we'll get 5^3 on the left side of the formula and $5 + 3 \cdot 20 + 6 \cdot 10 = 125$ on the right side.

Problems

15. Before collecting like terms, how many terms will appear in multiplying out the product

$$(a + b + c + d)(x + y + z)(u + v)?$$

(Hint: substitute 1's for the letters.)

16. Find the largest coefficient of the polynomials (a) $(a + b + c + d + e)^5$; (b) $(a + b + c + d)^5$.

17. Find coefficients K_1, K_2, K_3, K_4, K_5 such that $(a + b + c + d + e)^4 = K_1(a^4 + \ldots) + K_2(a^3b + \ldots) + K_3(a^2b^2 + \ldots) + K_4(a^2bc + \ldots) + K_5(abcd + \ldots)$. How many terms are there in each set of parentheses? Verify your answer by setting $a = b = c = d = e = 1$.

A way to compute probabilities

As we all know from our everyday experience, a buttered slice of bread falls butter down with a probability of 1/2 and butter up with the same probability. Some people, though, believe that the probability of the first outcome is 0.9 and that of the second 0.1. But hardly anybody will doubt that the probability of getting a 6 when you roll a die is 1/6, and that of getting two 6's in a row is 1/36.

These examples illustrate the general idea that a probability is a number between 0 and 1 that expresses quantitatively the chance of one or another outcome (like butter up or down) of some random event (dropping a buttered slice of bread), and that the sum of the probabilities of all possible outcomes is equal to 1.

When you're trying to solve a probability problem, the first thing to do is assign certain probabilities to all "elementary outcomes" of the random experiment in question, if they aren't specified explicitly in the statement of the problem. (For instance, when a die is said to be *fair*, it means that each of its six faces is equally likely to be rolled and so each has a probability of 1/6.) The probabilities of more complex events are calculated by using laws of mathematical probability theory. For a specific kind of problem, the calculation amounts to raising the sum of several numbers to some power by using the formulas considered in the preceding section. It turns out that the monomials that arise when you do this all have a definite probabilistic meaning.

By way of illustration, consider this (somewhat artificial) example. Imagine a huge pile of sweet peppers in your favorite grocery store. You're told that 1/3 of them are red, 1/2 are yellow, and 1/6 are green. If you take one pepper at random, we naturally assume the probabilities that it's red (r), yellow (y), or green (g) are equal to 1/3, 1/2, and 1/6, respectively. If you take two peppers, one by one, there are 9 possibilities: *rr, ry, rg, yr, yy, yg,*

gr, gy, gg. If the pile is big enough, and if you grab peppers without looking at them, the probability that the second pepper in a pair is, say, yellow, doesn't depend on the color of the first pepper and is equal to the proportion of yellow peppers in the pile—that is, it equals 1/2, no matter what color the first pepper was. So the probability of taking, say, a pair ry (which equals the portion of such pairs in the set of all possible pairs) is 1/2 the portion of pairs in which the first color is red—that is, $1/2 \cdot 1/3 = 1/6$. Similarly, the probabilities of all pairs of colors, listed in the same order as the pairs above, are equal to

$$\left(\frac{1}{3}\right)^3, \frac{1}{3}\cdot\frac{1}{2}, \frac{1}{3}\cdot\frac{1}{6}, \frac{1}{2}\cdot\frac{1}{3}, \left(\frac{1}{2}\right)^2,$$

$$\frac{1}{2}\cdot\frac{1}{6}, \frac{1}{6}\cdot\frac{1}{3}, \frac{1}{6}\cdot\frac{1}{2}, \left(\frac{1}{6}\right)^2.$$

The sum of these numbers is 1. This is easy to see if you notice that all these products crop up when the sum $1/3 + 1/2 + 1/6 = 1$ is squared (before like terms are collected).

To find the probability that one of the peppers is red and the other green, we have to add together the probabilities of rg and gr:

$$\frac{1}{3}\cdot\frac{1}{6}+\frac{1}{6}\cdot\frac{1}{3}=2\cdot\frac{1}{3}\cdot\frac{1}{6}=\frac{1}{9}.$$

Notice that this operation can be viewed as a part of the following: take the square of $(r + y + g)$, collect like terms (in our particular case, rg and gr, which yields $2rg$), and then replace the letters (r and g) with their respective probabilities.

The rule for calculating coefficients in a power of a sum enables us to find probabilities of longer combinations, too. For example, if our farmer chooses at random, one after another, five peppers, the probability that three of them are red, one is yellow, and one is green can be computed as

$$\frac{5!}{3!1!1!}\cdot\left(\frac{1}{3}\right)^3\cdot\frac{1}{2}\cdot\frac{1}{6}=\frac{5}{81},$$

since $(5!/(3!1!1!))r^3yg$ is the monomial arising in the simplified expansion of $(r + y + g)^5$ after all the anagrams of $rrryg$ are collected.

It's interesting that the cases of "3 red and 2 yellow peppers" and of "2 red, 2 yellow, and 1 green" have the same probability

$$\frac{5!}{2!3!}\left(\frac{1}{2}\right)^3\left(\frac{1}{3}\right)^2=\frac{5!}{2!2!1!}\left(\frac{1}{2}\right)^2\left(\frac{1}{3}\right)^2\frac{1}{6}$$

$$=\frac{5}{36}.$$

(These are the most probable combinations of five peppers given the initial probabilities we've chosen.)

Problem

18. Each edge of a cube measures 6 cm. The three edges passing through vertex A are divided into segments of 3 cm, 2 cm, and 1 cm, starting from point A. Then the cube is cut into 27 pieces along the planes parallel to its faces and passing through the points of division (fig. 4). (a) How many of the pieces will be "boxes" measuring 3 cm × 2 cm × 1 cm? (b) What are the other pieces, and how many pieces of each sort are there? What are their volumes?

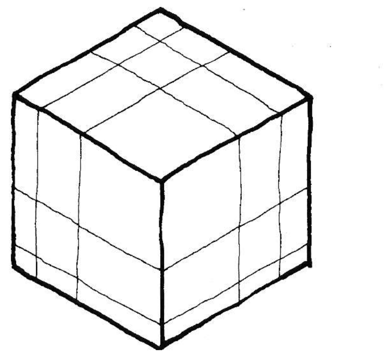


Figure 4

Binomial coefficients

In the foregoing we barely touched on the simplest and, perhaps, most important case of two letters. The coefficients in the expansion of $(a + b)^n$ have a special notation—in fact, they even have several standard notations, maybe because they're so important: $C(n, r)$, ${}_nC_r$, $\binom{n}{r}$ (and others). Here we'll use the first of these:[1]

$$(a + b)^n = C(n, 0)a^n + C(n, 1)a^{n-1}b + \ldots + C(n, r)a^{n-r}b^r + \ldots + C(n, n)b^n.$$

This formula is the famous *binomial theorem*. The coefficients $C(n, r)$ are called *binomial coefficients* (by the way, in the general case—with an arbitrary number of letters—the coefficients are often called *multinomial*). We know that $C(n, r)$—the number of n-letter words consisting of r letters b and $n - r$ letters a—is given by the formula

$$C(n,r)=\frac{n!}{(n-r)!r!}$$

$$=\frac{n(n-1)\ldots(n-r+1)}{r!}.$$

For instance,

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4,$$

because

$$C(4,0)=C(4,4)=\frac{4!}{4!0!}=1,$$

$$C(4,1)=C(4,3)=\frac{4!}{3!1!}=4,$$

$$C(4,2)=\frac{4!}{2!2!}=6.$$

Binomial coefficients are encountered in the most diverse problems of combinatorics, algebra, geometry, calculus, and probability. There are numerous connections between them, expressed by beautiful identities. For instance, from $C(n, r) + C(n, r - 1) = C(n + 1, r)$ (see problem 5), it follows that for all natural n and r, $0 \le r < n$, $C(r, r) + C(r + 1, r) + \ldots + C(n, r) = C(n + 1, r + 1)$ (for $r = 1$, it's just the formula that opens this article). And, of course, these numbers will crop up more than once in the pages of *Quantum*.

ANSWERS, HINTS & SOLUTIONS ON PAGE 62

[1]A while back our readers had an opportunity to become acquainted with these coefficients in "Summertime, and the Choosin' Ain't Easy" by Kurt Kreith (*Quantum*, July/August 1992). From that article you can learn how to find the number of *combinations* with repetitions.—*Ed.*

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M76

Applying classic inequalities. Prove that if the sum of two positive numbers is less than their product, then the sum is greater than 4. (N. Vasilyev)

M77

Curious rook. A rook visited all n^2 squares of an $n \times n$ chessboard. Prove that in so doing it had to change direction at least $2n - 2$ times. (Remember, the rook moves parallel to the sides of the chessboard.) (Y. Khodzinsky, 10th grader)

M78

Summing ordered distances. The numbers $1, 2, \ldots, 2n - 1, 2n$ are arbitrarily split into two groups of n numbers each. Let $a_1 < a_2 < \ldots < a_n$ be the numbers of the first group in ascending order and $b_1 > b_2 > \ldots > b_n$ the numbers of the second group in descending order. Prove that the sum of the distances between the corresponding numbers of the two groups, $|a_1 - b_1| + |a_2 - b_2| + \ldots + |a_n - b_n|$, equals n^2. (V. Proizvolov)

M79

Equal sides, equal circles. A line through the vertex B of an isosceles triangle ABC ($AB = BC$) cuts its base AC at D so that the radius of the incircle of triangle ABD equals that of the excircle of triangle CBD externally touching side DC (and the extensions of BC and BD—see figure 1). Prove that this radius is 1/4 the height h of the triangle dropped from a base vertex. (I. Sharygin)

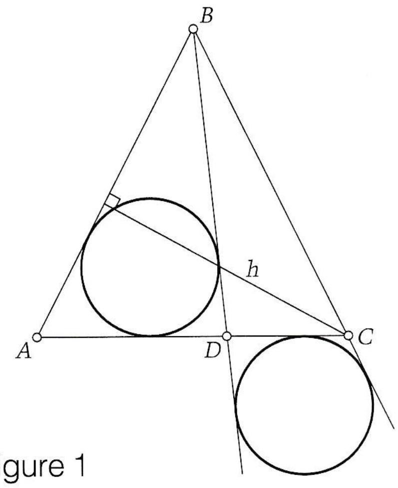


Figure 1

M80

Counting L-tiles. A square measuring 99×99 is tiled with figures of three types made of 3 or 4 unit squares, as shown in figure 2. (a) Prove that the number of figures of the first type ("*L*-tiles") is not less than 199. (b) Give an example of a tiling containing exactly 199 *L*-tiles. (D. Fomin)

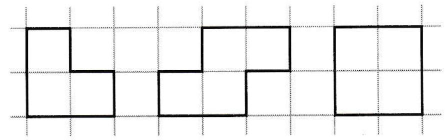


Figure 2

Physics

P76

Spring jacket. A spring of length L_0 has a great number of identical open windings. If the spring is hung by one of its ends, its length increases by a factor of 1.5. What will the length of the spring be if it is placed in a tall vessel with smooth walls? To what level must water be poured to cover the spring completely? The density of the spring is r times that of water. (S. Krotov)

P77

Puckish behavior. A puck of mass M slides along a smooth horizontal table with speed V_0. It hits another puck of mass $2M$. The first puck stops after a nonelastic collision, but the second one starts moving and hits still another puck of mass M (the same as the first one). What are the velocities of the pucks after colliding head-on? (Neglect friction.) (A. Vargin)

P78

Minimeltdown. A miniature oven for melting metals has a heating element

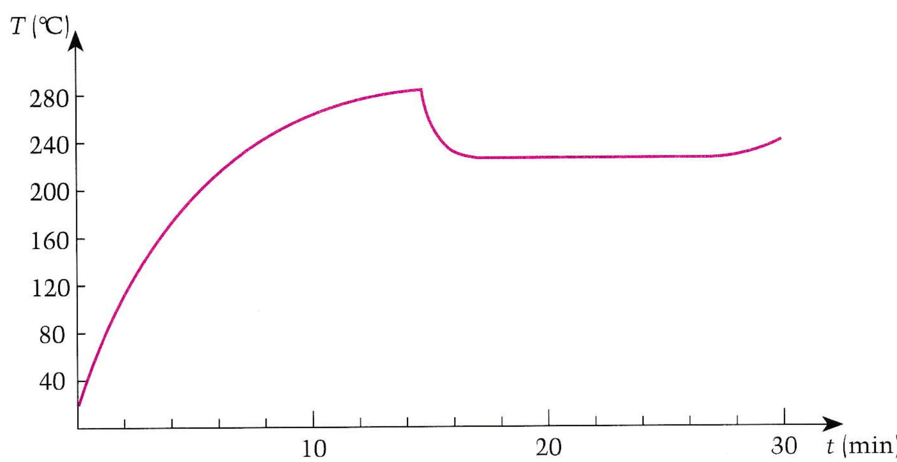


Figure 3

of constant power $P = 20$ W. The oven is switched on. After the temperature practically stops rising, some scraps of tin of mass $M = 50$ g are thrown into the oven and they begin to melt (the graph of temperature versus time is given in figure 3). Using these data, determine the latent heat of melting for tin. (A. Zilberman)

P79

Normal intensity. A system of stationary charges is symmetric relative to the axis OO_1. At a great distance from these charges, at point A on the axis, the electric field is $E_1 = 100$ V/m; at point B it is $E_2 = 99$ V/m. The distance between A and B is $L = 1$ m. Let's move from point A to a point $r = 1$ cm away from the axis. What is the perpendicular component of the electric field at this point? (A. Zilberman)

P80

Rite of passage. N converging lenses of focal length $2F$ and N diverging lenses of focal length $-F$ are placed alternately along an axis at a distance F apart (fig. 4). A beam of parallel light rays of diameter D enters the system along the axis. Determine the diameter of the exiting beam. (A. Yershov)

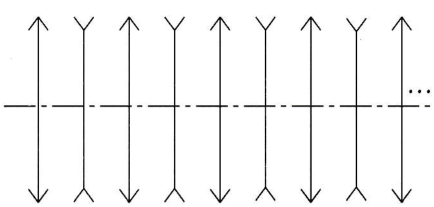


Figure 4

ANSWERS, HINTS & SOLUTIONS ON PAGE 56

Calling all modem maniacs!

What do you like in this or any issue of *Quantum*? If you find pen-and-paper communication too old-fashioned, you can send your comments, questions, and suggestions to the managing editor by electronic mail at the following address:

72030.3162@compuserve.com

We look forward to hearing from you.

The inevitability of black holes

In fact, why aren't there more of them?

by William A. Hiscock

Art by Sergey Ivanov

BLACK HOLES ARE TODAY an essential part of modern theoretical physics and astrophysics. Yet there are still some scientists (and many nonscientists) who find the notion of a black hole physically unacceptable. They believe the very idea is too outlandish to be real. In this article, we'll see that some of the most basic properties of a black hole can be understood in terms of Newton's law of gravity and that the formation of black holes need not involve ultrahigh-density matter or other areas of physics about which we currently know little. Black holes are a natural consequence of the nature of gravity. In fact, the odd thing is that there is anything in the universe *except* black holes!

Black holes are used in astrophysics to explain a number of different types of highly energetic astrophysical objects. Many galaxies seem to have extremely luminous and active nuclei. Depending on their appearance, such galaxies are classified as quasars, Seyfert galaxies, or BL Lac objects. It's widely believed that the source of energy powering these active galactic nuclei is a supermassive black hole, with a mass between 10^6 and 10^9 solar masses. Within our own galaxy, binary star systems that are bright X-ray sources are believed to contain either a neutron star or a black hole.

There exist well-defined (if not yet

precisely known) upper limits for the mass of any neutron star, certainly less than about 3.5 solar masses. The orbits of several of these binary systems are well enough determined to then rule out neutron stars. The best known is Cyg X-1, so called because it was the first X-ray source discovered in the constellation Cygnus.[1] For many years Cyg X-1 was the most promising candidate for a system containing a black hole. The best estimate of the mass of the invisible, X-ray-emitting object in the Cyg X-1 system is 16 solar masses, much greater than the maximum mass of a neutron star.

The actual physical notion of a black hole is a simple one that Newton would have understood in his time. A black hole is simply a region of space where gravity is so intense that nothing, not even a photon traveling at the speed of light, can escape the region. While Einstein's theory of gravity, "general relativity"—a name that hides the fact that it is a theory of gravity—is needed to correctly describe the gravitational physics of such an object, several key properties of black holes can be understood in terms of Newtonian gravity.

Newtonian black holes: Michell

The first person to consider the possible existence of an astronomical object from which light could not escape was the Reverend John Michell, a British amateur astronomer. In a letter to Henry Cavendish in 1783, Michell described a calculation using Newton's theory of gravity that showed that a spherical object 500 times the radius of the Sun, but with the same density, would have an escape velocity exceeding the speed of light.

Consider a spherical mass of radius R and mass M. It could be a planet, star, or soccer ball, but we'll call it a "star" for convenience. The gravitational potential energy of a particle of mass m on its surface is V, where V is given by

$$V = -\frac{GMm}{R},$$

and $G = 6.67 \times 10^{-11}$ N m^2/kg^2 is the gravitational constant. The particle will be able to reach infinity if it is projected from the surface with a velocity greater than or equal to the escape velocity from the surface v_e. The escape velocity is defined as the velocity that will allow the particle to just reach infinity with no residual velocity or kinetic energy. Thus, at infinity the total energy of the object should be zero. If the particle has velocity v_e at the surface of the star, then its total energy at the surface will be its kinetic energy $\frac{1}{2}mv_e^2$ plus the potential energy defined above. Since energy is conserved, this total energy must also be zero:

$$\tfrac{1}{2}mv_e^2 - \frac{GMm}{R} = 0.$$

This expression can be used to define the escape velocity from a star with mass M and radius R by solving for v_e:

$$v_e = \sqrt{\frac{2GM}{R}}.$$

Note that the mass m of the particle drops out, as it must, by Galileo's principle of equivalence (all bodies fall with the same acceleration in a gravitational field). So far, this is simply standard Newtonian gravitational physics, straight out of any textbook. In order to learn something about a "Newtonian black hole," we must incorporate one other idea: the notion that there is a maximum universal velocity at which particles can travel—namely, the speed of light c. Imagine holding the mass M constant in the equation for v_e above, while letting R decrease. As the radius of the star decreases, the escape velocity increases. There will be some special value of R at which the escape velocity will be equal to the speed of light. Any star of smaller radius would have an escape velocity greater than the speed of light, and nothing (light, spaceships, ...) could escape the object.

Relativistic black holes: Schwarzschild

The special value of R where the escape velocity is equal to c is called the Schwarzschild radius, named after Karl Schwarzschild, who discovered the solution that describes the simplest black hole in Einstein's theory of gravity. If we set v_e equal to c and solve for the Schwarzschild radius R_S, we obtain

$$R_S = \frac{2GM}{c^2},$$

a relation that is true in both Newtonian gravity and Einstein's theory. If we substitute in the values of G and c in order to find out just how small a black hole might be for a given mass, some remarkable numbers result. In SI units, we find

$$R_S\,[\text{meters}] = (1.48 \times 10^{-27})M\,[\text{kg}].$$

So, for example, the Schwarzschild radius of the Sun (mass 1.99×10^{30} kg) will be about 3 km. For the Earth (mass 5.98×10^{24} kg), the Schwarzschild radius is only about a centimeter. This doesn't mean that there is a black hole with these dimensions at the center of the Earth or the Sun. This is the radius to which we'd have to compress the entire Earth or Sun to cause it to become a black hole.

Problem 1. Calculate your own Schwarzschild radius. Is it larger or smaller than the size of an atomic nucleus? Calculate the Schwarzschild radius for the Milky Way Galaxy (mass 10^{11} times the mass of the Sun).

One might expect the *density* of an object the size of the Schwarzschild radius to be quite high (just imagine compressing the Earth until all its mass is confined to a sphere the size of a golf ball). If we assume that the object has a constant density ρ throughout, then $M = \frac{4}{3}\pi\rho R^3$, which again, thanks to spherical symmetry, holds exactly in Einstein's theory (despite the curvature of space–time) as well as in Newton's.

We can set $R = R_S$, solve for R_S, and then find the density ρ_S as a function of the mass M for an object at its Schwarzschild radius:

$$\rho_S = \frac{3c^6}{32\pi G^3 M^2}.$$

[1]You may recall Cyg X-1 from "The View through a Bamboo Screen" in the January/February 1992 issue of *Quantum*.—*Ed.*

Substituting in numbers, we again find some remarkable results: densities that are certainly far greater than any for which we have direct laboratory experience. If our Sun were to undergo gravitational collapse to form a black hole (which it would not do, but that's another story), its density as it passed the Schwarzschild radius is *inversely* proportional to the square of the mass—that is, smaller masses yield higher densities.

For comparison, recall that the density of water is 1 g/cm^3 and that the nucleus of the atom, which is the densest material studied in the laboratory, has a density of around 10^{14} g/cm^3. This is 1/100 the density the Sun would have at the Schwarzschild radius.

Problem 2. Calculate the density of a collapsing galactic-mass object as it crosses its Schwarzschild radius.

Principle of equivalence: Galileo to Einstein

The very large values for the density of matter at the Schwarzschild radius is one of the reasons some scientists refuse to take black holes seriously. How could anyone pretend that they know anything about the properties of matter at such high densities?[2] One could imagine all sorts of new physical laws coming into play at such densities that might prevent the formation of something as absurd as a black hole. After all, the physics of ice is rather different from the physics of steam. Might it not be reasonable to expect that at some high density, before the black hole forms, the matter creates a very large internal pressure, stopping the collapse and preventing the formation of a black hole? Many scientists who are unfamiliar with Einstein's theory of general relativity have proposed such ideas as a way to "escape" the idea that nature may contain such odd objects as black holes.

However, in the theory of general relativity, a very large pressure (such as it would take to halt the collapse at these extremely high densities) can actually *strengthen* the collapse rather than impede it. In order to understand this apparently paradoxical result, we must understand the Einstein equivalence principle, one of the key ideas on which he built his theory of general relativity.

The principle of equivalence has always been a cornerstone of our understanding of gravity, from Galileo through Einstein. The principle was first articulated by Galileo, who recognized that all types of matter fall with the same acceleration in a gravitational field. While Aristotle had proposed that heavier objects fall faster than light objects, Galileo's important insight led him to consider separately the effects of gravity and air resistance. A popular story (but probably untrue) has it that Galileo dropped cannonballs from the leaning tower of Pisa to show that the acceleration does not depend on the size or composition of the objects.

Testing the principle of equivalence: Braginsky and Panov

Today, the equality of the gravitational acceleration of different types of matter is one of the most precisely known quantities in physics. Experiments conducted by V. B. Braginsky and V. I. Panov in Moscow in 1971 showed that platinum and aluminum fall toward the Sun with the same acceleration to better than one part in 10^{12}. This means that if we wrote out the numerical values of the accelerations of platinum and aluminum, they would be the same number for at least the first 12 digits. Few properties of matter are known to such precision.

Einstein used Galileo's principle of equivalence—that all forms of matter respond to gravity (and create gravity) in the same way—and combined it with an insight gained from special relativity: energy and matter are equivalent ($E = mc^2$). Einstein's equivalence principle states that all forms of energy (including all forms of matter) respond to gravity and create gravitational fields in the same manner. Only the total amount of mass is important, not whether it is platinum or aluminum, "rest mass" energy, thermal energy, or even gravitational potential energy. The notion that the energy in the gravitational field itself *acts as a source for the gravitational field* was one of the most profound new ideas in general relativity.

For instance, if we have two cannonballs of the same initial mass and heat one up to a high temperature, then the hot cannonball will attract a test body more strongly than the cold cannonball, since it now has more total energy (and hence more total mass). The heat energy in the hot cannonball is equivalent to a certain additional amount of mass ($m = E/c^2$), and by the Einstein equivalence principle, all forms of matter and energy participate equally in gravity. Thus, the larger mass of the hot cannonball produces a stronger gravitational field. The difference in this case is unmeasurably small with current technology, but the principle applies to all bodies regardless of size and composition.

The relevance of all this to the formation of black holes is that the usual method proposed to prevent black hole formation is to hypothesize that matter at some high density develops an exceedingly large pressure, which then supports the star against collapse to form a black hole. It turns out, however, that in Einstein's theory of gravity, the large pressure that would be required to stop the collapse actually hastens the collapse! The internal pressure in the collapsing body represents a form of energy. By $E = mc^2$, there is a mass associated with that energy, and thus the pressure strengthens the gravity of the collapsing body, causing it to collapse even more rapidly than it would otherwise!

In the everyday world, pressure does act to support objects, such as you, me, the Earth, and the Sun, against gravitational collapse for two reasons. First, all of these objects are "nonrelativistic"—that is, they are much larger than their Schwarzschild radii, so that the gravitational force trying to collapse the body is not too large. Second, in all of these objects the

[2] See the excerpt from Galileo's "The Assayer" in the November/December 1992 issue.—*Ed.*

pressures are small compared to the mass density of the objects—that is, the ratio $p/\rho c^2$ is much less than one.

Problem 3. Calculate the ratio $p/\rho c^2$ for water at room temperature and atmospheric pressure. Is the number close to unity?

Einstein's equivalence principle, supported by the high-precision experiments of the Moscow group and others, makes it very unlikely that a large pressure (or any other form of internal energy) could prevent the formation of black holes. Still, we don't presently know all the details of the physics of matter at densities 10–100 times the density of the atomic nucleus. For this reason, the apparent discovery of black holes with extremely large masses in the centers of many galaxies is exciting because it provides unassailable evidence for black holes. For any collapsing object with a mass larger than about 10^8 solar masses, the density of the object as it crosses the Schwarzschild radius is less than the density of water—1 g/cm^3.

While there may be mysteries associated with the behavior of matter more dense than the atomic nucleus, we do have a good understanding of matter at ordinary densities such as 1 g/cm^3, and we know there is nothing in the physics of such matter that could halt the collapse and prevent the formation of a black hole. Even if we someday find that Einstein's theory is wrong for sufficiently strong gravitational fields (just as we know Newton's theory is wrong for strong fields), the low densities of matter needed to form black holes of very large mass makes their existence inevitable.

Suggestions for further reading

General Relativity from A to B, Robert Geroch (Chicago: University of Chicago Press, 1978).

Space, Time & Gravity: The Theory of the Big Bang & Black Holes, Robert M. Wald (Chicago: University of Chicago Press, 1992).

William A. Hiscock *is an associate professor of physics at Montana State University in Bozeman.*

ANSWERS ON PAGE 62

PHYSICS CONTEST

How about a date?

"Time is what prevents everything from happening at once."
—John Archibald Wheeler

by Arthur Eisenkraft and Larry D. Kirkpatrick

THE SEPTEMBER 1991 DISCOVERY of a frozen body in the Tirolean Alps has revived interest in radiocarbon dating (see "Physics Fights Frauds" in the January/February issue of *Quantum*). The body was that of a hunter (or possibly a shepherd). The discovery is particularly valuable to anthropologists because the body was virtually intact and fully clothed, and the hunter had been carrying various articles such as a bow and arrows. The date of the hunter's death was determined by carbon dating to be around 3300 B.C.

Modern techniques of carbon dating using mass spectroscopy have greatly reduced the size of the samples required for dating. The date for the death of the hunter was determined with a tissue sample the size of a tablet of artificial sweetener. Carbon dating techniques rely on the presence of two carbon isotopes in the atmosphere. The vast majority have a nucleus contain 6 protons and 6 neutrons. These are designated by the symbol ^{12}C, where the superscript denotes the total number of protons and neutrons. However, trace amounts of ^{14}C (containing two extra neutrons) are produced by collisions of cosmic rays with nitrogen atoms in the upper atmosphere. Currently the ratio of ^{14}C to ^{12}C in the atmosphere is 1.3×10^{-12} to 1.

All plant and animal life interact with the atmosphere, and the ratio of the two carbon isotopes in their bodies reaches equilibrium with the atmosphere. At the time of death, the ratio of ^{14}C to ^{12}C in a sample of the plant or animal is therefore equal to that in the atmosphere. After this the number of stable ^{12}C atoms in the sample remains constant. However, the radioactive decay of the ^{14}C causes their number to decrease exponentially according to the well-known decay law

$$N = N_0 e^{-\lambda t},$$

where N_0 is the original number of atoms in the sample, N is the number of atoms remaining after a time t, and the decay constant λ is a parameter that depends on how rapidly the atoms decay.

A useful measure of the decay rate is the half-life $t_{1/2}$. This is the time required for one half of any sample of radioactive atoms to decay and is independent of the size or age of the sample. For ^{14}C the half-life is measured to be 5,730 years = 1.81×10^{11} s. The decay constant and the half-life are related by

$$\lambda = \ln 2/t_{1/2}.$$

Instead of looking at the number of atoms remaining in a sample, we can focus on the rate R at which they decay. This can be obtained by differentiating the decay law to obtain

$$R = R_0\, e^{-\lambda t},$$

where $R_0 = \lambda N_0$.

This brings us to our contest problem.

A. Assume that we have isolated a 1-g sample of carbon from a frozen animal and that the atmospheric ratio of the two carbon isotopes was the same when the animal died as it is now. What was the decay rate in decays per minute of the ^{14}C shortly after the animal died?

B. If the current decay rate is 1 decay per minute, how many years ago did the animal die?

Unfortunately, the ratio of the two isotopes of carbon has not been constant throughout time. It's possible to determine the dependence by dating samples from objects with well-determined ages. These could be such things as dated historical documents or tree rings. Let's look at two simplified scenarios to illustrate the problems associated with this variability.

C. How does the age of our sample change if the ratio varied linearly in the past? Assume that the ratio decreases by 1% of the current value for each century that we go back in time.

D. How does the age of our sample change if the atmospheric ratio has varied sinusoidally in the past? Assume a cosine dependence with an average value equal to the current value, an amplitude that is 5% of the current value, and a period of 628 years.

CONTINUED ON PAGE 41

Art by Tomas Bunk

25 MILL
10 MILL
5 MILL
2 MILL
1 MIO
500,000 YEARS
300,000 YEARS
100,000
20,000
10 000
1
3
5
7
8
9
SNOW FROM YESTER YEAR
TOMAS BUNK

Miss or hit

More brainteasers from Perelman

IN THIS INSTALLMENT OF THE Kaleidoscope we present more examples from the vast legacy of a master puzzler.[1] They are all taken from Yakov Perelman's books *For Young Mathematicians* (the first and second hundred puzzles).

If you wish, you can mentally substitute dollars and cents for rubles and kopeks without bothering to consult the current exchange rates—one ruble contains the same number of kopeks as a dollar contains cents.

1. *Cover price.* The price of a bound book is 2 rubles 50 kopeks. The book is 2 rubles more expensive than its binding. How much does the cover cost?

2. *Lost discount.* Ms. Ivanova buys all her books from a bookseller she knows and gets a discount of 20 percent. Beginning January 1, all book prices will be raised 20 percent. Ivanova figures now she'll be paying the same for books as every other customer did before January 1. Is she right?

Art by Pavel Chernusky

[1]See "Hit or Miss" in the November/December issue.

3. *Rare coin.* A well-known collector of antiquities was told that a coin was dug up in Rome with the inscription (in Latin): 53 B.C. "This coin is definitely a fake," the collector said without missing a beat. How did she know, without seeing the coin or even a picture of it?

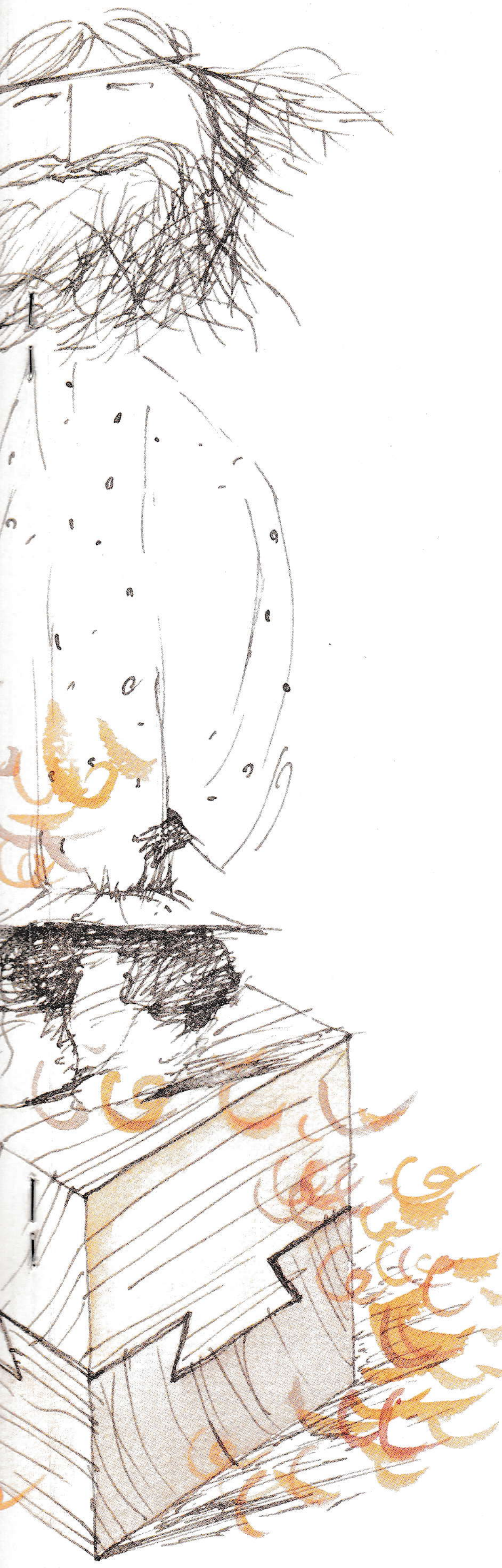
4. *Asparagus.* A customer used to buy asparagus from a greengrocer in large bundles, each 40 centimeters around. She would measure the bundles to be sure she wasn't being cheated. One day the greengrocer didn't have a 40-centimeter bundle, so he offered the customer two smaller bundles, each 20 centimeters around, for the same price. The customer measured both bundles and, convinced that each of them really was 20 centimeters around, paid the greengrocer the same amount she usually paid for the thick bundle of asparagus. Did she save money or lose money in this transaction?

5. *Carpenters and cabinetmaker.* Six carpenters and one cabinetmaker were hired do some work. Each carpenter earned 20 rubles, while the cabinetmaker earned 3 rubles more than the average wage of all seven. How much did the cabinetmaker earn?

6. *Addition and multiplication.* No doubt you've noticed the curious property of these two equations:

$$2 + 2 = 4,$$
$$2 \times 2 = 4.$$

It's the only example of the sum and product of two integers (by the way, two equal integers) being equal. But did you know there are fractions (not equal, though) that possess the same property? For instance,



$$3 + 1\tfrac{1}{2} = 4\tfrac{1}{2},$$
$$3 \times 1\tfrac{1}{2} = 4\tfrac{1}{2}.$$

Try to find other such numbers. Don't worry—your search won't be in vain. There are lots of them.

7. *How was it done?* The picture at left shows a wooden cube made of two pieces of wood: the upper half has protuberances that slide tightly into the grooves of the lower half. Look closely at the shape and placement of the protuberances and explain how the carpenter managed to join both pieces of wood.

8. *The game of 32.* This is a game for two persons. Put 32 matches on the table. To begin play, the first player takes 1, 2, 3, or 4 matches. Then the second player takes as many matches as she wants but no more than 4. And so on. The player who takes the last match wins.

As you can see, the game is very simple, but it has an interesting feature: the first player can always win by correctly calculating how many matches to take. What is the optimal strategy for the first player?

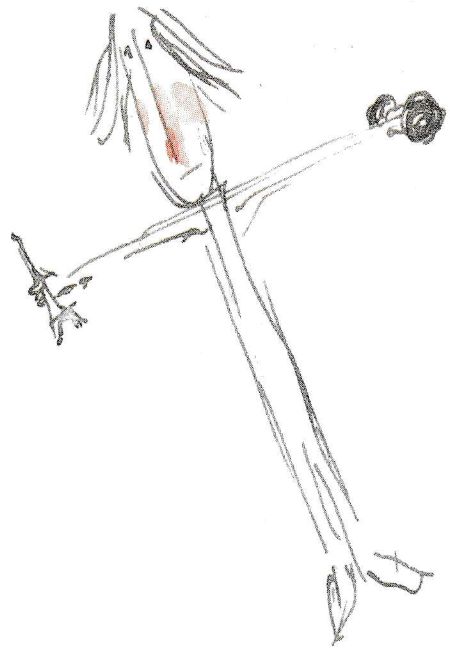
9. *The game of 32 reversed.* The game of 32 can be reversed: the player who takes the last match loses. What is the optimal strategy for the first player this time?

10. *The game of 27.* In this variant of the game of 32, each player alternately takes no more than 4 matches from the table. The player who has an even number of matches at the end of the game is the winner. Again, the first player can always win if the right strategy is chosen. What is it?

11. *The game of 27 reversed.* The player with an odd number of matches at the end wins in this version of the game of 27. What is the optimal strategy in this case?

12. *Mass of a bottle.* A bottle filled with gasoline has a mass of 1,000 grams. The same bottle filled with acid has a mass of 1,600 grams. The acid is twice as dense as gasoline. What is the mass of the bottle?

13. *Cherry.* The fleshy portion of a cherry surrounds the cherry pit with a layer of the same thickness as the pit



itself. We'll assume that both the cherry and the pit are spherical. Can you mentally estimate the ratio of the volume of the flesh to the volume of the pit?

14. *Model of the Eiffel tower.* The Eiffel tower in Paris is 300 meters tall and is made completely of iron (8,000,000 kilograms of it). A friend of mine has an exact copy of the famous tower. It has a mass of only 1 kilogram. Is it taller or shorter than an ordinary drinking glass?

15. *Sailboat race.* Two sailboats were involved in a race. They had to sail 24 miles in one direction and 24 miles back. The first boat sailed the entire course at a uniform speed of 20 mph; the second boat sailed the first half at 16 mph and the second half at 24 mph. The first boat won, although it would seem that the second boat would lag in the first half of the course by the same amount it gained in the second, so that the two boats would arrive at the finish line simultaneously. Why was the second boat late?

16. *River vs. lake rowing.* Rowing downstream, a rower covers 5 miles in 10 minutes. The return trip takes an hour. So the rower covers 10 miles in 1 hour 10 minutes. How long would it take the same person to row 10 miles in still water (say, a lake)?

CONTINUED ON PAGE 40

In the curved space of relativistic velocities

Where two pathfinding concepts intersect

by Vladimir Dubrovsky

WHEN THREE MATHEMATICAL geniuses—Bólyai, Gauss, and Lobachevsky—shook the world of classical geometry to its foundations with their theory of non-Euclidean geometry, the time was apparently ripe for this discovery. They made it independently and almost simultaneously. (The dramatic story of this great discovery was related by A. D. Alexandrov in the November/December 1992 issue of *Quantum*.) But even if the theory weren't created then, back in the first quarter of the 19th century, it would surely have emerged in the course of modern research into elementary particles, where relativistic effects become significant. The fates decreed that after a lapse of almost a hundred years, Lobachevsky's "imaginary geometry"—a mental construction that appeared in a purely speculative way out of efforts to prove Euclid's parallel postulate—has found a remarkable physical embodiment.

For Lobachevsky, the two versions of geometry—Euclidean and non-Euclidean[1]—were equally acceptable as the true geometry of the space around us, and it was a matter of experimentation to decide which one of the two describes our space more exactly. These ideas of the "Copernicus of geometry" were of paramount importance for the future development of science because they shattered once and for all the ingrained view that Euclidean geometry is the only possible geometry, that it is inherent in our space or in the way we apprehend it. And modern science accepts the possibility that space is indeed non-Euclidean.

But the "physical embodiment" I'm going to describe is not the space we live in. Nor is it an artificial "magic world," like the one invented by Poincaré (see Simon Gindikin's article in the November/December issue). This space is nevertheless very real. It shows up in virtually every problem related to collisions of elementary particles, and this sort of physical problem is perhaps the kind that is solved most often. (I'm not exaggerating—every day in dozens of laboratories throughout the world, thousands of experiments on particle scattering are conducted.) On the other hand, this is a very abstract space: you can neither feel nor see it. Even to imagine it is rather difficult. The points that constitute this space are . . . well, for the time being, let's say they're *velocities*—all possible velocities of moving objects. And this is why this space is called *velocity space*. Its non-Euclidean nature reveals itself only when the speeds we consider are so high that relativistic effects must be taken into account. But to get used to this space, and to understand what it might be good for, let's begin with the "Newtonian," nonrelativistic case.

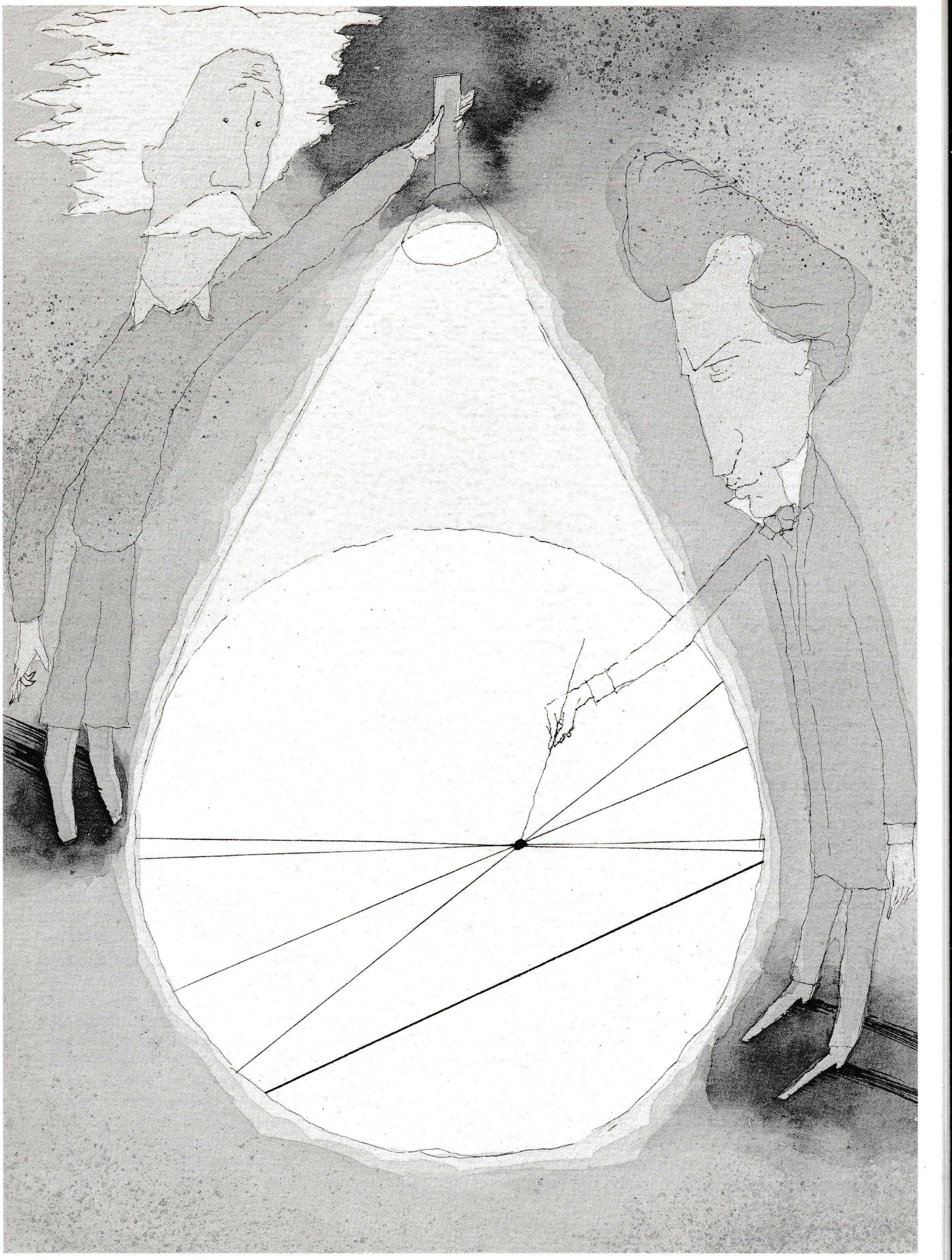
A close encounter of the stellar kind, or nonrelativistic kinematic graphs

I'll start with example 10-6 from *Physics* by Richard Wolfson and Jay M. Pasachoff (Little, Brown and Company). It's a fairly standard, comparatively easy problem but its routine conservation-law solution in the textbook takes almost a whole page of algebraic rigmarole. Here's the problem (in slightly modified form):

A star B of mass M is a great distance from another star A of equal mass and is approaching the second star at 680 km/s. The two stars undergo a close encounter, and much later star B is moving at a 35° angle to its initial direction (fig. 1). In the frame of reference in which star A is initially at rest, find the final speeds of both stars and the direction of motion of star A.

[1]As in previous *Quantum* articles on this subject, the term "non-Euclidean" geometry is understood here in the narrow sense of one particular kind of geometry other than the Euclidean: hyperbolic (or Lobachevskian) geometry.

Art by Pavel Chernusky



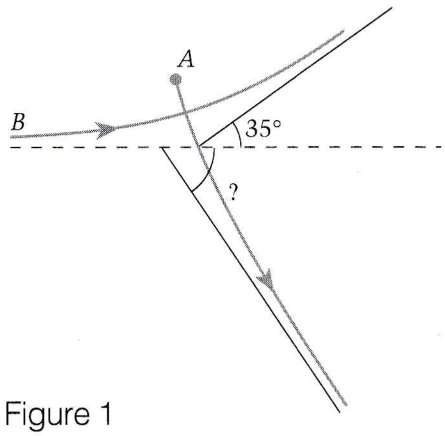


Figure 1

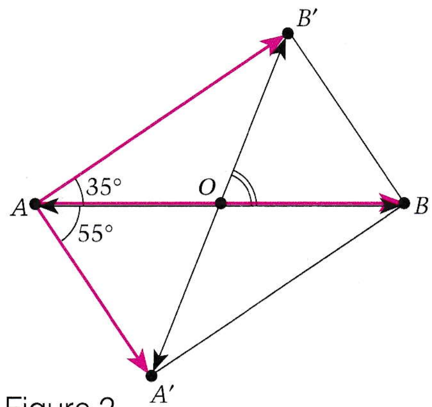


Figure 2

I'll solve this problem using a "kinematic graph" of the interaction. Mark an arbitrary point on the plane—this will represent the center of mass O of both stars (in this case it's simply the point halfway between the stars' centers), or rather, its velocity. Draw the velocities $\mathbf{v}_A$ and $\mathbf{v}_B$ of the stars with respect to center O as vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ on the plane (fig. 2). These are two opposite vectors ($\overrightarrow{OA} = -\overrightarrow{OB}$), because the net momentum of any system with respect to its center of mass is zero: $M\mathbf{v}_A + M\mathbf{v}_B = 0$; of course, this is true at all times, before, during, and after the interaction. The center of mass moves uniformly (this is a rewording of the conservation law for the net momentum), so it's always represented by the same point in the plane. The stars change their velocities during the encounter, but they don't change their speeds v_A and v_B with respect to O, because these speeds remain equal and the combined kinetic energy $Mv_A^2/2 + Mv_B^2/2$ in the reference frame O is preserved (the interaction can be considered elastic). So vectors $\overrightarrow{OA'}$ and $\overrightarrow{OB'}$, equal to the velocities of the stars a long time after the encounter, can be obtained simply by rotating $\overrightarrow{OA}$ and $\overrightarrow{OB}$ about point O through some angle. All four vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$, $\overrightarrow{OA'}$, and $\overrightarrow{OB'}$ form what's called the *kinematic graph* of the interaction. We see that the tips of the vectors are the vertices of a rectangle, and all four vectors issue from its center. But the problem deals with velocities in the reference frame A in which the second star was initially at rest. No problem! The velocity of O in this reference frame is, obviously, $-\mathbf{v}_A$, so by the velocity addition rule, the velocity of any object X with respect to A is the vector sum of its velocity with respect to O and $-\mathbf{v}_A$. In the graph, this is the sum of vectors $\overrightarrow{AO}$ and $\overrightarrow{OX}$—that is, $\overrightarrow{AO} + \overrightarrow{OX} = \overrightarrow{AX}$. So to obtain the kinematic graph in the reference frame A, we must simply draw vectors from point A to all the other points (the red vectors in figure 2).

Now, by the statement of the problem, $\angle BAB' = 35°$ and $AB = 680$ (with the proper unit length). Thus,

1. The final speed of star $B = AB' = AB \cos 35° = 680 \cos 35° \cong 557$ km/s;
2. The "angle of recoil" of star $A = \angle BAA' = 90° - 35° = 55°$;
3. The final speed of star $A = AA' = AB \cos 55° = 680 \cos 55° \cong 390$ km/s.

What we've done here works as well for any elastic collision of two bodies or particles of equal mass. In particular, if it's not a head-on collision, then in the reference frame where one of the colliding objects was initially at rest, they always bounce apart at right angles.

Further inspection of our diagram provides us with more information. For instance, write the Pythagorean theorem for the right triangle $AA'B'$: $A'B'^2 = A'A^2 + B'A^2$. The lengths of the legs AA' and AB' are the final speeds u_A and u_B of the respective bodies in A's reference frame; the length of the hypotenuse $A'B' = AB$ is the initial speed w_B of B with respect to A. Substituting, we get $w_B^2 = u_A^2 + u_B^2$; multiplying by $M/2$, we get

$$\frac{Mw_B^2}{2} = \frac{Mu_A^2}{2} + \frac{Mu_B^2}{2}$$

—the law of conservation of energy in A's reference frame (w_A—the initial speed of body A with respect to itself—is zero). Another observation is that the "scattering angle" $\angle BOB'$ (between the initial and final velocities of the colliding body B) in the center-of-mass reference frame is twice the scattering angle $\angle BAB'$ in the "laboratory" reference frame A (prove this simple geometric fact yourself).

These examples illustrate the close connection between ordinary (nonrelativistic) kinematics and ordinary (Euclidean) geometry. Kinematic graphs form the bridge between the two.

Now we'll look at the above construction from a more general point of view. To avoid misunderstandings in what follows, let's agree that all the moving objects move uniformly and parallel to one and the same plane. Confining ourselves to the two-dimensional case makes our arguments simpler and our diagrams clearer. The three-dimensional case can be treated by analogy.

Kinematics and geometry: the plane of velocities

So how do we render kinematic notions and facts in geometric language? We choose an arbitrary point O in the plane to represent some reference frame (it was the "center of mass" frame of reference in the problem above) and link every moving object with some point X so that the vector $\overrightarrow{OX}$ equals (in certain units of measurement) the vector of this object's velocity with respect to the reference frame. Thanks to the *classical* velocity addition rule, we can now find the velocity of any object with respect to any reference frame by drawing the vector from the point representing the frame to the point representing the object. We did it for the reference frame A in figure 2, but we could just as well do it for any

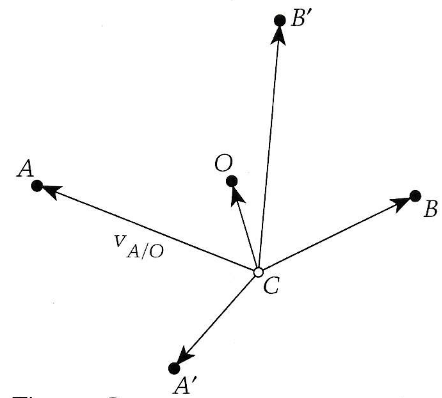


Figure 3

other reference frame C, as shown in figure 3. For instance, vector $\overrightarrow{CA'}$ in this figure is equal to the initial velocity of star A in the reference frame C—we'll denote it $\mathbf{v}_{A/C}$. To find the speed $v_{A/C}$ of A with respect to C, we can simply measure the distance CA on the plane, but we can also calculate it, given sufficient data and using well-known theorems from plane geometry—the law of cosines and so on.

All geometric notions can be expressed in terms of vectors, their lengths, and the angles between them. So, if we think of a vector $\overrightarrow{XY}$ as the vector $\mathbf{v}_{X/Y}$ of the relative velocity of corresponding uniformly moving objects, every geometric notion receives a certain kinematic interpretation.

Now I want to use these interpretations as the *definitions* of the notions of geometry in terms of kinematics.

Imagine some set V whose points are in a one-to-one correspondence with all possible velocities. Since velocities don't exist by themselves (they always exist "with respect to something"), it might be better to say that each of these points represents a whole "herd" of all the objects moving with one and the same velocity—that is, stationary with respect to one another.[2] You can visualize V as the Euclidean "plane of kinematic graphs," but it isn't necessary, and it will even turn out to be wrong in the relativistic case. I'll attach the tag "v-" to everything connected with the set V, so what I've just described is v-points. So what is the v-distance between two v-points A and B? Why, it's the relative speed $v_{B/A} = v_{A/B}$ of an object represented by one of the v-points with respect to any object corresponding to the other v-point. And v-angle BAC? It equals the angle between the corresponding velocity vectors $\mathbf{v}_{B/A}$ and $\mathbf{v}_{C/A}$.

In particular, when this angle measures 0° or 180°, v-points A, B, and C are v-collinear—that is, they belong to the same straight v-line. Put in a different but equivalent way, which will prove useful later, the v-collinearity of three different v-points means that *if some objects B and C corresponding to two of these v-points meet an object A corresponding to the third v-point as they move, then they necessarily meet each other*. Indeed, from the point of view of the observer moving with object A, this object is at rest and objects B and C visit its location, while the directions in which they move are the same or opposite—the angle between $\mathbf{v}_{B/A}$ and $\mathbf{v}_{C/A}$ is 0° or 180°. So for this observer, B and C move along one and the same straight path and therefore have to meet in the past or in the future, since their velocities are different (because B and C are distinct points).

The set V supplied with these definitions of basic geometric notions, which enable one to give a kinematic rendering of any statement of geometry, is called the *nonrelativistic velocity space*.

As for the "plane of kinematic graphs," it can be regarded as the exact map of the two-dimensional velocity space. We don't really need it to *define* this space, but it makes it evident that its geometry is Euclidean.

Introducing relativity

Now let's try to carry over the geometric method in kinematics to the relativistic case. I think many *Quantum* readers have heard something about special relativity, although I won't assume you have. I just need the two basic assumptions that form Einstein's relativity principle:

1. All the laws of nature are the same in all uniformly moving frames of reference.
2. A light ray moves with the same speed c regardless of whether it's emitted by a moving body or a body at rest, and this speed is the absolute speed limit for all matter.

Usually that final clause, about the speed limit, isn't included in the statement of the theory of relativity: it can be derived from the constancy of the speed of light. I could do without it, too. But it makes the subsequent discussion shorter and simpler, and after all, it's true—so why not take advantage of it? As to the first part of the principle, it's not especially relativistic and is obeyed in Newtonian mechanics, too. But classical mechanics implicitly assumes that interactions can propagate instantaneously—which is all right, because it deals with speeds so low compared to c that c can be considered infinite and relativistic effects negligible.

When we apply our construction of velocity space to relativity, everything works smoothly except the definition of v-distance as relative speed. Of course, we want v-distances to satisfy the usual addition rule: $AB = AC + CB$ whenever C lies on line AB between A and B. Then the relation $XY = v_{Y/X}$ would imply the ordinary velocity addition rule: $v_{B/A} = v_{C/A} + v_{B/C}$—at least in the one-dimensional case, where objects A and B move in opposite directions with respect to C. But this doesn't obey the speed limit! If the speeds of A and B with respect to C are, say, $0.6c$, then this rule would yield $\mathbf{v}_{B/A} = 1.2c > c$, which is impossible. So we must slightly generalize the definition and assume that the relativistic v-distance between A and B—which will be denoted by $r(A, B)$—is a certain function R of the relative speed of the moving objects A and B:

$$r(A, B) = R(\mathbf{v}_{B/A}).$$

In addition, we'll assume that v-distance satisfies the addition rule

[2]A more formal and, in a sense, simple way to put it would be to declare every such "herd" a point of V. I'm afraid, though, that such strange "points" would have confused some of our readers.

$$r(A, B) = r(A, C) + r(C, B)$$

for any v-point C on a v-segment AB. The unknown function $R(v)$ is intended to correlate the addition of distances along a line and the relativistic addition of velocities.

These assumptions will suffice for us to study the geometric structure of the velocity space V and to find an explicit expression for $R(v)$. To this end, let's draw a map of the space V as we did before.

Velocity mapping and the Klein model of non-Euclidean geometry

Choose a point O to represent some uniformly moving observer (and the corresponding v-point) in the plane. By the second part of the relativity principle, whatever velocity vector measured by the observer is drawn from the point O, its tip will lie within the circle of radius c centered at O, so the entire map M_O of V constructed by the observer O will coincide with this circle. Of course, the map is not drawn to scale, because the equality of relative speeds $\mathbf{v}_{B/A} = \mathbf{v}_{D/C}$, which is equivalent to the equality of corresponding v-distances $r(A, B) = r(C, D)$, does not imply the equality of Euclidean distances between corresponding pairs of points A, B and C, D on the map.[3]

For instance, according to the map M_O, the Euclidean distance OA equals the relative speed $v_{A/O}$ for any point A. However, if a segment has a first endpoint on the circumference of the circle and a second inside the circle, then, by the second part of the relativity principle, it corresponds to the speed of light c, no matter where the second endpoint may be.

Actually, such a discrepancy between the map and what's mapped isn't so unusual. The only kind of map most of us have ever used—geographic maps—distort distances and shapes on the Earth's surface if the mapped territory is large enough. And there's nothing we can do about it, because a piece of a sphere simply can't be mapped to scale on a plane (see "In Search of a Definition of Surface Area" in the March/April 1991 issue of *Quantum*). Similarly, relativistic velocity space is impossible to map to scale on the plane because, like the sphere, it's also curved. Nevertheless, its map will tell us a lot about its geometry. And it reproduces correctly some of the features of space V. In particular, the images on the map of any three collinear v-points also lie on one line—a chord of the circle. This fact isn't at all trivial, because now we *cannot* add velocities as vectors. Its proof is based on the definition of three collinear v-points as corresponding to three uniformly moving objects, every two of which meet (see figure 4).

Thus, we arrive at a crucial conclusion: *the relativistic velocity space can be mapped on a circle so that v-lines are mapped into chords of this circle.*

At this point, readers familiar with non-Euclidean geometry (through earlier *Quantum* articles, perhaps) must be heaving a big sigh of relief. Why? Because the set of all interior points of a circle, with straight lines defined as chords, and other geometric notions defined so as to comply with this concept of line, is nothing other than the Klein model of non-Euclidean (hyperbolic) geometry. The geometry so defined obeys all the axioms of Euclidean plane geometry except the parallel postulate, which means that a number (an infinite number!) of lines not intersecting a given line can be drawn through a given point. This is illustrated in figure 5, which shows several lines through O "parallel" to line AB.

We're not finished yet! We know that the geometry of space V is not Euclidean, because the parallel postulate is violated. We still can't be sure, though, that all the *other* Euclidean postulates are valid in velocity space. Taking for granted that they're satisfied in the Klein model (we'll come back to this later), we need only prove

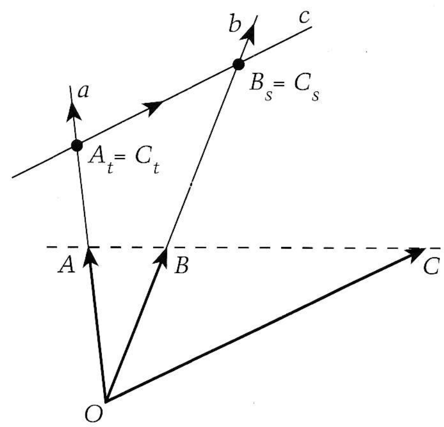


Figure 4

V-collinearity on a velocity map. Consider three objects A, B, and C moving with constant velocities $\mathbf{v}_A$, $\mathbf{v}_B$, and $\mathbf{v}_C$ relative to the observer O. In the figure, $\overrightarrow{OA}$, $\overrightarrow{OB}$, and $\overrightarrow{OC}$ are vectors drawn on the map M_O and equal to the respective velocities. Assume that objects A and B meet right at the location of the observer when his or her watch reads zero, and that A meets C at time t according to the observer's watch (the lines a, b, and c in the figure are the paths of the respective objects). If the v-points corresponding to the objects are collinear, C must also meet B at some moment s. Denoting by X_u the position of object X at time u, we have $A_t = C_t$, $B_s = C_s$, and so $\overrightarrow{C_tC_s} = \overrightarrow{A_tB_s} = \overrightarrow{OB_s} - \overrightarrow{OA_t}$. Substituting into the vector equation of uniform motion, we get $(s - t)\mathbf{v}_C = s\mathbf{v}_B - t\mathbf{v}_A$, or $s(\mathbf{v}_C - \mathbf{v}_B) = t(\mathbf{v}_C - \mathbf{v}_A)$. It follows that vectors $\overrightarrow{BC}$ and $\overrightarrow{AC}$ are proportional: $s\overrightarrow{BC} = t\overrightarrow{AC}$, so points A, B, C on the map lie on one line. The argument can be reversed to show that if points on the map are collinear, then the corresponding v-points are v-collinear.

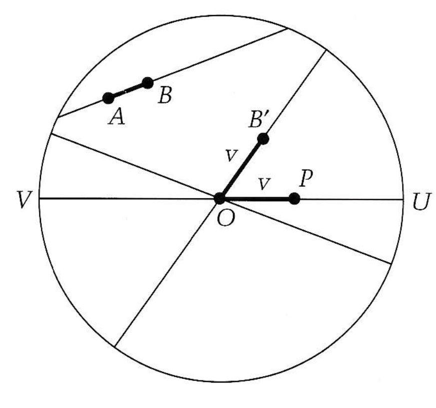


Figure 5

[3]Remember, we agreed to call a "v-point" any moving object represented by this v-point, and to denote its image on the map by the same letter, explicitly indicating exactly what is meant at each occurrence of the equation.

that *v-congruent segments in velocity space are depicted on the map as segments congruent in the sense of the Klein model*, because all geometric motions and their properties can be expressed in terms of points, lines, and congruence.

As we know, v-segments AB and CD have equal v-lengths if the corresponding relative speeds are equal: $v_{B/A} = v_{D/C}$. In the Klein model, two segments are considered congruent if one of them can be taken into the other by a transformation of the circle that preserves the collinearity of points—that is, takes chords into chords—because, by definition, all such transformations are isometries (distance-preserving transformations) in this model. Let $k(A, B)$ be the (non-Euclidean) distance between points A and B in the Klein model; for the time being, its explicit form doesn't matter for us. If I show that $k(A, B)$ is a function of relative speed $v_{B/A}$, it would mean that the equality $v_{B/A} = v_{C/D}$ implies $k(A, B) = k(C, D)$, and our goal will be attained.

Fix some point A of the map M_O and consider the transformation of M_O that sends an arbitrary point X in it onto the point X' such that $\overrightarrow{OX'} = \mathbf{v}_{X/A}$ (according to our agreement, X and A on the right denote moving objects represented by respective points on the map). Under this transformation point A goes into the center O of the map. Writing A' for O, we can express this as $\overrightarrow{A'X'} = \mathbf{v}_{X/A}$. Except for the primes, this is exactly the relation defining the map M_A drawn by an observer moving with object A! So our transformation turns the map M_O into M_A. Since every chord on map M_O represents some v-line, which in turn is represented by some chord on map $M_{A'}$, the transformation preserves lines. Therefore, from the point of view of the Klein model, it's an isometry: $k(X, Y) = k(X', Y')$ for any points X and Y. In particular, $k(A, B) = k(A', B') = k(O, B')$. Now it's not hard to see that $k(O, B')$ depends *only* on the Euclidean length v of OB' (fig. 5). Indeed, any other point P the same distance v from O can be obtained from B' by a rotation about O. Such a rotation certainly preserves lines and therefore preserves the non-Euclidean distance k. So $k(O, P) = k(O, B')$. Let $R(v) = k(O, P)$ (for any P such that $OP = v$). Then $k(A, B) = R(O, B')$. But, by the definition of point B', $OB' = v_{B/A}$; so, finally,

$$k(A, B) = R(v_{B/A}),$$

and we're done: we've conclusively established that, from the geometric point of view, *two-dimensional relativistic velocity space is the non-Euclidean (hyperbolic) plane.*

The formula for v-distance and relativistic velocity addition

We can see from the preceding that the non-Euclidean distance $k(A, B)$ between the images of two v-points on a velocity map can be taken as the v-distance between the v-points: $k(A, B)$ satisfies the addition rule and is a function R of the relative speed $v_{B/A}$. To find the explicit expression for the function $R(v)$, I'll use another model of non-Euclidean geometry—the one devised by Henri Poincaré, which was thoroughly described in the article "Inversion" (September/October 1992; see also "The Wonderland of Poincaria" in the subsequent issue). This model is also built from the interior points of some circle (let it be M_O again), but the lines are defined as arcs orthogonal to this circle, not chords, and isometries are defined as the mappings of the circle onto itself preserving such arcs (they're generated by reflections in these arcs—see "Inversion"). Figure 6 shows how the Klein model can be turned into the Poincaré model so that chords become arcs orthogonal to M_O. It also shows, together with figure 7, that the "cross ratio" of points A, B, A_0, B_0 (where A_0 and B_0 are the endpoints of the chord through A and B), defined as

$$\{AB, A_0B_0\} = \frac{AA_0}{A_0B} : \frac{AB_0}{B_0B} = \frac{A_0A \cdot BB_0}{A_0B \cdot AB_0},$$

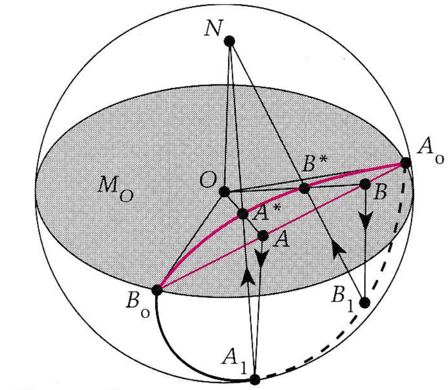


Figure 6

Transforming the Klein model into the Poincaré model. Consider the sphere built on the circle M_O as its equator. From any point A of the circle draw the perpendicular to the equatorial plane to meet the "southern hemisphere" at A_1. Join A_1 to the "North Pole" N to meet the circle M_O at A^*. We get the two-step transformation $A \to A_1 \to A^*$. The first step converts a chord A_0B_0 into the semicircle on diameter A_0B_0 lying in a vertical plane and thus orthogonal to circle M_O. The second step is in fact a "stereographic projection"—like the one considered in the article "Inversion" referred to in the text, except that the plane is passed not through the "South Pole" but through the sphere's center (which does not make a significant difference). By the properties of stereographic projection, it maps the semicircle onto the arc A_0B_0 orthogonal to M_O, which is considered a straight line in the Poincaré model. In addition, stereographic projection, as a particular case of inversion, preserves cross ratios—$\{A^*B^*, A_0B_0\} = \{A_1B_1, A_0B_0\}$; while the first step of our transformation "takes a square root" of the cross ratio of points A, B, A_0, B_0—$\{A_1B_1, A_0B_0\} = \{AB, A_0B_0\}^{1/2}$, as seen in figure 7. So $\{AB, A_0B_0\} = \{A^*B^*, A_0B_0\}^2$.

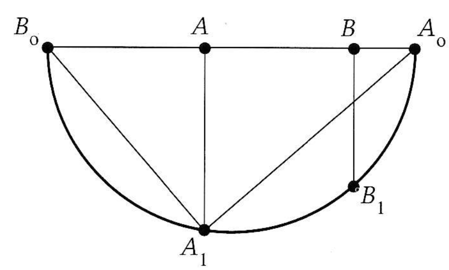


Figure 7

By the similarity of triangles $A_0A_1B_0$, A_1AB_0, and A_0AA_1, $AA_0/AB_0 = (AA_0/AA_1)(AA_1/AB_0) = (A_1A_0/A_1B_0)^2$. Similarly, $BB_0/BA_0 = (B_1B_0/B_1A_0)^2$. Multiplying the two equalities and taking the square root, we get $\{A_1B_1, A_0B_0\} = (A_1A_0 \cdot B_1B_0)/(A_1B_0 \cdot B_1A_0) = \{AB, A_0B_0\}^{1/2}$.

is equal to the square of the cross ratio $\{A^*B^*, A_0B_0\}$, where A^* and B^* are the points corresponding to A and B under the transformation of the models. You may know from "Inversion" (if from nowhere else) that the distance d between A^* and B^* in the Poincaré model is given by the formula $d(A^*, B^*) = \log\{A^*B^*, A_0B_0\}$ if the order of the points on the arc A_0B_0 is A_0, B^*, A^*, B_0. With this definition, all the axioms of non-Euclidean geometry turn out to be true for this model. So if we put

$$k(A, B) = d(A^*, B^*)$$
$$= \tfrac{1}{2}\log\{AB, A_0B_0\},$$

all of them will be satisfied for the Klein model, and this completes our construction of that model.

Now it remains to rewrite the last formula in terms of the relative speed $v = v_{B/A}$. Recall that $k(A, B) = k(O, P)$, where P is any point of the map M_O such that $OP = v$. In the notation of figure 5, the cross ratio in the definition of $k(O, P)$ equals

$$\{OP, UV\} = \frac{UO \cdot PV}{UP \cdot OV} = \frac{c(c+v)}{(c-v)c}$$
$$= \frac{1+v/c}{1-v/c},$$

because the radius of the map is $OU = OV = c$. We'll get simpler formulas if we take c as the unit speed and the logarithm to be natural (the base of the logarithm determines the unit length in the model). Finally, we come up with the following result:

$$r(A, B) = k(A, B) = \frac{1}{2}\ln\frac{1+v}{1-v},$$

where $v = v_{B/A}$. Note that with the above convention this formula turns into the nonrelativistic $AB \cong v_{B/A}$ for small values of v. Sometimes the v-distance $r(A, B)$ is called "rapidity." Though it doesn't have any direct physical sense, it's more convenient for calculations than the relative speed. Solving the last equation for v, we obtain the expression for relative speed v in terms of rapidity $r = r(A, B)$:

$$v = \frac{e^{2r}-1}{e^{2r}+1} = \frac{e^r - e^{-r}}{e^r + e^{-r}}.$$

The function on the right has a special name—*hyperbolic tangent*—and notation—tanh r. So

$$v_{B/A} = \tanh r(A, B).$$

This function is similar to the regular tangent in many ways. Try to prove, for instance, the following *addition formula*:

$$\tanh(x+y) = \frac{\tanh x + \tanh y}{1 + \tanh x \tanh y},$$

which differs from the formula for tan x only in the sign in the denominator.

This formula enables us to deduce the relativistic velocity addition rule. If object B moves with speed v_1 with respect to A, and C moves in the same direction with speed v_2 relative to B, what is the speed v of C relative to A?

Let r_1, r_2, and r be the corresponding rapidities. Since the v-points A, B, and C lie on one v-line, B between A and C, $r = r_1 + r_2$. Therefore, $v = \tanh r = \tanh(r_1 + r_2)$; substituting $x = r_1$, $y = r_2$, $v_1 = \tanh r_1$, $v_2 = \tanh r_2$ into the addition formula for the hyperbolic tangent, we get (for $c = 1$!)

$$v = \frac{v_1 + v_2}{1 + v_1v_2}.$$

Of course, it would be reckless to try to squeeze into one article a presentation of relativity with any completeness. The geometric method described above provides a tool to explain all the well-known relativistic effects—time dilation, length contraction, the twin paradox—and to treat relativistically more difficult two-dimensional problems like the star encounter we considered at the outset. But my goal was more modest: to show how Einstein's relativity principle is transformed into the negation of Euclid's parallel postulate in velocity space. The fact that these two great theories—relativity and hyperbolic geometry—prove to be intimately connected is pretty impressive, don't you think? ◘

"MISS OR HIT"
CONTINUED FROM PAGE 33

17. *From N-burg to X-ville.* Coming downstream, a steamboat moves at 20 mph; going upstream, it moves at 15 mph. The trip from N-burg to X-ville is 5 hours shorter than the trip from X-ville to N-burg. How far is N-burg from X-ville?

18. *All humanity in a square.* In 1924 the population of the Earth was 1,800,000,000. Let's imagine that all the people living at that time gathered in a compact crowd on a plain. You want to place them in a square area, allotting 1 square meter for every 20 persons (standing close together, 20 people can fit in such a square). Without doing any calculations, try to estimate the size of the square plot you'd need for this. For example, would a square with a side of 100 kilometers be big enough?

19. *Parquet-maker.* Cutting squares of wood, a parquet-maker checks them by comparing the lengths of the sides. If all four sides are equal, then he considers the square to be cut correctly. Is such a check reliable?

20. *Another parquet-maker.* Another parquet-maker checks her work in a different way: she measures not the sides but the diagonals of the square pieces. If both diagonals are equal, then the parquet-maker considers the square to be cut correctly. Do you agree?

21. *The third parquet-maker.* Yet another parquet-maker checks his work by comparing the four pieces created by the diagonals. In his opinion, the equality of all four parts proves that the quadrilateral he cut off is a square. What do you think? ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 59

"HOW ABOUT A DATE?" CONTINUED FROM PAGE 30

Please send your solutions to *Quantum*, 3140 North Washington Boulevard, Arlington, VA 22201 within a month after receipt of this issue. The best solutions will be noted in this space and their authors will receive special certificates from *Quantum*.

The tip of the iceberg

The contest problem in the September/October issue asked you to lower a pencil into a glass of water while holding it by the eraser. As you no doubt observed, the pencil will remain vertical when only a small part of it is submerged in the water. As more of the pencil descends, it reaches an equilibrium angle that is dependent on the depth.

Ben Davenport from the North Carolina School for Science and Mathematics submitted a correct solution to the problem and then provided an interesting extension that we'll tempt you with later.

A. The three forces acting on the pencil are the force of gravity, the force of buoyancy, and the force of the pivot. If we choose to take torques about the point at the top, the force of the pivot produces no torque and can be ignored (see figure 1):

$$\sum \tau_p = 0,$$
$$F_g = -\rho_r LAg,$$
$$F_b = \rho_w \left(L - \frac{H}{\cos\theta} \right) Ag.$$

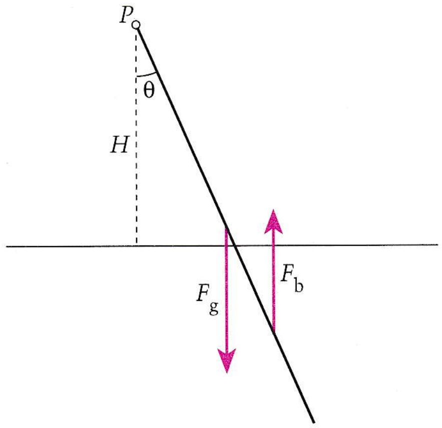


Figure 1

Using the moment arms for the torques about P, we get equation 1 in the box below. You can divide each side of equation 1 by $\sin\theta$, but you then must realize that $\sin\theta = 0$ is a solution to the equation and, therefore, θ may be equal to 0.

The second solution is

$$\cos\theta = \left(\frac{H}{L\sqrt{1-\rho_r/\rho_w}} \right),$$

where $H \le L\sqrt{1-\rho_r/\rho_w}$, since $\cos\theta$ must be less than or equal to 1.

B. In order to sketch a graph of potential energy versus theta, it's useful to derive the equation for potential energy—equation 2 in the box. The extrema of potential energy represent equilibria. You can get a sense of this by thinking about stable points for a roller coaster. For our pencil in water, we plot the potential energy versus theta on a spreadsheet for small H and for large H (fig. 2). From the graphs we can see that for large H, zero degrees is a stable equilibrium. For small H, zero degrees is an unstable equilibrium, and the angle given by the relation in part A is the stable equilibrium.

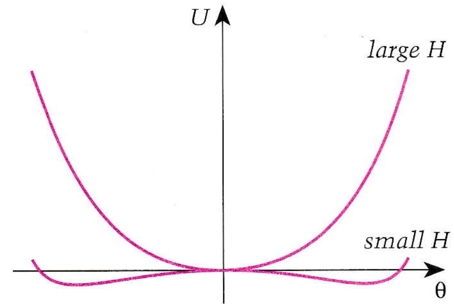


Figure 2

C. To solve the last part of the contest problem, let the submerged part of the pencil be referred to as S. Then

$$S = L - \frac{H}{\cos\theta},$$

where

$$\cos\theta = \frac{H}{L\sqrt{1-\rho_r/\rho_w}}.$$

Therefore,

$$S = L\left(1 - \sqrt{1-\rho_r/\rho_w}\right).$$

We then conclude that the submerged part of the stick is constant—that is, independent of H. You can verify this by actually performing the experiment.

$$1.\quad \sum \tau = -\left(\tfrac{1}{2}\sin\theta\right)(\rho_r LAG) + \left[L - \left(\frac{L - H/\cos\theta}{2} \right) \right] (\sin\theta)\,\rho_w \left(L - \frac{H}{\cos\theta} \right) Ag = 0$$

$$2.\quad \Delta U = -\int_0^\theta \tau\, d\theta$$
$$= -\int_0^\theta \frac{AgL^2\rho_r}{2} \left(-1 + \frac{\rho_w}{\rho_r} \left[1 - \left(\frac{H}{L\cos\theta} \right)^2 \right] \right) \sin\theta\, d\theta$$
$$= \frac{AgL^2\rho_r}{2} (1 - \cos\theta) \left[1 - \frac{\rho_w}{\rho_r} + \frac{\rho_w}{\rho_r} \left(\frac{H}{L} \right)^2 \frac{1}{\cos\theta} \right]$$

As we mentioned above, Ben Davenport challenged himself (and now we challenge the rest of our readers) with what happens if H remains constant but the pivot point of the pencil is lowered.

The details of this problem and the experimental verification of the solution can be found in an article by Joseph Priest and David F. Griffing in the April 1990 issue of *The Physics Teacher* (pp. 210–13).

The problem book of Anania of Shirak

"On the ancient peak of Ararat
The centuries have come like seconds
And passed on."
—Avetik Issahakian

by Yuli Danilov

SOME YEARS AGO JOURNALISTS INTERviewing celebrities liked to ask them: "What books would you take with you if you were to go off on a space flight?" And though the number of books allowed on the trip varied from 10 to 30, depending on the type of spacecraft and the generosity of the interviewer, and celebrities are people of the most varied tastes, ages, and professions, not one of them dared to say that he or she would want to take with them at least one book of arithmetic problems.

Some of these people certainly excluded this kind of literature because they were trained in the humanities and had nothing but scorn for "numbers" (though secretly afraid of them). Others steered clear of such puzzle books because they were masters of incomparably more difficult branches of modern mathematics and didn't mind saying for all the world to hear that they didn't know how to solve mere arithmetic problems. Professional mathematicians were no exception. Here's what the Russian mathematician Alexander Khinchin, a specialist in statistics, wrote about arithmetic: "I willingly confess that any time a fifth-grader asked me to help solve an arithmetic problem, it was a hard work for me, and sometimes I failed completely. Of course, like most of my friends, I could easily solve the problem by the natural algebraic route—constructing equations or sets of equations. But we were supposed to avoid using algebraic analysis at all costs! . . . By the way, it's a fact that is well known and oft repeated that, as a rule, neither high school graduates, nor students at teaching colleges, nor teachers beginning their careers (nor, I must add, scientific researchers) can solve arithmetic problems. It seems the only people in the world who are able to solve them are fifth-grade teachers."

Now, I'm not insisting that a book of arithmetic problems be included in the bookbag of anyone flying into space. But a sense of justice induces me to recommend one particular problem book, one that will satisfy the most fastidious taste and supply food for thought sufficient not only for a relatively short flight to the Moon but for a extended space voyage—say, to Venus and back.

One for the "road"

> They both took out the books they brought for the road. Kingsley glanced at the Royal Astronomer's book and saw a bright cover with a group of cutthroats shooting at each other with revolvers. "God knows what this kind of stuff leads to," thought Kingsley.
>
> The Royal Astronomer looked at Kingsley's book and saw the *History* of Herodotus. "Good Lord, next he'll be reading Thucydides," thought the Royal Astronomer.
>
> —Fred Hoyle, *The Black Cloud*

The book I'm talking about isn't very big, but its 24 problems constitute 24 elegant miniatures from seventh-century Armenia. Naive and wise at the same time, rich in striking detail and the bright coloration of the period, these problems are reminiscent of the reliefs on the famous monument of Armenian architecture, the church on the island of Akhtamar in Lake Van (in what is now Turkey). They are as inseparable from the image of Armenia as the elegant letters of the Armenian alphabet, invented by Mesrop Mashtots, or the songs of Komitas, or the paintings of Saryan.

An edition of these incredibly beautiful problems has long been a bibliographic rarity. It was published under the title *Problems and Solutions of Vardapet*[1] *Anania of*

[1]"Vardapet" (or "vartabed") means teacher or learned man in Armenian. (The Armenian language suffers in English from a dual transliteration scheme. Thus, Mesrop is often rendered as "Mesrob," Komitas as "Gomidas," and so on.)

Art by Yury Vashchenko

a
A
c
a
b
c
B
B

Shirak, Armenian Mathematician of the Seventh Century (translated and published by I. A. Orbeli, Petrograd, 1918).

The abundance of close observations and wide-ranging information about the way of life and customs of that remote epoch when Anania of Shirak lived and worked have actually rendered a disservice to his problem book. For many years the book was known only to researchers in the humanities—specialists in Armenian history who jealously guarded their treasure and wouldn't let just anyone see it. Even now, after research by K. P. Patkanov, the learned monk Father Haloust, J. I. Orbeli, A. Abramyan, V. K. Chaloyan, and others has brought the works of Anania of Shirak to light in scholarly circles, the general reader remains ignorant of the very existence of this remarkable problem book.

Vardapet Anania of Shirak

Once I fell in love with the art of calculation, I thought that no philosophical notion can be constructed without number, considering it the mother of all wisdom.

—Anania of Shirak

Among ancient Armenian thinkers, Vardapet Anania of Shirak stands out because of the breadth of his interests and the unique mathematical orientation of his work. Some of his works have been preserved. In addition to the *Problems and Solutions*, the following tracts have found a special place in the estimation of scholars: *On Weights and Measures*, *Cosmography and Calendrical Theory*, and *Armenian Geography of the Seventh Century A.D.* (the authorship of the last work was long attributed to another outstanding thinker of ancient Armenia, Movses of Khoren).

In his autobiography, Anania of Shirak has this to say about himself:

I, Anania of Shirak, having studied all the science of our Armenian land and having learned the Holy Scripture intimately, in the expression of the psalmist, 'every day I illuminated the eyes of my mind.' Feeling myself lacking in the art of calculation, I came to the conclusion that it is fruitless to study philosophy, the mother of all sciences, without number. I could find in Armenia neither a man versed in philosophy nor books that explained the sciences. I therefore went to Greece and met in Theodosiople a man named Iliazar who was well versed in ecclesiastical works. He told me that in Fourth Armenia[2] there lived a famous mathematician, Christosatur. I went this person and spent six months with him. But soon I noticed that Christosatur was a master not of all science but only of certain fragmentary facts.

I then went to Constantinople, where I met acquaintances who told me: "Why did you go so far, when much closer to us, in Trebizon, on the coast of Pontus,[3] lives the Byzantine vardapet Tyukhik. He is full of wisdom, is known to kings, and knows Armenian literature." I asked them how they knew this. They answered: "We saw ourselves that many people traveled long distances to become pupils of so learned a man. Indeed, the archdeacon of the patriarchate of Constantinople, Philagrus, traveled with us, bringing many young persons to become pupils of Tyukhik." When I heard this, I expressed my gratitude to God, who had quenched the thirst of His slave.

I went to Tyukhik at the monastery of St. Eugene and explained why I had come. He received me graciously and said: "I praise Our Lord that He sent you to learn and to transplant science in the domain of St. Gregory; I am glad that all your country will learn from me. I myself lived in Armenia for many years as a youth. Ignorance reigned there." Vardapet Tyukhik loved me as a son and shared all his thoughts with me. The Lord bestowed upon me His blessing: I completely assimilated the science of number, and with such success that my fellow students at the king's court began to envy me.

I spent eight years with Tyukhik and studied many books that had not been translated into our language. For the vardapet had an innumerable collection of books: secret and explicit, ecclesiastical and pagan, books on art, history, and medicine, books of chronologies. Why enumerate them by title? In a word, there is no book that Tyukhik did not have. And he had such a gift from the Holy Spirit for translating that when he sat down to translate something from the Greek into Armenian, he did not struggle as other translators did, and the translation read as if the work were written in that language originally.

Tyukhik told me how he had achieved such vast erudition and how he had learned the Armenian language. "When I was young," he said, "I lived in Trebizon, at the court of the military chief Ioannus Patricus, and for a long time, up to the accession of Mauritius to the throne, I served as a military man in Armenia and learned your language and literature. During one attack by Persian troops on the Greeks, I was wounded and escaped to Antioch. I lost all my possessions. Praying to the Lord to heal my wounds, I made a promise: 'If You prolong my life, I shall dedicate it not to accumulating perishable treasures but to collecting treasures of knowledge.' And the Lord heard my prayers. After I recovered I went to Jerusalem, and from there to Alexandria and Rome. Upon returning to Constantinople, I met a famous philosopher from Athens and studied with him for many years. After that I returned to my homeland and began to teach and instruct my people."

After some years that philosopher died. Not finding a replacement for him, the king and his courtiers sent for Tyukhik and invited him to assume the teacher's position. Tyukhik, citing the promise he made to God not to move far from the city, turned down the offer. But because of his wide learning, people came streaming from all countries to study with him.

And I, the most insignificant of all Armenians, having

[2]Fourth Armenia was one of fifteen provinces into which, according to *Armenian Geography in the Seventh Century A.D.*, so-called Great Armenia was divided.

[3]"Pontus" (or "Pontus Euxinus") was an old name for the Black Sea.

learned from him this powerful science, desired by kings, brought it to our country, supported by no one, obligated only to my own industry, God's help, and the prayers of the Blessed Educator. And no one thanked me for my efforts.

Problems and solutions

> A half and one sixth and one nineteenth of all the books were printed on vergé; one fifth and one two-hundred-eighty-fifth—on rag paper; one forty-fifth and one eight-hundred-fifty-fifth—on vellum; and forty-five inscribed copies—on Dutch paper. And so, find how many copies were printed in all.
>
> —Imitation of Anania of Shirak

A Latin proverb says *habent sua fata libelli* ("books have their own fate"). The fate of *Problems and Solutions* by Anania of Shirak is quite amazing. The manuscripts of Anania's book were preserved only because, according to Armenian historians, "in ancient and medieval Armenia manuscripts were guarded from invaders, like weapons, and cherished, like one's own children." Biding their time, the manuscripts lay in the Matenadaran, a renowned depository of ancient manuscripts (now the Mesrop Mashtots Institute of Ancient Manuscripts). And its hour finally arrived. In 1896 the learned monk Father Haloust used two manuscripts to publish the problem book, supplementing it with an introduction and commentary. In 1918 the book was translated into Russian, edited, annotated, and typeset by Iosef Orbeli, a prominent scholar (and later a member of the Academy of Sciences of the USSR).

In the translator's words, the problems of Anania are "amusing, full of life, and simple." Orbeli goes on to say: "The subjects of the problems are generally taken from everyday life. The scene is predominantly his homeland Shirak and the surrounding countryside, and the *dramatis personae*, if they are named, are the local princes—the Kamsarakans, including Nersekh, who was a contemporary of Anania." Like other ancient authors, Anania of Shirak used only "aliquots"—that is, fractions with a numerator of 1. When it is necessary to write fractions with numerators other than 1, one has to represent it as a sum of aliquots (see the epigraph above).

Like any true work of art, the problems of Anania suffer terribly in the retelling. You have to read the originals (albeit in translation) in their full glory. So let's open Anania's problem book—a gift from across the ages.

Problems 1 and 8 relate to the Armenian uprising against the Persians in A.D. 572.

Problem 1

My father told me the following story. During the famous wars between the Armenians and the Persians, prince Zaurak Kamsarakan performed extraordinary heroic deeds. Three times in a single month he attacked the Persian troops. The first time, he struck down half of the Persian army. The second time, pursuing the Persians, he slaughtered one fourth of the soldiers. The third time, he destroyed one eleventh of the Persian army. The Persians who were still alive, numbering two hundred eighty, fled to Nakhichevan. And so, from this remainder, find how many Persian soldiers there were before the massacre.

Problem 8

During the famous Armenian uprising against the Persians, when Zaurak Kamsarakan killed Suren, one of the Armenian azats[4] sent an envoy to the Persian king to report the baleful news. The envoy covered fifty miles in a day. Fifteen days later, when he learned of this, Zaurak Kamsarakan sent riders in pursuit to bring the envoy back. The riders covered eighty miles in a day. And so, find how many days it took them to catch the envoy.

Problem 18 mentions vessels made of varying amounts of metal. In the Russian translation, they are all called "dishes." But in the original Armenian, according to Orbeli's note, the dishes in the first and second instances are called *mesur*, and in the third instance *scutel*. *Scutel* is a common Armenian word, but *mesur* had not been encountered in Armenian literature before Anania's *Problems and Solutions*.

Problem 18

There was a tray in my house. I melted it down and made other vessels from the metal. From one third I made a mesur; from one fourth, another mesur; from one fifth, two goblets; from one sixth, two scutels; and from two hundred ten drams, I made a bowl. And now, find the weight of the tray.

Several of the problems reflect the richness of the Caucasian fauna in Anania's time—for instance, problem 7.

Problem 7

Once I was in Marmet, the capital of the Kamsarakans. Strolling along the bank of the river Akhuryan, I saw a school of fish and ordered that a net be cast. We caught a half and a quarter of the school, and all the fishes that slipped out of the net ended up in a creel. When I looked in the creel, I found forty-five

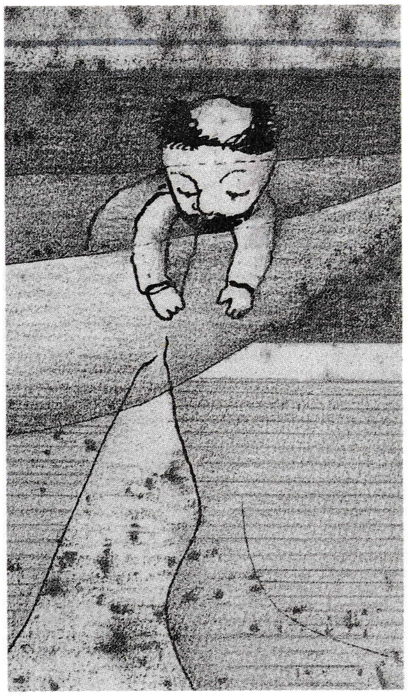
---

[4] "Azats" were members of one of several strata of freemen in ancient Armenia.

fishes. And now, find how many fishes there were in all.

The temptation is great to present all 24 problems. But I'll restrain myself and offer you just one more.

Problem 20 provides some interesting information about the wild animals that inhabited Armenia at one time but now extinct for so long that there is no mention of them even in zoological reference books. The wild donkey, according to the generally accepted view, never roamed the Armenian lands. Yet Anania of Shirak offers evidence to the contrary.



Problem 20

The hunting preserve of Nersekh Kamsarakan, ter[5] of Shirak and Asharunik, was at the base of the mountain called Artin. One night great herds of wild donkey entered the preserve. The hunters could not cope with the donkeys and, running to the village of Talin, told Nersekh about them. When he arrived with his brothers and azats and entered the preserve, they began killing the wild beasts. Half of the animals were caught in traps, one fourth were killed by arrows. The young, which constituted one twelfth of all the animals, were caught alive, and three hundred sixty wild donkeys were killed by spears. And so, find how many beasts there were at the start of this massacre.

"Set in type by me, Iosef Orbeli"

His biography could not be squeezed into the framework of a bibliography.

—K. Uzbashyan, *Academician Iosef Abgarovich Orbeli*

Anyone who is lucky enough to hold a copy ($1/n$ of the small printing—n is the solution to the epigraph in the previous section) of the Russian translation of Anania of Shirak's *Problems and Solutions*, a thin book with yellowed pages, has probably noticed the variety of the fonts, the elegance of the borders, and the high quality of the design, printing, and binding. Such great attention to detail is characteristic of works that fulfill a requirement for a degree in bookmaking. And this problem book was indeed a kind of diploma attesting to the professional maturity of the man who created it. An advertisement at the end of the book reads: "This book was typeset in December 1917 at the printing offices of the Russian Academy of Sciences by me, Iosef Orbeli; the text was also proofread, laid out, and decorated with borders by me. Various circumstances prevented me from carrying this project to the end; the final pages of the book were typeset by M. Strolman."

Typesetting was neither the first nor the only profession of the renowned orientalist Iosef Orbeli, who later became the director of the Hermitage Museum in Leningrad. He was also a cabinetmaker and a locksmith. Orbeli had already become acquainted with the famous academic printing house Typis Academiae, founded in 1728 and known all over the scientific world for its rich collection of fonts and its virtuoso typesetters. In preparing to publish the corpus of ancient inscriptions preserved on the walls of Armenian churches, Orbeli found it necessary to create a new font that would preserve the unique signs and ligatures. This complicated work was done by M. G. Strolman. (Unfortunately the entire set of letters was destroyed during the blockade of Leningrad in World War II.)

When Orbeli came to the printing offices of the Academy of Sciences, times were hard. The only way to publish the newly translated *Problems* of Anania was for Orbeli to learn typesetting (he had always been attracted to the printer's craft). In 1922 Orbeli became the director of printing at the Academy of Sciences. Even after he retired, he remained a tireless champion of Russian academic typography.

Back to Earth

This book by definition does not exhaust all the most important works in this domain. The editor hopes that those who are guilty of this incompleteness will read these lines and, stung by shame, will work up, if not a collection like this, at least a monograph.

—V. Bonch-Bruyevich, introduction to the Russian translation of *Solid-Body Symmetry* by R. Knox and A. Gold

Let's imagine a time when space flight is an everyday thing, and high schoolers will spend their breaks as astronauts-in-training in the Perelman crater on the far side of the Moon. Maybe one of the space travelers will take this very copy of *Quantum*, and another, looking over her shoulder, will read this article and say to himself: "This Anania from Shirak seems like a pretty interesting guy. When I get home I'll try to find his problems."

Good luck, my young friend! Anania is sure to entertain you. Perhaps by then there will be more than n copies of his timeless *Problems and Solutions*. And we can hope they will be as lovingly printed as the masterpieces created by Iosef Orbeli.

[5] "Ter" was the title of the heads of sovereign royal families in ancient Armenia.

Landau's license plate game

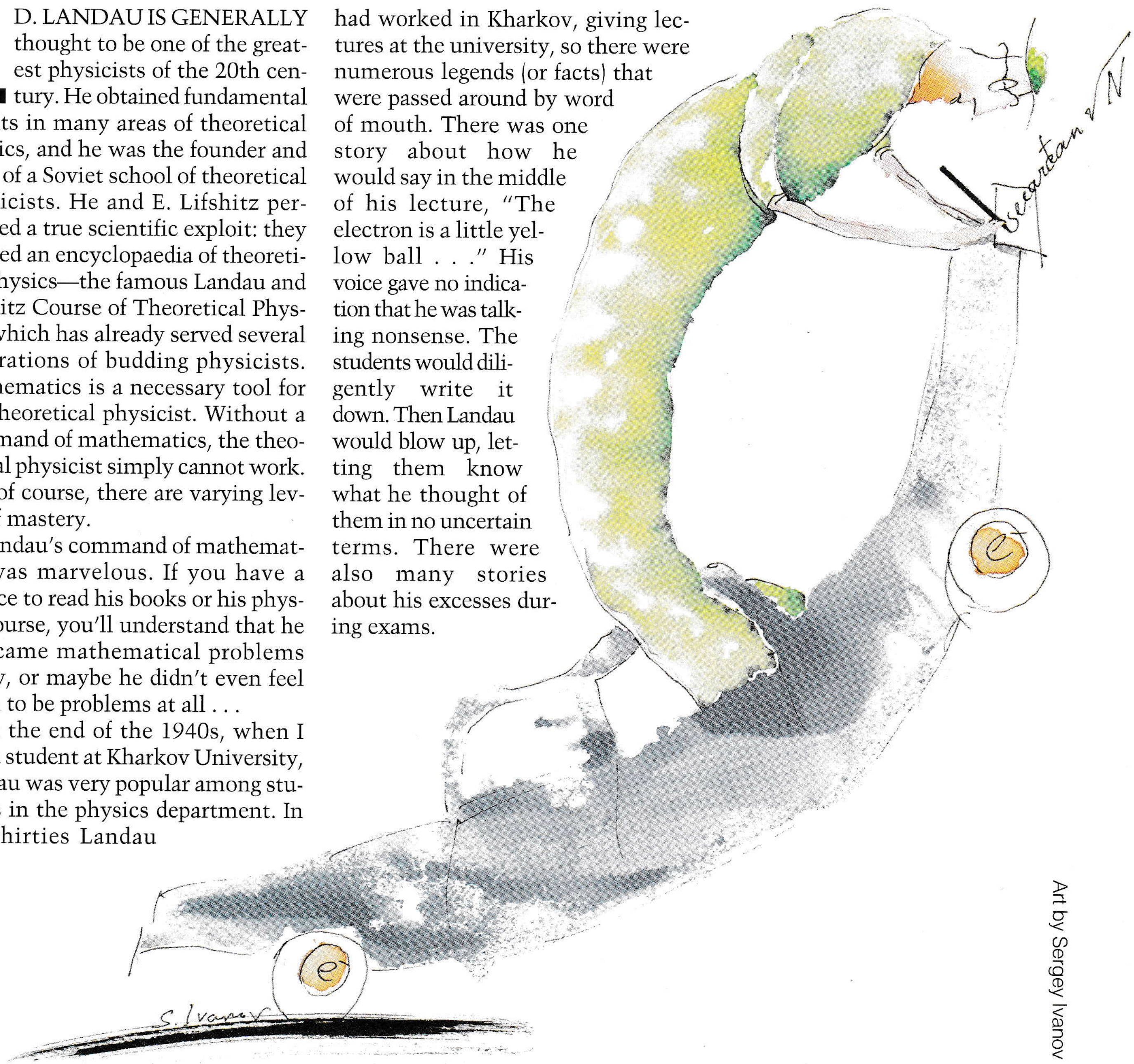
And other mathematical feats of a great physicist

by M. I. Kaganov

L. D. LANDAU IS GENERALLY thought to be one of the greatest physicists of the 20th century. He obtained fundamental results in many areas of theoretical physics, and he was the founder and head of a Soviet school of theoretical physicists. He and E. Lifshitz performed a true scientific exploit: they created an encyclopaedia of theoretical physics—the famous Landau and Lifshitz Course of Theoretical Physics, which has already served several generations of budding physicists. Mathematics is a necessary tool for the theoretical physicist. Without a command of mathematics, the theoretical physicist simply cannot work. But, of course, there are varying levels of mastery.

Landau's command of mathematics was marvelous. If you have a chance to read his books or his physics course, you'll understand that he overcame mathematical problems easily, or maybe he didn't even feel them to be problems at all . . .

At the end of the 1940s, when I was a student at Kharkov University, Landau was very popular among students in the physics department. In the thirties Landau had worked in Kharkov, giving lectures at the university, so there were numerous legends (or facts) that were passed around by word of mouth. There was one story about how he would say in the middle of his lecture, "The electron is a little yellow ball . . ." His voice gave no indication that he was talking nonsense. The students would diligently write it down. Then Landau would blow up, letting them know what he thought of them in no uncertain terms. There were also many stories about his excesses during exams.



Art by Sergey Ivanov

Especially popular were the stories about Landau's mathematical talents. At that time we studied mathematics from a book of problems that we students nicknamed *The Ten Authors.* It was considered pretty difficult. In fact, we didn't always manage to solve the assigned problems. Well, there was a legend (or was it true?) that Landau had solved all the problems in *The Ten Authors* twice: once any old way, and once the right way.

And here's what I heard from my teacher, I. M. Lifshitz, a theoretical physicist who knew his math cold.

Landau thought there was no need to make a special study of probability theory. If you understand the problem, you can always obtain the answer using standard logic and, if necessary, differential and integral calculus.

One time Lifshitz was arguing with Landau, and he gave him a tough problem in probability theory. Landau couldn't give him an answer off the top of his head. That bothered him. But that evening he called Lifshitz and gave him an original, and correct, solution to the problem. Unfortunately I've forgotten the problem, but I do remember that it wasn't a trivial one.

Landau believed in his mathematical abilities and, no doubt, he had good reason. But sometimes his self-confidence would lead him to value his intuition too highly (at least, when the problem didn't require serious thought).

Of all number games for one player, I remember the license plate game best of all. No doubt that's because L. D. Landau taught it to me. The point of the game is to make an equality out of any four-digit license plate number. (Soviet license plate numbers are of the form "*AB-CD*"—for example, "12-34.") Here are the rules: you can use only the arithmetic, algebraic, and trigonometric operations learned in school; you're not allowed to rearrange the numbers; and you have to work out the solution in your head. In other words, you have to turn "-" into "=" by inserting signs known to any high school student ($+$, $-$, $\times$, $\div$, $\sqrt{}$, log, cos, and so on) between the numbers. Some numbers are easy—for instance, 75-31 ($7 - 5 = 3 - 1$). Or 38-53 ($\sqrt[3]{8} = 5 - 3$). And here's a number that doesn't cry out for ink: 27-33 ($27 = 3^3$). But there are harder numbers—for example, 75-33. Both ways of solving it will seem arcane to the novice player: $7 - 5 = \log_{\sqrt{3}} 3$ or $7 - 5 = 3! \div 3$.

Passionate players of the license plate game often argued about which operations are allowed and which are prohibited: the problem was, they didn't know the exact boundaries of the high school curriculum. In particular, they argued about the propriety of using the factorial sign "!" (which came in handy with certain numbers that stubbornly resisted being made into an equation).

At the time Landau described the game to me, he was a brilliant player. He would come up with an answer almost as soon as he saw a license plate, *any* license plate.[1] Still, there were stumpers—for instance, 75-65. Of course, one could use the function $E(x)$ equal to an integral part of x: $E(7 \div 5) = E(6 \div 5)$, but this function wasn't studied at school. Besides, if the function $E(x)$ were allowed, the game would get pretty dull.

The question of an "existence theorem" came up. I put the question to Landau: "Is it always possible to make an equality out of a license plate number?"

"No," came the emphatic answer.

I was surprised. "So you've proven the nonexistence theorem?!"

"No," Landau said with conviction, "but I haven't been able to solve all the license plates."

Infected by the license plate game, I spread the infection among the young mathematicians of my acquaintance. One of them, Y. Gandel, took the game quite seriously and proved the existence theorem. He showed that, using functions taught in high school, you can "equate" any two integral numbers, since there is a formula for reducing $N + 1$ to N.

The proof of the reduction formula requires knowledge of one trigonometric formula and skill in dealing with inverse trigonometric functions—"arc . . ." Indeed, $\sqrt{N+1} = \sec\arctan\sqrt{N}$.

Alas, after the existence theorem was proved the game lost its charm, because it became possible to equate any numbers by applying the reduction formula several times.

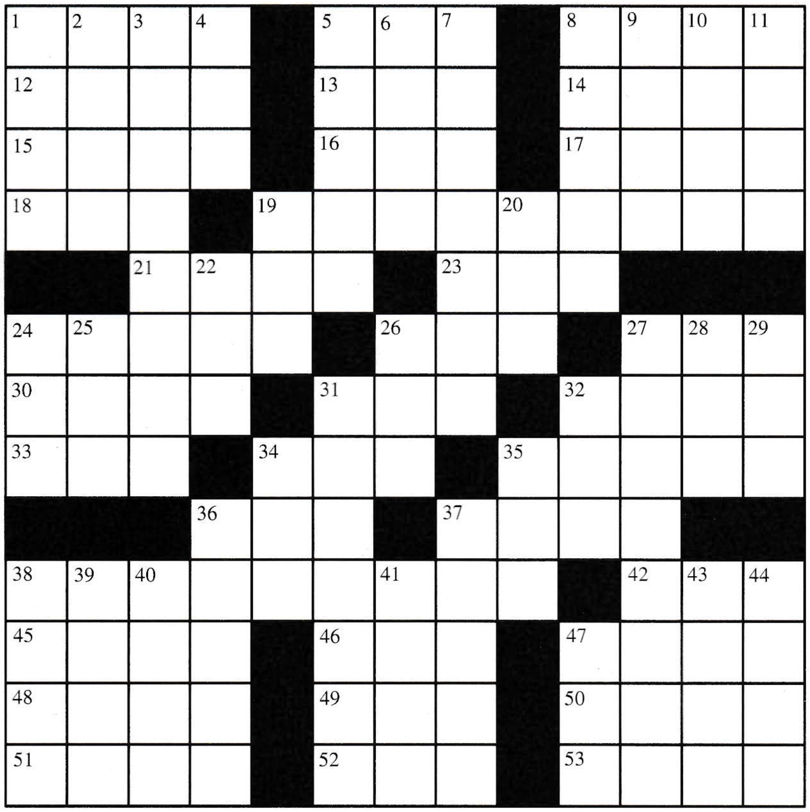
I brought the proof to Landau. He liked it very much, and we discussed (half in jest) whether it would be worth publishing it in some scientific journal. "Maybe we shouldn't," Landau said. "The mathematicians will be offended. They're already mad at me!"

Before I stop, I'd like to stress once again that Landau believed in his mathematical abilities, and his confidence helped him solve difficult, important problems—much more difficult than those in the license plate game. ■

[1]It's not advisable to play the license plate game while crossing the street or driving a car.

Criss × cross science

by David R. Martin



Across

1 Before coulomb or volt
5 Standard time plus one hour (abbr.)
8 Seaport in Lebanon
12 European capital
13 Small European deer
14 One who inherits
15 Pigment
16 After benz or propyl
17 College near London
18 Latin for openings
19 An iron oxide
21 Element 82
23 Title of respect in Turkey
24 Airship
26 Before now
27 ___ Lincoln
30 Bills
31 Natural source of metal
32 Computer time expender
33 Am. Indian
34 Calendar abbr.
35 Star seen in mitosis
36 Belonging to us
37 Arm bone
38 $C_{10}H_{13}N_5O_4$
42 Greek letter
45 1,003
46 Pull
47 Fishing reel (to a Scot.)
48 Site of Lima
49 Person in Togo
50 Med. school sub. (abbr.)
51 Stalk
52 Father
53 Cruel Roman

Down

1 Lone performance
2 Russian emperor
3 Low pH
4 Foot digit
5 Feeling about finals
6 ___ and dance (cheap)
7 Adolescent
8 Greek letter
9 Abominable Snowman
10 Violent uproar
11 Sea eagle
19 Transform (as in math)
20 Self
22 Printer's measures
24 Unit of energy (abbr.)
25 Allow
26 Part of a circle
27 Element 85
28 Member of superfamily Apoidea
29 Make a mistake
31 Unit of magnetic field strength
32 Military branch (abbr.)
34 Pair
35 Beerlike drink
36 Isologue of ammonium (suff.)
37 Single
38 Units of current
39 Food regimen
40 Ireland
41 Home to Hawkeyes
43 Moroccan tree
44 To (poetically)
47 Layer of hard soil

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

P	H	O	B	E		M	O	A	T	S		S	E	T	U	P
A	A	R	O	N		A	R	I	E	L		A	L	I	N	E
C	L	A	S	S	I	C	A	L	M	E	C	H	A	N	I	C
E	L	L	S		B	E	N		W	E	C	A	N			
				H	I			C	A	P	S			A	D	A
A	B	A	C	U	S		N	A	V	Y		M	O	D	E	L
T	E	X	A	S		D	O	N	E		N	E	E	D	L	E
O	T	I	S		F	E	E	T		K	U	R	D			
M	A	S	T	E	R	A	L	I	E	N	B	E	I	N	G	S
			I	R	A	N		L	A	O	S		P	O	L	E
S	T	I	N	G	Y		F	E	R	T		L	U	B	E	S
P	R	O	G	S		F	A	N	S		P	A	S	S	E	S
Y	I	N			E	L	I	A			L	R				
			O	S	C	A	R		E	W	E		S	A	G	A
Q	U	A	N	T	U	M	M	E	C	H	A	N	I	C	A	L
U	T	I	C	A		E	A	G	R	E		O	M	E	G	A
M	E	T	E	R		S	N	O	U	T		W	I	S	E	R

The satellite paradox

So you think the atmosphere slows a satellite down . . .

by Y. G. Pavlenko

As it moves in the Earth's upper atmosphere, a satellite is slowed by the air. Before satellites were launched, data on the atmospheric density at altitudes up to 200 km were obtained by geophysical rockets. Later, analysis of satellite motion provided these data.

The force of resistance acting on a satellite in the upper atmosphere is due to the collisions of air molecules. The direction of this force is opposite the satellite's velocity and can be written in the form

$$\mathbf{F} = -k\mathbf{v}, \qquad (1)$$

where the factor k is positive and generally depends on the velocity. The greater the cross-sectional area S of a body and the greater the density ρ of the atmosphere, the greater k is.

Let's compare the acceleration imparted to the satellite by the air resistance and the Earth's attractive force. Suppose a satellite of mass m = 100 kg is shaped like a ball with a cross-sectional area S = 1 m^2 and moves in a circular orbit with a speed v = 8 km/s at an altitude h = 160 km, where the density of the atmosphere is approximately $\rho = 10^{-9}$ kg/m^3. To get a good approximation, we can take $k = S\rho v$. In this case, the absolute value of the acceleration caused by the resistance force is

$$a = \frac{kv}{m} \cong \frac{S\rho v^2}{m} \cong 6.4 \cdot 10^{-4} \ \text{m/s}^2.$$

The attractive force

$$F = G\frac{Mm}{(R+h)^2}$$

acting on the satellite from the Earth ($M = 5.976 \cdot 10^{24}$ kg is the Earth's mass, R = 6,371 km is the average radius of the Earth, and $G = 6.673 \cdot 10^{-11}$ N $\cdot$ m^2/kg^2 is the gravitational constant) gives it an acceleration directed toward the center of the Earth that is equal to

$$a_g = G\frac{M}{(R+h)^2} \cong 9.35 \ \text{m/s}^2.$$

We see that the so-called "braking" acceleration is about 1/15,000 of the centripetal acceleration. Nevertheless, the impact of the atmosphere on the satellite at an altitude of about 160 km is so significant that a satellite begins to descend rapidly after one or two orbits around the Earth.

The first Soviet satellite was shaped like a ball with a diameter of 58 cm and a mass of 83.6 kg. The booster rocket was much larger. You'd think that the rocket, after achieving its orbit, would fall behind the satellite, since the rocket has a greater S and so is affected by greater air resistance. Observations showed, however, that the booster was far ahead of the satellite. Let's try to investigate this paradox.

Nowadays the motion of satellites is calculated on computers, since it is difficult to solve Newton's laws when the resistance depends on a body's velocity. However, we can resolve the satellite paradox by looking at the energy of the satellite. If there is no air resistance, the total mechanical energy E of the satellite is constant:

$$E = \frac{mv^2}{2} + \left(-G\frac{Mm}{r}\right) = \text{constant}, \qquad (2)$$

where $-GMm/r$ is the potential energy of the satellite at a distance r from the Earth's center. When air resistance is present, the mechanical energy no longer remains constant, but depends on time: $E = E(t)$. The change in mechanical energy ΔE during a small displacement Δs is equal to the work performed by the resistance:

$$\Delta E = W = F\Delta s, \qquad (3)$$

where $\Delta s = v \cdot \Delta t$. Substituting this expression and the force from equation (1), we obtain the expression for the rate of change in the satellite's mechanical energy:

$$\frac{\Delta E}{\Delta t} = -kv^2. \qquad (4)$$

Assume that the satellite has been placed in a circular orbit with a radius r. In the absence of any air resistance, the satellite's speed can be found by

Art by Mariya Kokorina

equating the centripetal force required for circular motion to the gravitational force:

$$m\frac{v^2}{r} = G\frac{Mm}{r^2} \Rightarrow v^2 = G\frac{M}{r}. \qquad (5)$$

The influence of the atmosphere leads to a distortion of the satellite's circular trajectory, transforming it into a spiral. If the trajectory does not deviate much from a circle, the relationship between the speed $v(t)$ and the radius $r(t)$ is still given by equation (5) at any moment. Now, however, the speed and the "radius of the circle" are functions of time.

Substituting equation (5) into equation (2), we can write the potential energy of the satellite in the form

$$U = -G\frac{Mm}{r} = -mv^2.$$

Therefore, the total mechanical energy in equation (1) can be written in the form

$$E = \frac{mv^2}{2} + \left(-mv^2\right) = -\frac{mv^2}{2}.$$

Let's now calculate the change in the total mechanical energy ΔE of the satellite that occurs for a very small increase in its speed Δv:

$$\begin{aligned}\Delta E &= \frac{-m(v+\Delta v)^2}{2} - \left(-\frac{mv^2}{2}\right)\\ &= \frac{-2mv\Delta v}{2} - \frac{m\Delta v^2}{2} \\ &\cong -mv\Delta v,\end{aligned} \qquad (6)$$

where we have neglected the term Δv^2 because Δv is very small. Dividing both sides of equation (6) by a small time interval Δt, we obtain

$$\frac{\Delta E}{\Delta t} = -mv\frac{\Delta v}{\Delta t}. \qquad (7)$$

Equating the expressions in equations (4) and (7), we get an expression for the change in the speed of our satellite:

$$\frac{\Delta v}{\Delta t} = \frac{k}{m}v. \qquad (8)$$

So we can see that the satellite's speed increases with time. The greater the ratio k/m, the faster it increases. The value of k/m is greater for the booster rocket than for the satellite, and so the booster's speed increases more rapidly. That's why the booster passes up the satellite after carrying it into orbit.

From the relationships we've been examining, we can draw a conclusion about how the atmosphere changes an orbiting satellite's energy. The force of resistance causes the satellite to begin to fall. Its speed and, consequently, its kinetic energy increase as it approaches the Earth, and its potential energy decreases (remaining negative). From equation (4) or (7) we see that the total mechanical energy also decreases. So the reduction in potential energy occurs more rapidly than the increase in kinetic energy.

Thus, because of the interaction with the atmosphere, satellites accelerate when they descend even though they don't have any engines. Upon entering the dense layers of the atmosphere, they burn up just like a "shooting star" (actually a meteorite). And that's why people nowadays have more opportunities to "wish upon a star."

Behind the scenes at the IMO

A look at the subtle way problems are selected and graded

by Vladimir Dubrovsky

ALMOST A YEAR AGO, WHEN it became more or less clear that in spite of all unforseen obstacles and difficulties Moscow would be able to host the 33rd International Mathematical Olympiad (IMO), as scheduled several years ago, a group of mathematicians from Moscow, St. Petersburg, and some other places in the former Soviet Union were invited to take part in the IMO as official coordinators. These were people who have been engaged for years in all kinds of math contests. I happened to be among them, and it wasn't until this invitation that I gave any thought to how elaborate and carefully designed an event like this must be. Just imagine: 350 participants from 65 nations, writing their papers in their own languages—as different as Bulgarian and Japanese, Greek and Hebrew, Indonesian and Danish . . . And all this diversity, enhanced by different courses of study, different educational traditions, different mentalities, after all, must be reduced to the common denominator of some unified system of scoring and placing that should satisfy everybody! It was interesting for me to get to know how this concealed mechanism works (and to help it work), and I thought this would be interesting for *Quantum* readers, too—all the more so because usually the reports of Olympiads are written by the participants, so the reader seldom has the opportunity to look at these competitions from the other side. (The article by Cecil Rousseau and Daniel Ullman in the previous issue of *Quantum* gives a good idea of what a team—in particular, the US team—does before and during the Olympiad.)


XXXIII
INTERNATIONAL
MATHEMATICAL
OLYMPIAD
JULY 10-21

Certificate

XXXIII Международная математическая Олимпиада
Россия, Москва, 1992

Chairman of the 1992 IMO Jury

The problems for the exam are proposed by participating countries (except the organizers). They began to arrive in Moscow in May, and that was the starting point of the coordinators' work. At this stage, the task was to select from the 60–70 problems received (not as many as one might expect, though) a preliminary list of 20–30 problems with detailed solutions (sometimes with commentary) for the jury, which consists of all team leaders, to make the final choice.

As usual, a number of problems were added to this list at the last moment, a few days before the official opening of the competition, when the team leaders brought them to Moscow.

The jury spent two days discussing the problems for the exam. Here are the results of the final vote (three problems for each of two days of competition).

IMO problems . . .

1. Find all integers a, b, c ($1 < a < b < c$) such that $(a - 1)(b - 1)(c - 1)$ is a divisor of $abc - 1$. (New Zealand)

2. Let $\mathbf{R}$ denote the set of all real numbers. Find all functions f: $\mathbf{R} \to \mathbf{R}$ such that

$$f(x^2 + f(y)) = y + (f(x))^2$$

for all x, y in $\mathbf{R}$. (India)

3. Consider nine points in space, no four of which are coplanar. Each pair of points is joined by an edge (that is, a line segment), and each edge is either colored blue or red or left uncolored. Find the smallest value of n such that whenever exactly n edges are colored, the set of colored edges necessarily contains a triangle all of whose edges have the same color. (China)

4. In the plane let C be a circle, L a line tangent to circle C, and M a point on L. Find the locus of all points P with the following property: there exist two points Q, R on L such that M is the midpoint of QR and C is the inscribed circle of triangle PQR. (France)

5. Let S be a finite set of points in three-dimensional space. Let S_x, S_y, S_z be the sets consisting of the or-

thogonal projections of the points of S onto the yz-plane, zx-plane, and xy-plane, respectively. Prove that

$$|S|^2 \leq |S_x| \cdot |S_y| \cdot |S_z|,$$

where $|A|$ denotes the number of elements in the finite set A. (Note: the orthogonal projection of a point onto a plane is the foot of the perpendicular from that point to the plane.) (Italy)

6. For each positive integer n, $S(n)$ is defined as the greatest integer such that, for every positive integer $k \leq S(n)$, n^2 can be written as the sum of k positive square integers. (a) Prove that $S(n) \leq n^2 - 14$ for each $n \geq 4$. (b) Find an integer n such that $S(n) = n^2 - 14$. (c) Prove that there are infinitely many integers n such that $S(n) = n^2 - 14$. (United Kingdom)

The solutions to these problems will appear in the next issue.

... and the best of the rest

Of course, it was impossible to include all the interesting problems in the final list, so the jury was forced to weed out some of them. And in some cases I was surprised at the selection—tastes (and motives) differ. So I'd like to share with you at least one problem from the "leftovers." I think it was the most interesting—though perhaps a bit too difficult, even for an IMO. It was a geometry problem proposed by China, and it presents new properties of a rather well-known construction.

On the sides of a convex quadrilateral $ABCD$ with perpendicular diagonals, the squares ABB_1A_2, BCC_1B_2, CDD_1C_2, and DAA_1D_2 are constructed externally. Prove that two convex quadrilaterals, one bounded by straight lines AB_2, BC_2, CD_2, DA_2 and the other one bounded by AD_1, DC_1, CB_1, BA_1, are congruent.

We'll publish the solution, along with other remarkable properties of this construction, in a future issue of *Quantum*.

The delicate task of scoring

As soon as the questions for the exam were selected, the coordinators set to work devising the guidelines for scoring the solutions, trying to anticipate possible achievements and flaws in the papers, and assigning the number of points to be added—or taken away—for them. The guidelines were explained to the jury in more detail at a special meeting. Question 4 generated a heated argument. It was the only geometric problem, but even this unique opportunity for "geometers" to steal the spotlight was spoiled because the problem had an algebraic (coordinate) solution. So the "algebraically minded" contestants gained a certain advantage over "pure geometers." Since there was only one geometric problem in the set, this partiality was impossible to even up, and we came up with a rather symbolic suggestion: to take away one point if the answer to problem 4 is described only by an equation without being rendered in geometric terms. Most of the jury members agreed there was certain unfairness and that, in general, geometry should be supported, but they argued that a solution is a solution no matter how it was obtained or expressed. According to the regulations, however, the official coordinators have the upper hand in formulating the rules for grading, so the discussion was somewhat theoretical. After all, our biased grading really didn't influence the final results. But maybe it will influence the selection of problems in future Olympiads and make it more balanced.

Until the second round of the exam, the jury and the teams stayed in separate hotels. But the time for a reunion had come, and while the contestants were still working, the team leaders moved to join their assistants (who were taking care of the teams) and, with the guidelines in hand, began to check the students' papers. And in the evening the "coordination" began.

Coordinating the results

The coordinators, divided into small groups responsible for particular problems, settled down around a dozen tables in a big hall. There they received the team leaders and assistants, who took turns presenting their teams' papers. Contrary to any doubts, it turned out that even without knowing the language, but asking when necessary (usually, in English) questions like "What is written where?" one could form a definite enough opinion about every answer. So this system proved to be quite efficient. The coordinators tried to be strict (one of them, Dmitry Tereshin, even earned the nickname "Dmitry the Terrible"), but they were never reproached for being unfair.

Perhaps the most memorable moment came during the coordination of question 5, and it involved the Chinese team (it was the last problem on their coordination schedule). The team leaders figured on a total of 34 points for this question (four complete 7-point solutions and a 6-point solution containing a minor flaw). But the coordinators discovered that the sixth member of the team, whom the leaders thought had utterly failed on this problem, had submitted a perfect solution, and they gave him a full 7-point score. At this moment everybody at the table stood up and shook hands: the coordinators congratulated the Chinese team.

Medals galore

Although it had been absolutely clear by then that nobody could challenge the Chinese (not this time), it was an impressive finishing touch to the scenario of their overall victory—six gold medals and a total of 240 points out of a possible 252. (By the way, last year Chinese teams won handily not only the math but the physics and computer science international olympiads, too!) The US team placed second with 181 points—3 gold and 3 silver medals; Romania was third (177 points—2 gold, 2 silver, and 2 bronze medals); the CIS, or former Soviet Union, was fourth (176 points—2 gold and 3 silver medals); next came the United Kingdom (168 points—2 gold, 2 silver, and 2 bronze medals) and Russia (158 points—2 gold, 2 silver, and 2 bronze medals).

Three special prizes were established by *Kvant*, the Russian sister magazine of *Quantum*, which was represented at the Olympiad by a group of members of its editorial board. The prize for the most elegant solution to question 1 was awarded to Pinai Linvong of Thailand; the prize for the best coordination (checking and presentation of solutions) went to Lisa McShine, the leader of the team from Trinidad and Tobago; and the prize for the most interesting problem at the Olympiad was given to Anthony Gardiner, the leader of the UK team, who created question 6.

The coordinators, naturally, didn't get any medals. But we got great satisfaction from the favorable comments on our work from all the team leaders. And it was a joy to work with such an outstanding group of mathematicians, young and—not quite so young.

Bulletin board

Jobs in space

"I told my kids that they will be living and working in space. They don't believe me. So, we made a show about it," says Jaime Escalante, the celebrated math teacher portrayed in the Academy Award–nominated film "Stand and Deliver." Escalante, Kathy Bates, Pat Morita, "Weird Al" Yankovic, and others host "Living and Working in Space," a prime-time special airing on PBS Wednesday, March 31, at 8:00 P.M. (check local listings).

"Living and Working in Space" opens the door to the abundance of opportunities for today's young people to live and work in space. It also promotes the study of math and science in order to cope in a technology-based society. "Don't think that you have to train to be an astronaut in order to be in space thirty years from now," says Dr. John Lewis at the University of Arizona Lunar & Planetary Lab. "It's not true. If you train yourself as a chemist or an engineer or as a restaurant manager, there will be a niche for you. Humanity is going into space."

The one-hour program features a series of interviews with today's space professionals, ranging from a space doctor and the designers of ground vehicles for Mars, to a woman who designs space clothes and a scientist known as the "lunar lettuce man." Also woven throughout the program are inventive dramatic and humorous vignettes that play on the interviews with industry professionals. The vignettes explore what life in a lunar or Martian habitat might be like.

This program for the entire family is funded by ARCO and the US Department of Energy. Additional funding for educational materials and activities is made possible by NASA. The program is produced by the Foundation for Advancements in Science and Education (FASE).

Sky Awareness Week

The week of April 25–May 1 has been designated Sky Awareness Week, a national celebration of the sky. The week provides opportunities for teachers, students, parents, nature center staff, television meteorologists, and others to look toward the sky and learn how to read and understand sky processes. Sky Awareness Week falls during the same week as National Science and Technology Week, and around the same time as Earth Day and other events focusing on our planet and its environment.

For more information about Sky Awareness Week 1993, contact Barbara G. Levine, THINK WEATHER, Inc., 1522 Baylor Avenue, Rockville, MD 20850. For a 10-page guide entitled "101 Ways to Celebrate Sky Awareness Week," send $3.00 to cover printing and mailing costs.

Thinking computers

In December, humans were pitted against computers once again in the second annual Quest for the Thinking Computer. In the contest, judges converse at computer terminals to try to determine which terminals are controlled by people and which by computers. In 1991, Programmer Joseph Weintraub's program "whimsical conversation" fooled half the judges into thinking it was a person. Weintraub was also the 1992 winner—his program "Men vs. Women" fooled two of eight judges.

The contest is administered by the Cambridge Center for Behavioral Studies, with assistance from The Computer Museum in Boston. It was inspired by a paper published in 1950 by the brilliant English mathematician Alan Turing, one of the fathers of the modern computer. The tests so far are restricted, requiring computers to be conversant on only one topic.

The 1993 contest will be held on Tuesday, September 21. The deadline for receipt of 1993 submissions is August 1. For more information and contest requirements, write to Dr. Robert Epstein, Contest Director, Cambridge Center for Behavioral Studies, 11 Waterhouse Street, Cambridge, MA 02138, or call 617 876-2716.

ANSWERS, HINTS & SOLUTIONS

Math

M76

This problem has a lot of different solutions that use simple and well-known inequalities for positive numbers:

$$\frac{x+y}{2} \ge \sqrt{xy}, \quad \frac{x}{y}+\frac{y}{x} \ge 2,$$

and suchlike. I'll give four of them: choose the one most to your liking.

(1) Rewrite the condition $ab > a + b$ as

$$(a-1)(b-1) > 1.$$

Both factors must be positive, because otherwise the inequalities $0 < a < 1$ and $0 < b < 1$ would imply $(a-1)(b-1) < 1$. Using the inequality between the arithmetic mean and the geometric mean of $a - 1$ and $b - 1$ (see, for instance, the Kaleidoscope in the last issue), we get

$$a-1+b-1 \ge 2\sqrt{(a-1)(b-1)} > 2,$$

and so $a + b > 4$.

(2) Another rewording of the condition is that the *harmonic mean* h of a and b is greater than 2:

$$h = \left(\frac{a^{-1}+b^{-1}}{2}\right)^{-1} = \frac{2ab}{a+b} > 2.$$

But the arithmetic mean is not less than the harmonic mean (see again the last Kaleidoscope, or simply note that $(a + b)/2 \ge 2ab/(a + b)$ is equivalent to $(a - b)^2 \ge 0$), so

$$\frac{a+b}{2} \ge h > 2,$$

or $a + b > 4$.

(3) Dividing the given inequality by a and by b, we get

$$a > \frac{a}{b}+1, \quad b > \frac{b}{a}+1,$$

and, adding these up,

$$a+b > \frac{a}{b}+\frac{b}{a}+2 \ge 4.$$

(4) Squaring the arithmetic mean–geometric mean inequality and using the condition, we obtain

$$\frac{(a+b)^2}{4} \ge ab > a+b.$$

It remains to divide the left and right sides by $a + b$ (which is positive) and multiply by 4 to arrive again at $a + b > 4$.

M77

Note that an endpoint of any straight segment of the rook's route is either the starting point of the entire route, its endpoint, or a point where it turns. So it suffices to prove that the route has at least n *disjoint* straight segments (these have $2n$ endpoints, 2 of which at most can be the endpoints of the route, so the remaining $2n - 2$ must be turning points).

If each of n horizontal rows of the chessboard contains a segment of the rook's route, then these are just the required n disjoint segments. If there is a row with no segments on it, then each of the n squares of this row must be crossed by a vertical segment of the route, and so we get n disjoint segments again.

The simplest conceivable route—all the way from left to right along the bottom row, one step up, all the way back along the second row, one step up, and so on—provides an example of exactly $2n - 2$ turns.

M78

The trick is that one of the two numbers a_k and b_k is always greater than n, and the other is not. Indeed, assume, for instance, that both a_k and b_k are not greater than n. Then all the numbers $a_1, a_2, \ldots, a_k$ and $b_k, b_{k+1}, \ldots, b_n$ are not greater than n, which is impossible because we'd have $k + (n - k + 1) = n + 1$ positive integers not exceeding n. In the same way we can show that a_k and b_k can't both be greater than n.

So every term of the sum in question is the difference of a number greater than n and a number not exceeding n. Clearing the absolute values and rearranging terms, we obtain

$$\begin{aligned}|a_1 - b_1| + \ldots + |a_n - b_n| &= [(n+1) + (n+2) + \ldots + 2n] \\ &\quad - (1 + 2 + \ldots + n) \\ &= (n+1) - 1 + (n+2) - 2 \\ &\quad + \ldots + 2n - n \\ &= n^2.\end{aligned}$$

M79

Let's draw the tangent to the incircle of triangle ABD parallel to AB (fig. 1); M is the point where it cuts the base AC. It will suffice to prove

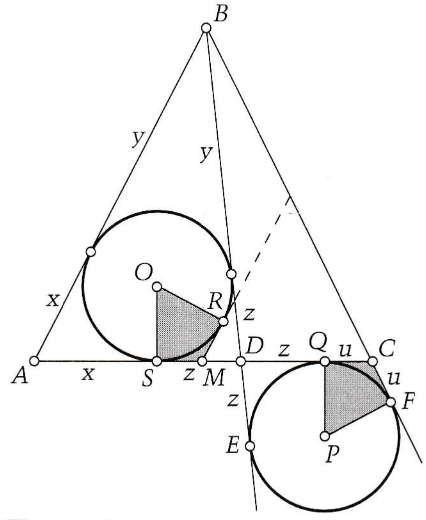


Figure 1

that the tangent runs along the triangle's midline, or that M is the midpoint of AC, because the distance between the midline and the side AB is half the height h.

Now angle AMR = angle DCF (each is supplementary to a base angle of the original triangle). Since the circles are congruent, it follows that quadrilaterals $OSMR$ and $PQCF$ are congruent. Therefore, $AM = x + u$, where x is the tangent length from A to the left circle and u is the tangent length from C to the right circle (see the figure). So we must prove that $x + u = AC/2$.

Using the fact that tangents drawn to a circle from the same point are congruent, we can label congruent segments as in the figure with the letters x, y, z, u (all four tangents to both circles drawn from D can be marked the same way, because the circles are congruent). Now we have

$$\begin{aligned} x + u &= (AB - y) + (BF - BC) \\ &= BF - y \\ &= BE - y \\ &= 2z \\ &= AC - (x + u) \end{aligned}$$

($AB = BC$ by the statement of the problem, $BE = BF$ since they are tangents to a circle from the same point), so $2(x + u) = AC$, and we're done. (N. Vasilyev, I. Sharygin)

M80

(a) We'll prove a more general fact: the number of L-tiles in any tiling of a $(2n - 1) \times (2n - 1)$ square with tiles of the three given types is not less than $4n - 1$ for any $n \geq 4$. In particular, for $n = 50$ this yields the desired estimate of 199.

Let's color the big square in four colors 1, 2, 3, 4 as shown in figure 2. Clearly, no matter how we fill this square with our tiles, any of the bigger tiles will always cover four unit squares of different colors. So all these tiles will cover an equal number of unit squares of each color. However, the total numbers of unit squares of different colors are different, and this is what underlies the solution.

If a is the number of L-tiles and b the number of all the other tiles, then

1	2	1	2	1	2	1
3	4	3	4	3	4	3
1	2	1	2	1	2	1
3	4	3	4	3	4	3
1	2	1	2	1	2	1
3	4	3	4	3	4	3
1	2	1	2	1	2	1

Figure 2

$3a + 4b = (2n - 1)^2$. On the other hand, each tile contains not more than one square colored 1, and the total number of such squares is n^2, so $a + b \geq n^2$. Therefore,

$$\begin{aligned} 4a &\geq 4n^2 - 4b \\ &= 4n^2 - (2n - 1)^2 + 3a \\ &= 4n - 1 + 3a, \end{aligned}$$

or $a \geq 4n - 1$, completing the proof.

(b) Figure 3 presents, for the case $n = 4$, two ways to tile a 7×7 square with one 4-square tile of either type and $15 = (49 - 4)/3 = 4n - 1$ L-tiles (we don't show the obvious division of a 2×3 rectangle into two L-tiles). Figure 4 shows how a tiling of a $(2n - 1) \times (2n - 1)$ square can be extended to a tiling of a $(2n + 1) \times (2n + 1)$ square by adding exactly 4 L-tiles (forming two 2×3 rectangles) and a number of small squares. So if the initial tiling contained $4n - 1$ L-tiles, the extended tiling contains $4n + 3 = 4(n + 1) - 1$ L-tiles. In particular, for $n = 50$ (a 99×99 square) we thus obtain a tiling with 199 squares.

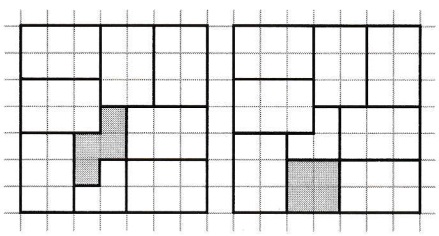


Figure 3

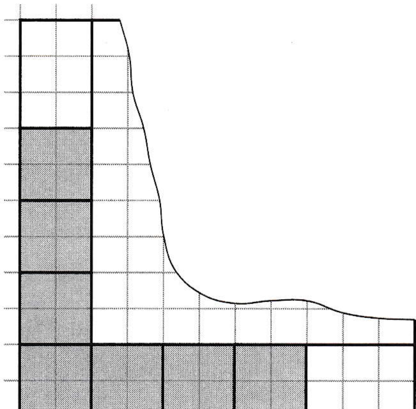


Figure 4

For $n = 2$ or 3, $3(4n - 1) > (2n - 1)^2$, which means that the number of squares covered by $4n - 1$ L-tiles is greater than the total number of squares in the figure. This shows that the number of L-tiles in these cases must be less than $4n - 1$.

Physics

P76

Let's hang the spring by one end. The tension will vary along the length of the spring. We'll use the following notation: M is the mass of the spring, N is the total number of windings, k is the spring constant for the entire spring, and n is the number of windings from the lower end of the spring to the point being considered. At this point the tension is equal to

$$T_n = \frac{Mg \cdot n}{N}.$$

The spring constant for one winding of the spring is N times that of the entire spring. Then the increase in the length of this winding will be

$$\Delta l_n = \frac{T_n}{kN} = \frac{Mg \cdot n}{k \cdot N^2}.$$

Summing the change in the lengths of all the windings, we get

$$\begin{aligned} L - L_0 = \sum \Delta l_n &= \frac{Mg \cdot N \cdot (N+1)}{2 \cdot k \cdot N^2} \\ &\cong \frac{Mg}{2k} = \frac{L_0}{2}. \end{aligned}$$

If the spring is put into a vessel, the result will be almost the same, but instead of lengthening the spring will decrease in length:

$$L_0 - L = \frac{Mg}{2k}.$$

If the spring is submerged completely, then the change of its length can be estimated by taking into account the reduction in its apparent weight due to the buoyant force:

$$L_0 - L = \frac{M^{\cdot} \cdot g}{2k} = Mg\frac{\rho - \rho_0}{2k \cdot \rho},$$

where ρ is the density of the spring and ρ_0 is the density of water. This means that the water should reach to a height

$$L = L_0 \frac{1 + \rho_0/\rho}{2} = L_0 \frac{1 + 1/r}{2}.$$

This formula is valid only when the density of water is less than the density of the spring; otherwise the spring will float.

P77

After the first collision, the puck of mass $2M$ ("$2M$-puck") will have a velocity $V_0/2$ (according to the law of conservation of momentum). To determine the velocities after the next collision, it's convenient to use a coordinate system that moves to the right at the velocity $V_0/2$. In this system the $2M$-puck is stationary and is struck by the puck of mass M ("M-puck") once again (but now the velocity of the moving puck is half its previous velocity). It's clear that the moving puck will stop (in this system!); the velocity of the $2M$-puck will be directed to the left and will equal $V_0/4$. In the laboratory coordinate system, the $2M$-puck moves to the right with a velocity $V_0/4$, and the velocity of the M-puck is $V_0/2$.

This calculation can be done without using a moving coordinate system. To determine the velocities directly, one must take into consideration that in a nonelastic collision a certain fraction of the maximum elastic energy of the pucks is converted into heat. The calculation is pretty messy, though.

P78

On the graph we can clearly see when the scraps of tin were put into the oven: the temperature dropped quickly to approximately 230°C—the melting temperature of tin. The temperature didn't change for almost 12 minutes, since all the heat absorbed by the tin was used to melt the tin. To determine the latent heat of melting, one needs to calculate what fraction of the heating element's power is dissipated to the surroundings and what fraction is used to melt the tin. Here the first part of the graph will be of help to us: at low oven temperatures there is little heat exchange with the surroundings—that is, we can assume that all the power P goes into heating the oven. On the graph we can see that initially the temperature increases approximately 1 degree per minute (it's convenient to draw a tangent to the curve and measure its slope). When the temperature reaches 230°C (which corresponds to the melting temperature), the slope of the curve is about 0.25 degree per minute. This means that at this temperature only 1/4 of the power remains inside the oven and the remaining 3/4 is dissipated to the surroundings. Therefore, the latent heat is

$$C = \frac{0.25 \cdot P \cdot t}{M}$$
$$= \frac{0.25 \cdot 20\ \text{W} \cdot 12\ \text{min} \cdot 60\ \text{s/min}}{50\ \text{g}}$$
$$\cong 70\ \text{J/g}.$$

In calculating the thermal balance we didn't take into account the heat exchange at the moment the oven is loaded with the tin scraps—the graph is rather rough and the precision of our calculations can't be that high. Nevertheless, we can estimate the heat capacity of the oven and convince ourselves that it isn't high (we can compare the slope of the heating curve at the melting temperature when the oven is empty and when the tin has finished melting).

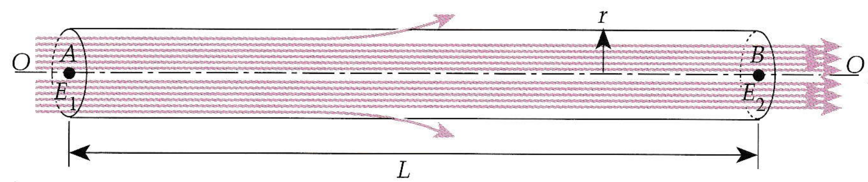


Figure 5

P79

If we draw the lines of force for the electrostatic field correctly (fig. 5), the number of lines crossing a unit area of a surface is proportional to the perpendicular component of the field at the surface. Consider a long thin cylinder along the axis of the field with a length L = 1 m and a radius r = 1 cm. It's on the walls of the cylindrical surface that we want to determine the perpendicular component of the electric field. The number of lines of force entering the left end of the cylinder is $N_1 = kE_1\pi r^2$, and the number of lines of force leaving the right end is $N_2 = kE_2\pi r^2$. Since the number of lines entering the cylinder must be equal to the number leaving the cylinder, the number of lines N passing through the walls must be given by

$$N = N_1 - N_2.$$

Therefore,

$$kE2\pi rL = k\pi r^2(E_1 - E_2),$$

and

$$E = \frac{r}{2L}(E_1 - E_2) = 0.005\ \text{V/m}.$$

P80

If a converging lens happens to be the first in the beam's path (fig. 6a), then after passing through two lenses the light

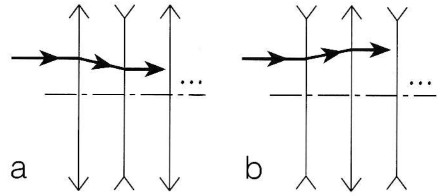


Figure 6

rays will be parallel to the axis again; but distances from the axis will be cut in half. Thus, after passing through N pairs of lenses the beam's diameter will be

$$d = \frac{D}{2^N}.$$

If the first lens is diverging (fig. 6b), the rays will move away from the axis and the beam's diameter will be

$$d = D \cdot 2^N.$$

It's clear that the diameter of the beam can't become larger than the diameter of the lenses. There are limitations in the first case as well: the light beam can't become arbitrarily small, because the wave properties of light prevent it.

Brainteasers

B76

Let a be the number of blond persons with blue eyes, b the number of all blonds, c the number of all blue-eyed people, and n the entire population. Then by the statement of the problem, $a/c > b/n$, or $an > bc$, or $a/b > c/n$. This means that the answer to the question is yes.

B77

The only kind of triangle that satisfies the condition is an isosceles right triangle. To prove this, consider a triangle ABC whose side $a = BC$ is not longer than the corresponding height h_a and $b = CA \le h_b$. Obviously no height is longer than any of the sides drawn from the same vertex of a triangle. So we can write the following string of inequalities:

$$a \le h_a \le b \le h_b \le a,$$

which means that $a = b = h_a = h_b$. But the equality $a = h_b$ is possible if and only if a coincides with h_b—that is, a is perpendicular to b.

B78

A drop falls off the end of a pipette when the surface tension no longer counterbalances gravity. When the temperature of the water increases, the coefficient of surface tension and the force of surface tension decrease. The decrease is noticeable—about 20% when the temperature increases from 20°C to 100°C. Thus, the weight of each hot drop is less than that of a cold drop, and the number of hot drops is therefore greater.

There is another process that occurs during heating: the density of water decreases because of expansion. Generally speaking, this phenomenon plays a contrary role here. But the coefficient of thermal expansion for water is small, so this effect is much weaker than that due to surface tension and is practically absent in this problem.

B79

TWO = 426. The last digit of the number in question, O, is such that O^2 ends in O. So it can be equal to 0, 1, 5, or 6. It's not zero, because the product TWO · O isn't zero (for a similar reason, W and T aren't zero either). It's not 1, because this product is a 4-digit number. It's not 5, because the products O · O, O · W, and O · T end in three different numbers. So O = 6. Now we know that 6 · W ends in W and 6 · T ends in T. An easy check shows that T and W can equal 2, 4, or 8. Since TWO · W and TWO · T are a 3-digit and a 4-digit number, respectively, the only possibility left is T = 4, W = 2.

B80

The answer is shown in figure 7.

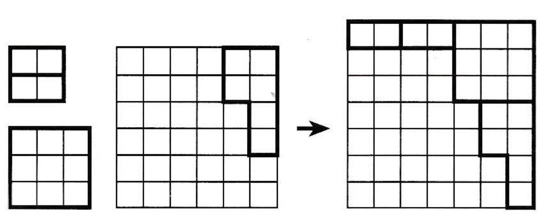


Figure 7

Kaleidoscope

1. The usual, unthinking answer is that the cover costs 50 kopeks. But then the price of the book would be 2 rubles—that is, only 1 ruble 50 kopeks more than the cover! The right answer is that the cover costs 25 kopeks, and the price of the book is 2 rubles 25 kopeks.

2. It may seem rather strange, but nevertheless Ivanova will still pay less than all the other customers paid before January 1. She'll receive a 20% reduction on a price that increased 20%—in other words, she'll receive a reduction of 20% of 120%, so she'll pay not 100% but only 96% of the former price of a book. She'll get a three-ruble book for 2 rubles 88 kopeks.

3. How could Romans know, when they supposedly coined that piece of money in "53 B.C.," that Christ would be born 53 years later?

4. The customer guessed wrong. We'll assume that the asparagus are all the same length, so the amount of the vegetable in a bundle depends on the cross-sectional area of the bundle, not on the circumference. Thus, a bundle with double the circumference contains not two but four times the asparagus as the thin bundle. She should either have paid half the usual sum or demanded not two but four thin bundles of asparagus.

5. We can find the average earnings of the seven workers by redistributing their wages among them. This is easy to do. We just take the three "extra" rubles away from the cabinetmaker. By the statement of the problem, this leaves him with an amount equal to the desired average. To make the carpenters just as rich, we need only distribute the extra three rubles in equal shares among them. Therefore, we need to add 50 kopeks to the 20 rubles earned by each carpenter—this is the average earnings of each of the seven. From this we find that cabinetmaker earned 20 rubles 50 kopeks plus 3 rubles—that is, 23 rubles 50 kopeks.

6. There are infinitely many pairs of such numbers. Here are only a few examples:

$$4 + 1\tfrac{1}{3} = 5\tfrac{1}{3},$$

$$4 \times 1\tfrac{1}{3} = 5\tfrac{1}{3};$$

$$5+1\tfrac{1}{4}=6\tfrac{1}{4},$$
$$5\times1\tfrac{1}{4}=6\tfrac{1}{4};$$

$$11+1.1=12.1,$$
$$11\times1.1=12.1;$$

$$9+1\tfrac{1}{8}=10\tfrac{1}{8},$$
$$9\times1\tfrac{1}{8}=10\tfrac{1}{8};$$

$$21+1\tfrac{1}{20}=22\tfrac{1}{20},$$
$$21\times1\tfrac{1}{20}=22\tfrac{1}{20};$$

$$101+1.01=102.01,$$
$$101\times1.01=102.01.$$

In general, if a and b are two such numbers, then

$$\frac{a+b}{ab}=\frac{1}{a}+\frac{1}{b}=1.$$

So if we start with any number a and let $b=1/(1-1/a)=a/(a-1)$, we'll arrive at a pair of numbers with the given property.

7. The explanation is quite simple, as you can see from figure 8. The protuberances and grooves aren't arranged in the shape of a cross, as it seemed when we saw the cube as put together, but diagonally. Such protuberances can easily be inserted into the corresponding grooves.

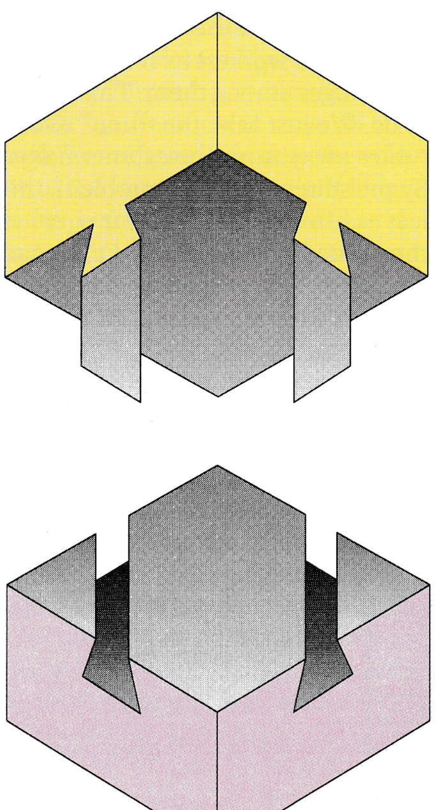


Figure 8

8. The winning strategy can easily be found if we analyze the game from the end backward. We then see that if after your next-to-last move there are 5 matches on the table, you're assured of victory: the other player can't take more than 4 matches, and so you can take all the remaining matches. But how do you arrange it so that you leave 5 matches after your next-to-last move? To this end you have to leave 10 matches after your previous move. In this case your opponent can't leave you less than 6 matches, and you can always leave 5 matches on your next turn. But how can you manage to force your opponent to select from 10 matches? Why, you have to leave 15 matches after your previous turn. So, subtracting 5 matches each time, we learn that 20 matches must remain on the table, and 25 before that, and, finally, 30 matches after the first round, which means you must take 2 matches in your first turn.

So this is the winning strategy: first, take 2 matches; then, after your opponent has taken some matches, take enough matches to leave 25; after your next turn you leave 20, then 15, 10, and, finally, 5 matches. You always get the last match. (For more on this style of reasoning, see "Jewels in the Crown," on mathematical induction, in the July/August 1992 issue of *Quantum*.)

9. If the player who takes the last match loses, then after your next-to-last move you must leave 6 matches on the table. In this case your opponent can't leave you less than 2 or more than 5 matches, so in your next turn you can leave the last match. But how can you manage to leave 6 matches? You have to leave 11 matches in your previous turn, and before that, 16, 21, 26, and 31 matches.

So, when you begin the game, you take only 1 match, and then you leave 26, 21, 16, 11, and 6 matches in your subsequent turns. The last match unfailingly goes to your opponent.

10. It's a bit harder to find the winning strategy for this game than for the game of 32. We need to take the following two points into account.

(1) If before the end of the game you have an *odd* number of matches, then you have to leave your opponent 5 matches to ensure your victory. After the next move your opponent will leave 1, 2, 3, or 4 matches. If 4 matches are left, then you take 3 matches and win the game. If 3 matches are left, then you take them all and win the game. If 2 matches are left, then you take 1 match and win the game.

(2) If before the end of the game you have an *even* number of matches, then you have to leave to your opponent 6 or 7 matches. Let's work through the rest of the game. If your opponent leaves 6 matches after the next turn, then you take 1 match and the number of your matches becomes odd. In this case you leave 5 matches, and your opponent is doomed to defeat. If your opponent leaves not 6 but 5 matches, then you take 4 matches and win the game. If your opponent leaves 4 matches, then you take them all and win the game; if 3 matches, you take 2 and win the game. Finally, if your opponent leaves 2 matches, then you win the game. Your opponent can't leave fewer than 2 matches.

Now it's not hard to lay out the winning strategy. If you have an *odd* number of matches, you must leave a number of matches that is 1 less than a multiple of 6—that is, 5, 11, 17, and 23. If you have an *even* number of matches, you must leave a number of matches that is a multiple of 6 or 1 more than that—6 or 7, 12 or 13, 18 or 19, 24 or 25. Zero can be considered an even number. From the original 27 matches you have to take 2 or 3 matches, then proceed according to plan. You cannot lose with this strategy. Just don't let your opponent take the initiative.

11. If the player who holds an odd number of matches at the end wins, then you must proceed as follows. If you have an *even* number of matches, you leave a number of matches that is one shy of a multiple of 6. If you have an *odd* number of matches, you leave a multiple of 6 or a multiple of

6 plus 1. This will invariably bring you to victory. When you begin the game, you have 0 matches (an even number). That's why you take 4 matches on your first move and leave 23 to your opponent.

12. From the statement of the problem we know that, first, the mass of the bottle plus the mass of the gasoline equals 1,000 g; second, because the acid is twice as dense as gasoline, the mass of the bottle plus twice the mass of the gasoline equals 1,600 g. It's clear that the difference in mass (600 g) is the mass of the bottle's volume of gasoline. But the mass of the bottle with gasoline is 1,000 g; therefore, the bottle's mass is 1,000 g – 600 g = 400 g. Indeed, the mass of the acid (1,600 g – 400 g = 1,200 g) is twice that of the gasoline.

13. The thickness of a layer of cherry flesh is equal to the diameter of a cherry pit. Therefore, a cherry's diameter is three times that of its pit. A cherry's volume is $3 \times 3 \times 3 = 27$ times that of its pit. So the volume of cherry flesh is $27 - 1 = 26$ times the volume of a cherry pit.

14. A one-kilogram model of the Eiffel Tower is far taller than a drinking glass—you may be surprised to learn that it's 1½ meters tall! In fact, the volume of the model is to the original as 1 kg is to 8,000,000 kg. So the model's height is to the height of the actual Eiffel Tower as 1 is to a number that, when cubed, equals 8,000,000. This number is 200. Dividing the height of the Eiffel Tower (300 m) by 200, we get 1½ m. At first blush this result may seem strange: a 1½-m piece of iron whose mass is only 1 kg! But here's the rub: in addition to being tremendously tall, the Eiffel Tower is remarkably airy, and so it's relatively unmassive.

15. The second boat finished second because it sailed at 24 mph for *less time* than it did at 16 mph. This boat sailed at a speed of 24 mph for 24 mi/24 mph = 1 hr, and at a speed 16 mph for 24 mi/16 mph = 1½ hr. So on the first leg it lost more time than it gained on the second leg.

16. Traveling downstream, the rower covers 1/2 mile per minute; going upstream, only 1/12 mile per minute. The first speed includes the velocity of the current, which is subtracted from the speed upstream. Therefore, 1/2 + 1/12—that is, 7/12 mile—divided by 2—that is, 7/24 mile per minute—is the rower's speed in still water. So in still water the rower will cover 10 miles in

$$\frac{10\ \text{mi}}{7/24\ \text{mi/min}} = 34\tfrac{2}{7}\ \text{min}.$$

People usually answer that the rower covers the 10 miles in the same time as on the river, arguing that the loss in speed upstream is compensated by a gain in speed downstream. But that's faulty reasoning (see the previous problem).

17. Coming downstream, the steamboat covers 1 mile in 3 minutes; going upstream, it covers 1 mile in 4 minutes. In the first instance the steamboat gains 1 minute with each mile. Over the entire distance the steamboat gains 5 hours, or 300 minutes. Therefore, the distance from N-burg to X-ville is 300 miles. Indeed,

$$\frac{300\ \text{mi}}{15\ \text{mph}} - \frac{300\ \text{mi}}{20\ \text{mph}} = 20\ \text{hr} - 15\ \text{hr}$$
$$= 5\ \text{hr}.$$

18. The square's side length must be one tenth of 100 km. A square with a 10-km side contains $10{,}000 \times 10{,}000 = 100{,}000{,}000$ square meters. If each square meter has room for 20 people, then such a square contains $20 \times 100{,}000{,}000 = 2{,}000{,}000{,}000$ people—that is, more than the entire population of the Earth in 1924 (1,800,000,000). Consequently, all of humanity could be placed inside a square with a side length of 10 km.

19. This kind of check is insufficient. A quadrilateral can satisfy it without being a square. Figure 9a shows an example of such a quadrilateral. It has equal sides but not right angles. In geometry such a figure is called a rhombus. Every square is a rhombus, but not every rhombus is a square.

20. This check is as unreliable as the previous one. The diagonals of a square, of course, are equal, but not every quadrilateral with equal diagonals is a square (see figure 9b). The parquet-makers have to use both checks simultaneously. Then they could be sure that they've cut a real square: any rhombus with equal diagonals is a square.

21. Such a check can only show that the quadrilateral under consideration has right angles—that is, it's a rectangle (see figure 9c). But it says nothing about the equality of its sides.

Coding

1. There are p^n words. (See, for instance, "Combinatorics–polynomials–probability" in this issue.)

2. The distances are 4, 2, and 2, respectively.

3. If $d(C) \geq 2r + 1$, there cannot be a word whose distance to two different words is $\leq r$. If there were, the triangle inequality would guarantee that these two words would be $2r$ or fewer units apart, which is impossible. So decoding is always unique,

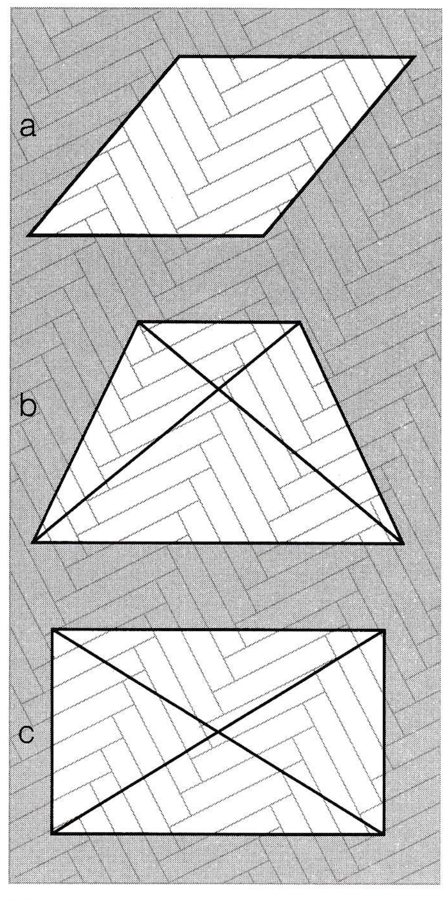


Figure 9

provided that the number of errors doesn't exceed r.

4. The numbers $a_1, a_2, \ldots, a_6$ should be relatively prime to 10—that is, they shouldn't be divisible by 2 or 5.

5. The volumes are $(p-1)n$ and $(p-1)^2n(n-1)/2$.

6. Substitution of the given word into the equation yields the first column (three ones) of the matrix. So the error was made in the first character, and the correct word is 0010101.

Black holes

1. If your mass is 50 kg, then $R_S = 7.4 \times 10^{-27}$ m, which is *much* smaller than an atomic nucleus (typical nuclear radii are of the order of 10^{-15} m). As for the Schwarzschild radius of the Milky Way Galaxy, $M_{Sun} = 2 \times 10^{30}$ kg, so $M_{Galaxy} = 2 \times 10^{41}$ kg. We then find $R_S = 2.96 \times 10^{14}$ m. This is about 2,000 astronomical units (1 AU equals the average distance from the Earth to the Sun) or about 0.03 light-year.

2. $\rho_S = 1.8 \times 10^{-3}$ kg/m^3 = 1.8×10^{-6} g/cm^3. This is about one thousandth the density of air.

3. 1 atm = 1.013×10^5 N/m^2 = 1.013×10^5 kg/(m · sec^2); the density of water = ρ = 1 g/cm^3 = 1,000 kg/m^3. We then find $p/\rho c^2 = 1.13 \times 10^{-15}$, which is certainly much less than one.

Combinatorics

1. 11! = 39,916,800; 12! = 479,001,600.

2. The number of zeros at the end of $n!$ is equal to the number of factors of 5 in $n!$. For $n = 24$ this number is still 4, but for $n = 25$ it makes a double leap and becomes equal to 6. So $n!$ never ends in exactly 5 zeros, and the smallest n such that $n!$ ends in 6 zeros is $n = 25$.

3. The given formula is equivalent to $(n + 1)! = n! + n!xn = (n + 1)n!$, which is formula (1) in the article.

4. Using the formula from problem 3, the given sum can be rewritten as

$$(2! - 1!) + (3! - 2!) + (4! - 3!) + \ldots + (n + 1)! - n!.$$

This sum "telescopes": the second term of each group of two cancels with the first term of the previous group. After all the cancellations, the sum reduces to $(n + 1)! - 1! = (n + 1)! - 1$.

5. We leave this verification to our readers.

6. Making the suggested substitution, we find that $(k!)! = (k! - 1)!(k!)$, which is true for any natural number k. Letting $x = k! - 1$, $y = k$, $z = k!$ gives us triples of numbers satisfying the required condition. Setting $k = 1$ or 2 gives the trivial solutions (0, 1, 1), (1, 2, 2). Letting $k = 3$, we obtain (5, 3, 6); and if $k = 4$, we obtain (23, 4, 24).

7. (a) 8!; (b) SPANIEL.

8. 123456789 × 8 = 987654312.

9. (a) 6! = 720, assuming that rotations of a given lettering are considered distinct from that lettering; (b) 12 (there are 6 ways to choose vertex A and 2 ways to choose the direction of labeling, or vertex B).

10. 8! (the first rook may sit in any of 8 squares of file a, the second—in any of 7 squares of file b that do not lie in the row where the first rook sits, and so on).

11. (a)7!/3!; (b) 8!/(2!)^2.

12. We must find anagrams for the "word" BBBPPOO, where each letter is the initial of one of the fruits. There are 7!/3!2!2! = 35 of these.

13. READ QUANTUM EVERY DAY.

14. To get the correct answer, we divide the number of permutations (with repetitions) of the given colors, $N = 11!/5!3!2!1!$, by $11 \cdot 2 = 22$, because when we compare the order of colors on two necklaces, we can start with any of the 11 beads and "read" the colors in either of two directions (clockwise or counterclockwise). So in the number 11! every way of making a necklace is counted 22 times. The answer is $N/22$ = 1,260.

15. Substituting 1's for each letter, each term in the sum will have a value of 1, so the numerical value of the sum will be the number of terms (before simplification). Here, this number is 24.

16. The largest coefficient will be the one for which we have the largest number of ways to choose the letters that make up the term. For (a) this is $5!/(1!)^5$ = 120 (the coefficient of $abcde$). For (b) it is 5!/2! = 60 (the coefficient of a^2bcd, ab^2cd, and similar monomials).

17. $K_1 = 1$, $K_2 = 4$, $K_3 = 6$, $K_4 = 12$, $K_5 = 24$; the numbers of terms are equal to 5, 20, 10, 30, and 5, respectively.

18. (a) 3! = 6. (b) The number of pieces of each sort is related to the number multiplying the probabilities of each color of pepper. In fact, if the green, red, and yellow peppers occurred in the ratio 3 : 2 : 1, the number of pieces of each sort would be identical to the coefficient of the corresponding probability. Each piece is a box whose dimensions are 1 cm, 2 cm, or 3 cm; if n_1 of the dimensions are equal to 1, n_2 to 2, n_3 to 3 ($n_i \geq 0$; $n_1 + n_2 + n_3 = 3$), then the volume of the box equals $1^{n_1}2^{n_2}3^{n_3}$, and the number of such boxes equals $3!/n_1!n_2!n_3!$.

Stomachion

(See the Toy Store in the January/February issue of *Quantum*)

1. See figure 10.

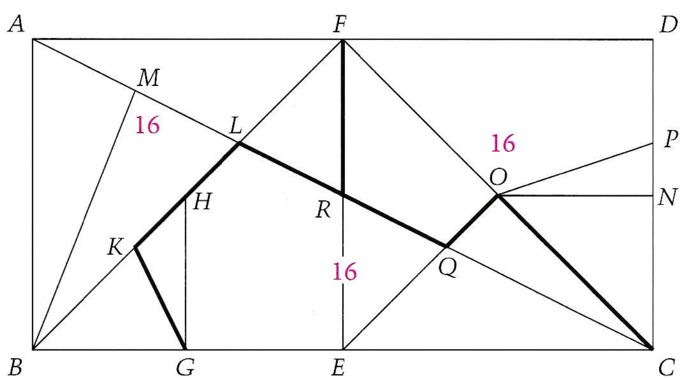


Figure 10

2. See figure 11.
3. See figure 12.

Toy Store

The pyramid at the bottom of figure 13 corresponds to the top drawing in figure 1 in the article, the one above it to the other drawing in figure 1; the pyramid at the top of figure 13 corresponds to the quadrilateral pyramid in figure 3 in the article.

The two tetrahedrons in question are developed from the same "blueprint," but it's as if one of them is turned inside out to get the other—the inside surface of one tetrahedron corresponds to the outside of the other.

The unknown number of spots on the face of the die is 3 (for the "clockwise" die) or 4 (for the "counterclockwise" die).

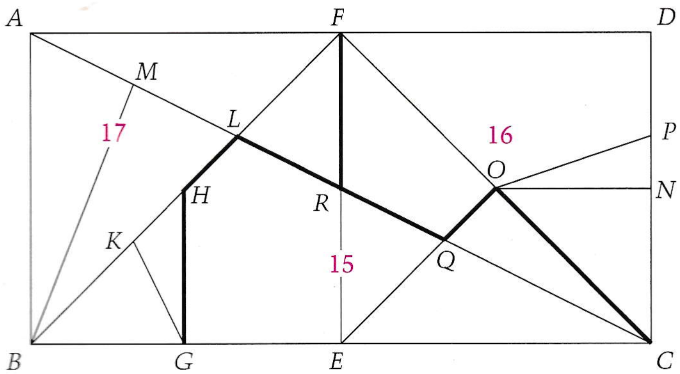


Figure 11

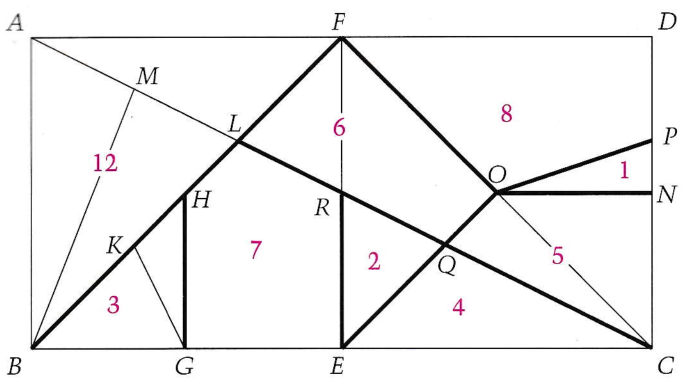


Figure 12

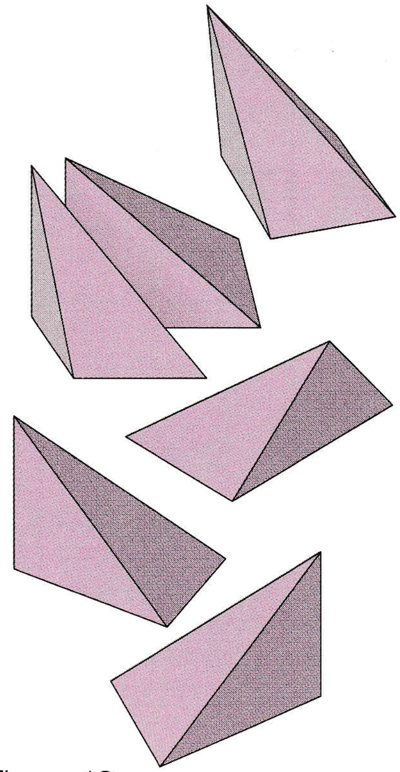


Figure 13

Index of Advertisers

American Association of Physics Teachers	23
Beloit College	29
Clarke College	9
Drexel University	25
Embry-Riddle Aeronautical University	Cover 4
Independent Educational Services	29

Corrections

Vol. 3, No. 3:

Gallery Q: Titian reworked Bellini's painting fifteen *years* later.

p. 19: Due to an editorial oversight, the source of the proof of the Euclid–Euler Theorem in "A Magnificent Obsession" was omitted. It was adopted from V. Klee and S. Wagon, *Old and New Unsolved Problems in Plane Geometry and Number Theory* (Washington: Mathematical Association of America, 1991), p. 179.

p. 27, col. 2, footnote: A production/software glitch caused an ellipsis (...) in the second displayed equation to be printed as the letter K.

p. 60, col. 1, solution 6: The same glitch caused five ellipses in the displayed equations to be printed as the letter L.

p. 32, col. 3, last line: *for* $\sqrt{(a+b)/2}$ *read* $\sqrt{(a^2+b^2)/2}$.

p. 60, col. 1, l. 19: *for* AA_0 *read* A_0.

p. 60, col. 2, l. 13: *for* $1/n$ *read* $1/(n+1)$.

Wacky pyramids

Trisecting the stately cube

by Yakov Smorodinsky

YOU SEE, I WANTED TO PLAY with a simple but nifty little puzzle that I had come across in one of Martin Gardner's articles.

So I went to a machine shop and asked them to make six pieces according to the two drawings in figure 1—three of each kind. The machinist I was dealing with was almost offended—he thought I was trying to make a fool of him. But joking was the furthest thing from my mind. The pieces were drawn according to the strictest rules of mechanical drawing (see figure 2 for another example).

Without reading any further, try to draw these pieces or at least visualize them.

In an attempt to simplify the task, I decided to have three pieces made (from another drawing—figure 3) instead of six. But since this drawing also proved hard to make out immediately, I chose not to try the machinist's patience any longer and to prepare paper models of the required pieces myself. It turned out that the paper versions were good enough to use. You can construct these pieces using the cutouts in figure 4.

Can you explain how two cutouts yield three different pieces?

I've talked enough about the pieces—now let's talk about the puzzle. The pieces made according to figure 1 are tetrahedrons, two of whose faces are congruent 45°–45° right triangles having one leg in common and fixed at a right angle. Such an arrangement of triangles can be achieved in two different ways that mirror-reflect each other.

The pieces made according to figure 3 are quadrilateral pyramids with square bases, whose heights are congruent to a side of the base and fall on one of the base vertices. (All the pieces are pictured in figure 13 on page 63.) The problem is to make a cube out of six of the tetrahedrons or three of the quadrilateral pyramids. This isn't difficult, but it's instructive—try it!

Now hold a cube assembled from three pyramids of the second type between your thumb and forefinger

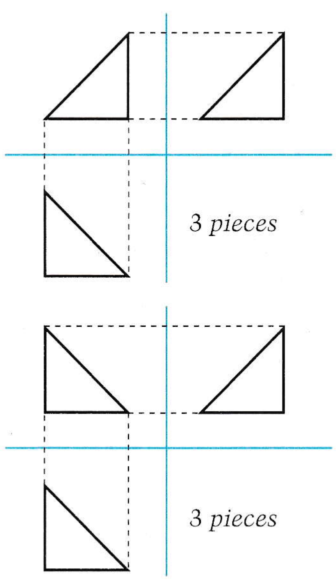


Figure 1

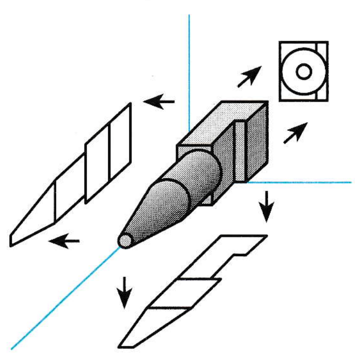


Figure 2

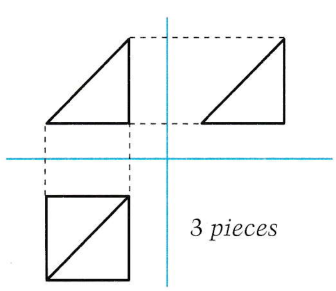


Figure 3

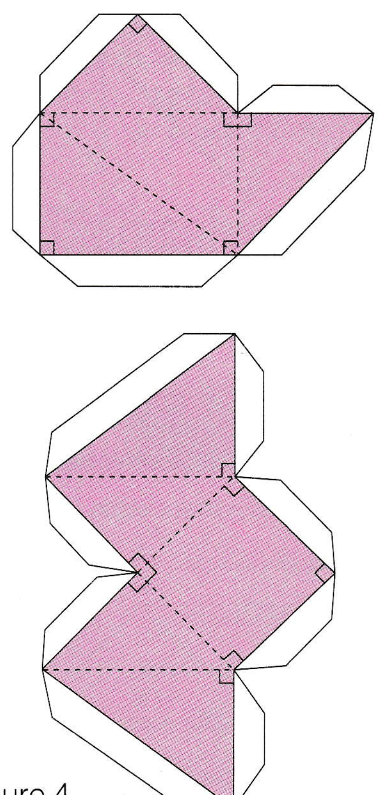


Figure 4

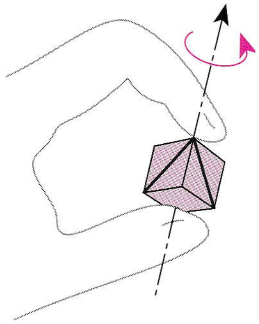


Figure 5

(fig. 5) and carefully rotate it about the axis thus arranged—the diagonal of the cube formed by the longest edges of the three pyramids. Under 120° and 240° rotations each pyramid will take the place of its neighbor. In a case like this we say that the figure (our assembly of pyramids) has a symmetry axis of the third order. It's easy to imagine an apple sliced into three equal parts, but you've probably never seen the "trisection" of a cube.

Each piece in this trisection (that is, each of our quadrilateral pyramids) has a plane of symmetry—it can be cut along a plane perpendicular to the base of the pyramid so that the base is dissected into two congruent triangles, and the entire pyramid into two triangular pyramids, which are just the pieces from figure 1. These pyramids are mirror images of each other, so they're congruent, but they're not identical: you can never, remaining in our three-dimensional space, move one of them so that it exactly fits the place that was occupied by the other.

It's curious that mirror images in space are usually rather hard to tell apart. Ask your friends whether our two mirror tetrahedrons, set a certain distance apart, are exactly the same. You'll see that this question won't be a simple one for many of them.

Here's another way to test your sense of orientation. You've seen dice many times, no doubt, and know that the sum of the spots on opposite faces of a die is seven. So, looking at one face, you can tell how many spots there are on the opposite face. *How many spots should there be instead of the question mark in figure 6?* After you get the answer, compare it with your die.

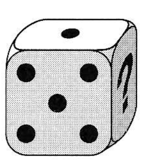


Figure 6

In answering the last question, perhaps you noticed that there are two mirror-symmetrical ways to arrange the spots on a die (so that the sums on opposite faces are all equal). This means there are two types of dice: equal sums ensure that faces 1, 2, and 3 of a regular die always have a common vertex, and they can be arranged around this vertex clockwise or counterclockwise (fig. 7). This uniquely determines the spots on the remaining faces. The dice used in games of chance are usually "clockwise"—at least, that's the standard in casinos (so they say). You can find counterclockwise dice in board games (or even irregular dice, with different sums on opposite faces). Not everyone knows there's a difference.

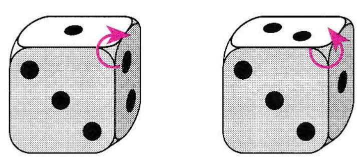


Figure 7

Yakov Smorodinsky, *who passed away in October of last year, belonged to that special class of persons who grow up but never become adults. Mark Twain, Albert Einstein, and Lewis Carroll (to name only a few) are other examples of that breed of "perpetual children." As they grow older, they keep the child's ability to be astonished by the miracles of the world around us, to do strange things (from the point of view of the usual grown-up), and to ask questions that seem silly at first glance but are actually profound. Like all the world's children, they are the real sages, the keepers of wisdom.*

A well-known physicist, Professor Smorodinsky used to play with toys, solve brainteasers, and look for shortcuts in olympiad problems. He was one of the founders of Kvant, *the sister magazine and forerunner of* Quantum. *He was always so energetic and curious, younger in his soul than his much younger colleagues. Now suddenly he is gone, but he has left much behind.*