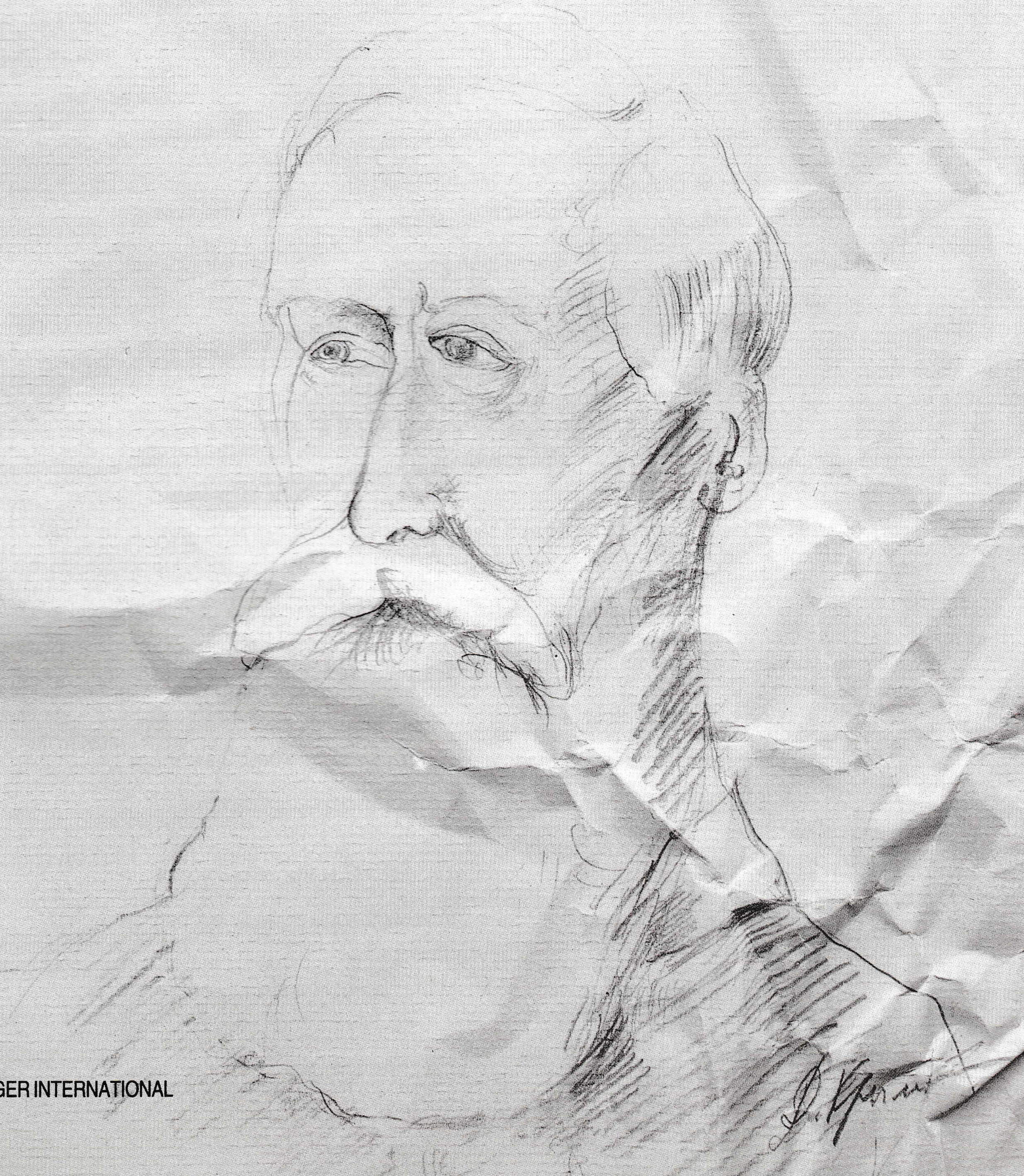
QUANTUM

SEPTEMBER/OCTOBER 1994 $5.95



SPRINGER INTERNATIONAL

The Lackawanna Valley (1855) by George Inness

WE OFTEN THINK OF ARTISTS AS FREE SPIRITS, unencumbered by the mundane considerations that hold the rest of us firmly earthbound. This painting by George Inness (1825–1894) gives the lie to that relatively modern conceit. The work was commissioned by the Delaware and Lackawanna Railroad, and when Inness had completed his vision of the roundhouse at Scranton, he found that his client was far from satisfied. Inness had painted only one line of rails—plans called for three or four more, so the president of the railroad demanded that they be put in the painting in advance of reality. Also, this powerful patron of the arts wanted the entire rolling stock of the company depicted, and he wanted the railroad's initials painted on the sides, and perhaps he had other ideas as well. He was, after all, a practical man.

Inness was faced with a dilemma: paint what he liked and let his family starve, or accommodate the whims of his patron. He gave in. And after everything he went through, Inness later learned the company had sold the painting, and as an old man he recovered it in a junk shop in Mexico.

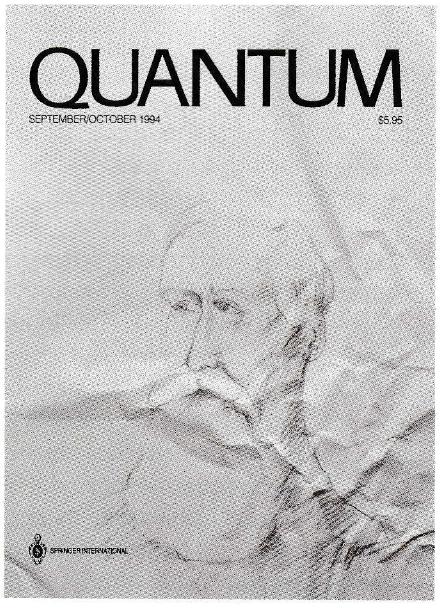
"Who is right, railroad or artist?" asks John Walker, curator emeritus of the National Gallery of Art. He surmises that most of us will stand behind the artist and berate the patron. "Yet many of the greatest works of art were executed in accordance with the strictest contracts," Walker writes—"how many figures to be shown, where they were to stand, how much gold, how much blue, how much red to be used." It is Walker's opinion that Inness benefited from the railroad magnate's restrictions on his artistic freedom: "Today *The Lackawanna Valley* is more highly prized than the misty landscapes he painted at the end of his life, when he had no patron to dictate."

Not only artists are thus constrained. Scientists who receive government research grants, writers for hire, teachers, corporate climbers—they all must tack their sails to some extent if they hope to be paid. And in a less venal light, many of us need the limitations of a specific task to get our creative juices flowing—or to keep us from spinning our wheels.

Questions of practicality keep popping up in this issue of *Quantum*, as do locomotives.

QUANTUM

SEPTEMBER/OCTOBER 1994 VOLUME 5, NUMBER 1



Cover art by Dmitry Krymov

The centenary of the death of the great Russian mathematician Pafnuty Lvovich Chebyshev is marked in this issue with a feature article inspired by his work with a certain class of polynomials and an Anthology piece written by him on the topic of map drawing. (When Chebyshev mentions Watt's parallelogram on page 38, veteran *Quantum* readers may recall the article "Making the Crooked Straight" in the November/December 1990 issue, where Chebyshev makes an appearance.)

Chebyshev was a prodigious engineer, and much of his mathematics arose from practical considerations. In "The Drawing of Geographic Maps" Chebyshev writes that "most practical problems can be reduced to problems of maximum and minimum values"—in others words, "how to dispose of one's means in order to achieve the greatest gain." You may find these words echoing in the back of your mind as you tackle questions of efficiency in the Kaleidoscope.

Vladimir Arnold argues, on the other hand, that "there is nothing more practical than a good theory." His thought-provoking article begins on page 24.

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Cyberspace exploration

Cheap thrills and real science in the computer age

THE NATIONAL SCIENCE Teachers Association has recently installed a T1 Internet node. All of us are very excited about the remarkable explosion of interest in electronic communications, but there are reasons to look with consternation on some aspects of this thrust.

The Internet offers a unique opportunity to bring people from all parts of the world together. To gain some small appreciation for the power of this new technology, one need only sit at his or her computer, bring up MOSAIC, and start looking at images and information from around the world. Or even the simple matter of sending and receiving mail messages or documents can be handled with far greater efficacy. The mix of multimedia and the Internet offers even more interesting and exciting possibilities.

But what are the concerns? The computer is a remarkable device, capable of perfect memory and logical reasoning. Yet it can only remember what it has been given by humans, and its reasoning is limited to drawing logical conclusions from premises provided by those humans.

There are those who call for the use of Internet, multimedia, and of similar computer-based technology to replace much of what is done in science education, particularly laboratory work. For example, in the July issue of *Wired* magazine, Nicholas Negroponte of MIT argues that "since computer simulation of just about anything is now possible, one need not learn about a frog by dissecting it. Instead, children can be asked to design frogs, to build an animal with froglike behavior, to modify that behavior, to simulate the muscles, to play with the frog." Negroponte goes on to emphasize the design aspects of learning.

> "As the great Russian poet Vladimir Mayakovsky said: 'The man who first formulated that "two plus two is four" was a great mathematician, even if he obtained this result from adding two cigarette butts to two cigarette butts. All his followers, even though they may have added immeasurably greater things—say, locomotives to locomotives—are not mathematicians.' . . . Locomotive mathematics lags behind practice: while we teach how to calculate with abacuses, computers appear. We must teach how to think, not how to push buttons."—Vladimir Arnold, "Why Study Mathematics?"

Computer simulations of natural phenomena do not teach science! They represent a form of exposition of science—of unsupported assertion about science. What is worse, they separate the person learning science from nature, imposing an intermediary device that is programmed to model the phenomenon under consideration. This means that the programmer can re-create nature in any way she or he desires, so that science can match nature, or not, as the programmer chooses.

Lest the reader think this is absurd, let me call attention to two such examples. I have seen simulations that allow one to study gas laws. The problem with these simulations is that there are perhaps 50 or more different gas law equations one might use, and the ideal gas law is very wrong for much of what happens to gases. Yet such simulations more often use that law. Thus, the student, removed from the real world of gases, is examining someone else's model of what gases do. That is not science.

I know of another instance where children were examined in terms of their predictions of the motion of a projectile in comparison to an "expert system," which was a computer simulation. Drawings from the children were compared with the expert (computer) so that the children could learn how the motion occurs. As it turned out, the computer used a parabolic path, where the children drew a diagram that was far closer to reality (due to air resistance). The so-called expert system was wrong because the programmers did not bother to include a few lines of code for air resistance. The kids were right because they had observed the motion in real situations.

Science is learned through development of concepts drawn from experience with real phenomena. Certain of those concepts are related to one another in relationships that can be determined through controlled experiments with natural phenomena. When several related empirical relationships are found for which explanations are needed, a theory is created by the human mind. Empirical laws can often be determined more easily by data analysis with a computer, and some models or theories are very amenable to computer modeling. Also, once there are theories and empirical laws, they can be applied in areas of engineering design. But engineering design is not science—it is engineering. And modeling is a high-level skill that should properly be preceded by experience with the phenomena and experimentation to arrive at empirical relationships before computer modeling and design are even included.

Modern technology has the power to greatly improve the ability of students to access raw data, analyze it, and create for themselves some of the natural laws scientists have found. The technology also offers great opportunities to create models and theories and suggest tests for them—which requires real-world measurements or observations. Modern computer technology, along with the proper transducers and coupling devices, can allow much better and easier access by students to real-time data and rapid analysis of those data. Such use of the technology allows far greater opportunity for changing variables and testing hypotheses. *This* is how computer technology should be used. It becomes part of the measuring or observing instrument.

Science is a study of natural phenomena, not a study of what some person has decided is natural phenomena. The technology offers great promise, but some real dangers. (The respected Russian mathematician Vladimir Arnold also sounds the alarm in his article in this issue.) We at NSTA want to use that technology in all of the best ways. With our new Internet node, my address is bgaldridge@nsta.org. Let me hear from you!

—Bill G. Aldridge

QUANTUM

THE 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)

The National Science Teachers Association is an organization of science education professionals and has as its purpose the stimulation, improvement, and coordination of science teaching and learning.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York Inc. Volume 5 (6 issues) will be published in 1994–1995. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau of the Russian Academy of Sciences, as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000, telephone (703) 243-7100. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (202) 328-5800; (New York) Bob Vrooman (212) 460-1700; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone (0) 30-82 07-1, telex 185 411.

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, P. O. Box 2485, Secaucus NJ 07096-2485.

Subscription Information:

North America: Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $34; Single Issue Price: $5.95. 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 07096-2485; or call 1-800-SPRINGER (777-4643) (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, Postfach 31 13 40, D-10643 Berlin, Germany.


S. Ivanov

Foiled by the Coanda effect

"In aerodynamics, theory is what makes the invisible plain. Trying to fly an airplane without theory is like getting into a fistfight with a poltergeist."—David Thornburg

by Jef Raskin

A SOUND THEORETICAL UNderstanding of the phenomenon of "lift" had been achieved within two decades of the Wright brothers' first flight (Ludwig Prandtl's work was most influential),[1] but the most common explanation of lift seen in elementary texts and popular articles today is problematical. Here is a typical example of what is found. Figure 1 is based on an entry in a popular book explaining machines and technology. The reasoning there implicitly involves the Bernoulli effect, which correctly states that the faster air moves over a surface, the lower the air pressure on that surface.

Now, most airplane wings do in fact have considerably more curvature on the top than the bottom, lending credence to this explanation. But even as a child, I found that it presented me with a puzzle: how can a plane fly upside down (something I'd seen at air shows)? When I pressed my teacher on this point, he just got mad, denied that planes can fly inverted, and tried to continue his lecture. I was frustrated and tried to argue until he said, "Shut up, Raskin!" I'll tell you what happened later in the article.

A few years later I carried out a calculation according to a naive interpretation of the common explanation of how a wing works. Using data from a model airplane, I found that the calculated lift was only 2% of that needed to fly the model (see the box on page 8). Given that Bernoulli's equation is correct (indeed, it's a form of the law of conservation of energy), I was left with a second puzzle: where does most of the lift come from?

AIRFOIL

The cross-section of a wing has a shape called an airfoil. As the wing moves through the air, the air divides to pass around the wing. The airfoil is curved so that air passing above the wing moves faster than air passing beneath. Fast-moving air has a lower pressure than slow-moving air. The pressure of the air is therefore greater beneath the wing than above it. This difference in pressure forces the wing upward. The force is called lift.

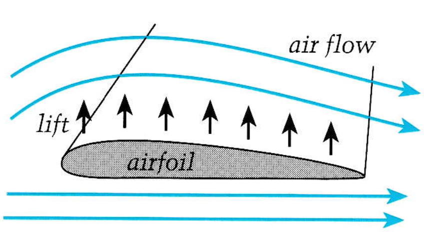


Figure 1

The common explanation. From The Way Things Work *by David Macaulay.*

Let's look at attempts to explain two phenomena—what makes a spinning ball curve and how a wing's shape influences lift—and see how the common explanation of lift has led a surprising number of scientists (including some famous ones) astray.

The spinning ball

The path of a ball spinning around a vertical axis and moving forward through the air is deflected to the right or the left of a straight path. Experiment shows that this effect depends both on the fact that it's spinning and that it's immersed in fluid (air). Nonspinning balls or spinning balls in a vacuum go straight. Before continuing, you might want to decide for yourself which way a ball spinning counterclockwise (when seen from above) will turn.

Let's see what five books say about this problem. Three are by physicists, one is a standard reference work, and the last, just for kicks, is from a book by my son's soccer coach. We'll start with physicist James Trefil, who writes:

[1]Ludwig Prandtl (1875–1953), a German physicist, is often called the "father of aerodynamics." His famous book on the the theory of wings, *Tragflugeltheorie*, was published in 1918.

Art by Sergey Ivanov

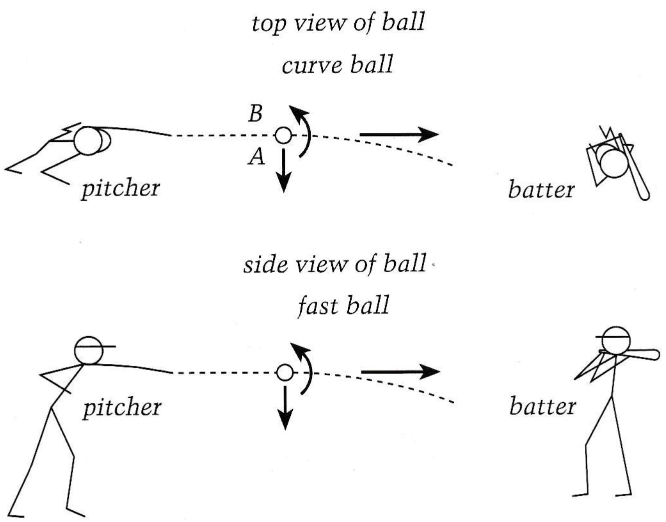


Figure 2

Based on a figure from James Trefil's book A Scientist at the Seashore. *It does not agree with some other sources. (For instance, the direction of spin on the left-hander's curve ball is problematical.)*

Before leaving the Bernoulli effect, I'd like to point out one more area where its consequences should be explored, and that is the somewhat unexpected activity of a baseball. Consider, if you will, the curve ball. This particular pitch is thrown so that the ball spins around an axis as it moves forward, as shown in the top of figure 2.[2] Because the surface of the ball is rough, the effect of viscous forces is to create a thin layer of air which rotates with the surface. Looking at the diagram, we see that the air at the point labeled A will be moving faster than the air at the point labeled B, because in the first case the motion of the ball's surface is added to the ball's overall velocity, while in the second it is subtracted. The effect, then, is a "lift" force, which tends to move the ball in the direction shown.[3]

Baseball aficionados would say that the ball curves toward third base. Trefil then shows a diagram of a fast ball, shown as deflecting downward when spinning so that the bottom of the ball is rotating forward. It is the same phenomenon with the axis of rotation shifted 90 degrees.

In *The Physics of Baseball*, Robert K. Adair imagines a ball thrown toward home plate so that it rotates counterclockwise as seen from above, as in Trefil's diagram. To the left of the pitcher is first base, to the right is third base. Adair writes:

> We can then expect the air pressure on the third-base side of the ball, which is traveling faster through the air, to be greater than the pressure on the first-base side, which is traveling more slowly, and the ball will be deflected toward first base.

This is exactly the opposite of Trefil's conclusion, though they agree that the side spinning forward is moving faster through the air. We have learned from these two sources that going faster through the air either increases or decreases the pressure on that side. I won't take sides in this argument as yet.

The *Encyclopaedia Britannica* (1979) gives a different reasoning that introduces the concept of drag into the discussion:

> The drag of the side of the ball turning into the air (into the direction the ball is traveling) retards the airflow, whereas on the other side the drag speeds up the airflow. Greater pressure on the side where the airflow is slowed down forces the ball in the direction of the low-pressure region on the opposite side, where a relative increase in airflow occurs.

Now we have read that spinning the ball causes the air to move either faster or slower past the side spinning forward, and that faster-moving air increases or decreases the pressure, depending on the authority you choose to follow. Speaking of authority, it might be appropriate to turn to one of the giants of physics of this century, Richard Feynman. He and his coauthors take the side of Trefil and uses a cylinder rather than a sphere (the italics are theirs, and the lift force referred to is shown pointing upward):

> The flow velocity is higher on the upper side of a cylinder [shown rotating so that its top is moving in the same direction as its forward travel] than on the lower side. The pressures are therefore *lower* on the *upper* side than on the lower side. So when we have a combination of a circulation around a cylinder and a net horizontal flow, there is a net *vertical force* on the cylinder—it is called a *lift force*.

Now for my son's coach's book. The coach in this case is the world-class soccer player George Lamptey. No theory is given, but we can be reasonably sure that Lamptey has repeatedly tried the experiment and should therefore report correctly the direction the ball turns. He writes:

> The banana kick is more or less an off-center instep drive kick which adds a spin to the soccer ball. Kick off center to the right, the soccer ball curves to the left. Kick off center to the left, the soccer ball curves to the right. . . . The amount the soccer ball curves depends on the speed of the spin.

As you can see from figure 3, Lamptey, like Adair, has the high pressure on the side moving into the air. I won't relate other accounts, some having the ball turn one way, some the other. Some explanations depend on the author's interpretation of the Bernoulli effect, some on viscosity, some on drag, some on turbulence.

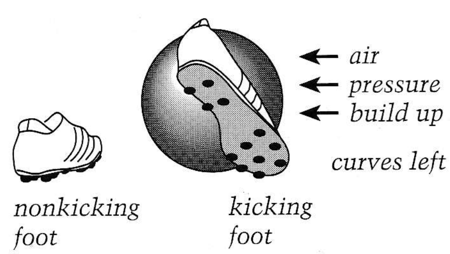


Figure 3

We'll return to the subject of spinning balls, but we're not yet finished finding problems with the common explanation of lift.

Other paradoxes

The traditional explanation of how a wing works leads us to conclude, for example, that a wing that is somewhat concave on the bottom—often called an "undercambered" wing—will always generate *less* lift, under

[2]I have renumbered Trefil's figure.

[3]The surface roughness is not essential. The effect is observed no matter how smooth the ball.

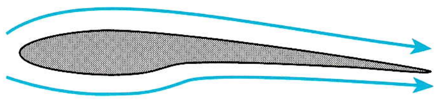


Figure 4
Undercambered airfoil.

otherwise fixed conditions, than a flat-bottomed one (fig. 4). With an undercambered airfoil, the bottom path of the air is longer than it is with the flat-bottomed airfoil in figure 1. Therefore, less lift—right? Wrong!

Then we have to ask how a flat wing like that of a paper airplane, with no curves anywhere, can generate lift (fig. 5). Note that the flat

Figure 5
Flat wing.

wing has been drawn at a tilt. This tilt is called the "angle of attack" and is necessary for the flat wing to generate lift. We'll return to this topic later in the article.

The cross-sectional shape of wings, like those illustrated here, are called "airfoils." A very efficient airfoil for small, slow-flying models is an arched piece of this sheet material (fig. 6), but it's not clear at all

Figure 6
Curved wing used in small model airplanes.

from the common explanation how it can generate *any* lift, since the top and bottom of the airfoil are the same length.

If the common explanation were correct, we should be making the tops of wings even curvier than they now are. Then the air would have to go even faster, and we'd get more lift. In figure 7 the wiggliness is exaggerated. (We'll encounter more realistic lumpy examples below.) If we make the top of the wing like in figure 7, the air on top has a lot longer path to follow, so the air will go even faster than with a conventional wing. You might conclude that this kind of airfoil should have lots of lift. In fact, it's a disaster.

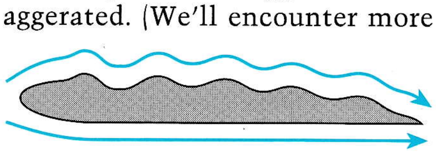


Figure 7
Wiggly wing.

Enough examples. While Bernoulli's equations are correct, their proper application to aerodynamic lift differs greatly from the common explanation. Applied properly or not, the equations offer no convenient visualization that links the shape of an airfoil with its lift and reveal nothing about drag. This difficulty, combined with the prevalence of the plausible-sounding common explanation, is probably why even some excellent physicists have been misled.

Albert Einstein's wing

My friend Yesso, who works for the aircraft industry (though not as a designer), came up with a proposed improved airfoil. Reasoning along the lines of the common explanation, he suggested that you should get more lift from an airfoil if you restarted the top's curve part of the way along (fig. 8). This is just a "reasonable" version of the wiggly airfoil we looked at earlier. Yesso's idea was, of course, based on the concept that a longer upper surface should give more lift. I was about to tell Yesso why his foil idea wouldn't work when I happened to talk to Jörgen Skogh, who worked on aircraft design for Saab in Sweden. He told me of a humped airfoil Albert Einstein designed during World War I, based on much the same reasoning Yesso had used (fig. 9). It had no aerodynamic virtues. This meant that instead of telling Yesso merely that his idea wouldn't work, I could tell him that he had created a modernized version of Einstein's error! Einstein later noted, with chagrin, that he had goofed![4]

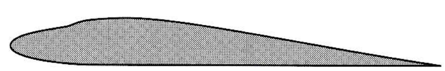


Figure 8
Lumpy wing.

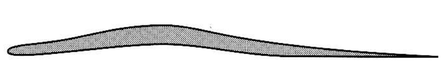


Figure 9
Albert Einstein's airfoil.

Evidence from experiments

If it were the case that airfoils generate lift solely because the airflow across a surface lowers the pressure on that surface, then if the surface is curved, it doesn't matter whether it's straight, concave, or convex. The common explanation depends only on flow parallel to the surface. Here are some experiments that you can easily reproduce to test this idea.

Experiment 1. Make a strip of writing paper about 5 cm × 25 cm. Hold it in front of your lips so that it hangs out and down, making a convex upward surface. When you blow across the top of the paper, it rises (fig. 10a). Many books attribute this to the lowering of the air pressure on top due to

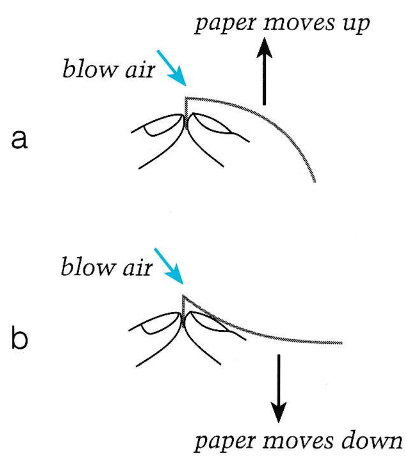


Figure 10

[4]Jörgen Skogh writes: "During the First World War, Albert Einstein was for a time hired by the LVG (Luft-Verkehrs-Gesselschaft) as a consultant. At LVG he designed an airfoil with a pronounced mid-chord hump, an innovation intended to enhance lift. The airfoil was tested in the Göttingen wind tunnel and also on an actual aircraft and found, in both cases, to be a flop." In 1954 Einstein wrote: "Although it is probably true that the principle of flight can be most simply explained in this [Bernoullian] way, it by no means is wise to construct a wing in such a manner!"

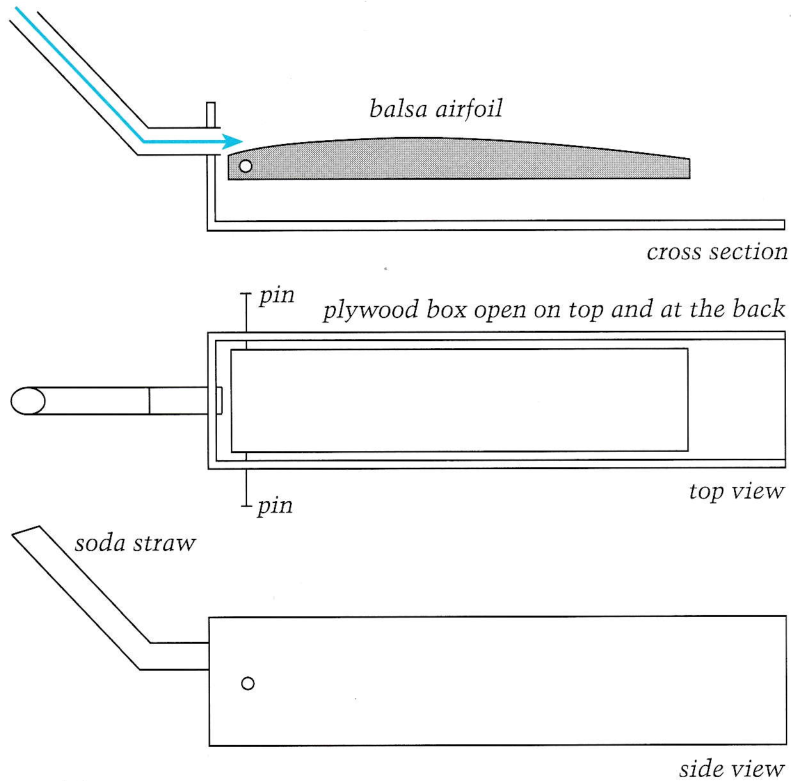


Figure 11

the Bernoulli effect. Now use your fingers to form the paper into a curve that is slightly concave upward along its whole length and again blow across the top of the strip. The paper now bends downward (fig. 10b).

Experiment 2. Referring to figure 11, build a box of thin plywood or cardboard with a balsa airfoil held in place with pins that allow it to flap freely up and down. Air is introduced with a soda straw. That's one of the nice things about science. You don't have to take anybody's word for a claim—you can try it yourself![5] In this wind tunnel the air flows only across the top of the shape. I made another where a vacuum cleaner blew on both the top and bottom, and I got the same results; but that design takes more effort to build and the airfoil models require refinement of the leading and trailing edges. Incidentally, I tried to convince a company that makes science demonstrators to include this in their offerings. They weren't interested because "it didn't give the right results." "Then how does it work?" I asked. "I don't know," said the head designer.

An experiment may be difficult to interpret, but unless it's fraudulent, it cannot give the wrong results.

Experimental results

When air is blown through the straw, the normal airfoil (see figure

[5]In some fields—for instance, the study of subatomic particles—you might need megabucks and a staff of thousands to build an accelerator to do an independent check, but the principle is still there.

A quantitative application of an incorrect explanation

If the pressure in newtons per square meter ($N/m^2 = kg/(m \cdot s^2)$) on the top of a wing is notated p_{top}, the pressure on the bottom p_{bottom}, the velocity (m/s) on the top of the wing v_{top}, and the velocity on the bottom v_{bottom}, and where ρ is the density of air (approximately 1.2 kg/m^3), then the pressure difference across the wing is given by the first term of Bernoulli's equation:

$$p_{top} - p_{bottom} = \frac{1}{2}\rho\left(v_{top}^2 - v_{bottom}^2\right).$$

A rectangular planform (top view) wing of one-meter span was measured as having a length chordwise along the bottom of 0.1624 m, while the length across the top was 0.2636 m. The ratio of the lengths is 1.0074. This ratio is typical for many model and full-size aircraft wings. Since the top and bottom of the wing are part of the same rigid object, they are moving at equal velocities; thus, according to the common explanation, the air velocities on the top and bottom are also in the ratio of 1.0074.

A typical speed for a model plane of 1-m span and 0.16-m chord with a mass of 0.7 kg (a weight of 6.9 N) is 10 m/s, so v_{bottom} is 10 m/s, which means that v_{top} = 10.074 m/s. Given these numbers, we find a pressure difference from the equation of about 0.9 $kg/(m \cdot s^2)$. The area of the wing is 0.16 m^2, giving a total force of 0.14 N. This is not nearly enough—it misses lifting the weight of 6.9 N by a factor of about 50. We would need an air velocity difference of about 3 m/s to lift the plane.

This calculation is an approximation, since Bernoulli's equation assumes nonviscous, incompressible flow, and air is both viscous and compressible. But the viscosity is small, and at the speeds we are speaking of air does not compress significantly. Accounting for these details changes the outcome at most a percent or so. This treatment also ignores the second term (not shown) of the Bernoulli equation—the static pressure difference between the top and bottom of the wing due to their trivially different altitudes. Its contribution to lift is even smaller than the effects already ignored. The use of an average velocity assumes a circular arc for the top of the wing. This is not optimal but it will fly. None of these details affects the conclusion that the common explanation of how a wing generates lift—with its naive application of the Bernoulli equation—fails quantitatively.

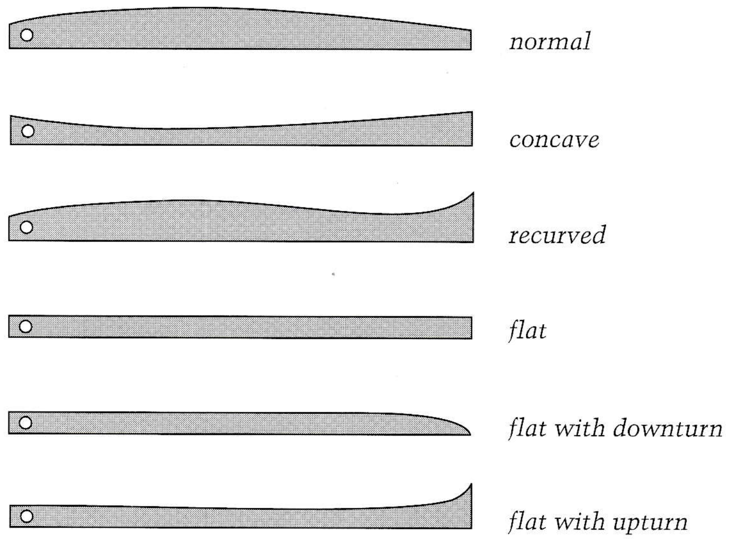


Figure 12

12) promptly lifts off the bottom and floats up. When the blowing stops, it goes back down. This is exactly what everybody expects. Now consider the concave shape. The curve is exactly the same as the first airfoil, though turned upside down. If the common explanation were true, then, since the length along the curve is the same as with the "normal" example, you'd expect this one to rise, too. After all, the airflow along the surface must be lowering the pressure, causing lift. Nonetheless, the concave airfoil stays firmly down. If you hold the apparatus vertically, it will be seen to move *away* from the airflow.

In other words, an often cited experiment that is usually taken as demonstrating the common explanation of lift doesn't do so—another effect is far stronger. The rest of the airfoils are for fun—try to anticipate the direction each will move before you put them in the apparatus. James Gleick has noted that "progress in science comes when experiments contradict theory," although in this case the science has long been known, and the experiment contradicts not aerodynamic theory but the often taught common explanation. Nonetheless, even if science doesn't progress in this case, an individual's understanding of it may. Another simple experiment will lead us toward an explanation that may help give us a better feel for these aerodynamic effects.

The Coanda effect

If a stream of water is flowing along a solid surface that is curved slightly away from the stream, the water tends to follow the surface. This is an example of the Coanda effect[6] and is easily demonstrated by holding the back of a spoon vertically under a thin stream of water from a faucet (fig. 13). If you hold the spoon so that it can swing, you will feel it being pulled *toward* the stream of water. The effect has limits: if you use a sphere instead of a spoon, you'll find that the water will follow only a part of the way around. Also, if the surface is to sharply curved, the water will not follow but will just bend a bit and break away from the surface.

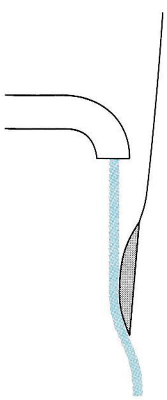


Figure 13

[6]In the 1930s the Romanian aerodynamicist Henri-Marie Coanda (1885–1972) observed that a stream of air (or other fluid) emerging from a nozzle tends to follow a nearby curved or flat surface, if the curvature of the surface or the angle the surface makes with the stream is not too sharp.

The Coanda effect works with any of our usual fluids (such as air—see figure 14) at usual temperatures, pressures, and speeds. I make these qualifications because liquid helium, gases at extremes of low or high pressure or temperature, and fluids at supersonic speeds may behave rather differently. Fortunately, we don't have to worry about all of these extremes with model planes.

Another thing we don't have to worry about is *why* the Coanda effect works—we can take it as an experimentally given fact. But if you're curious, we can touch on it lightly. On a microscopic scale we note that most gas molecules that come close to one another generate a small attractive force, called the van der Waals force, that tends to keep them together. There are a number of sources of these van der Waals forces, but the main one is due to the fact that the charge distribution of electrically neutral molecules, such as hydrogen, oxygen, and nitrogen, is distorted by their proximity to one another in such a way as to create a dipole (the electrical equivalent of a magnet). These dipoles arrange themselves so that the positive end of one is near the negative end of another. Thus, they develop a minuscule electrostatic "cling." For inherently polar molecules, such as water or carbon dioxide, their mutual attraction is increased by proximity. For similar

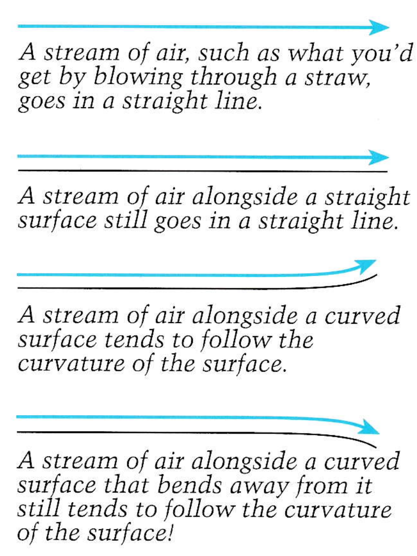


Figure 14

reasons, gas molecules tend to cling to liquid or solid surfaces. On the other hand, molecules resist being pushed closer together than the van der Waals forces bring them. The resistance to compression due mostly to electrostatic repulsion (the electrons surrounding each atomic nucleus are all negative and —if forced too close—repel each other), along with the motion of the molecules, are the mechanisms underlying pressure. The van der Waals forces explain why at least a thin layer of a fluid follows a surface.

A word often used to describe the Coanda effect is to say that the airstream is "entrained" by the surface. One advantage of discussing lift and drag in terms of the Coanda effect is that we can trace the forces involved in a rather straightforward way. The common explanation (and the methods used in serious texts on aerodynamics) are anything but clear in showing how the motion of the air is physically coupled to the wing. This is partly because much of the approach taken in the 1920s was shaped by the need for the resulting differential equations (mostly based on the Kutta–Zhukovsky theorem[7]) to have closed-form solutions or to yield useful numerical results with paper-and-pencil methods. Modern approaches use computers and are based on only slightly more intuitive constructs. We will now develop an alternative way of visualizing lift that makes it easier to predict the basic phenomena associated with it.

Visualizing lift and drag

As is typical of physicists, I have often spoken of the air moving past the wing. In fact, aircraft wings usually move through the air. It makes no real difference, as flying a slow plane into the wind so that its ground speed is zero demonstrates. So I will speak of the airplane moving or the wind moving, whichever

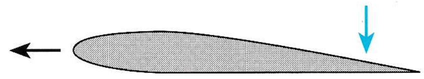


Figure 15

The surface moves to the left and the air molecules, attracted to the surface, are pulled down.

makes the point more clearly at the time. In the next illustration, it becomes convenient to look at the air from the point of view of a moving airplane.

In figure 15, think of the wing moving to the left and the air standing still. The air gets pulled toward the wing by the attractive and pressure forces just discussed, much as if they were attached to the wing with invisible rubber bands. It's often helpful to think of lift as the action of the rubber bands that are pulling the wing up.

Another detail is important: the air gets pulled along in the direction of the wing's motion as well. So the action is really more like figure 16.

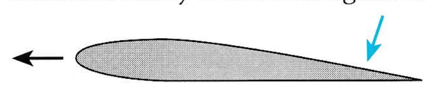


Figure 16

The air is pulled forward as well as down by the motion of the wing.

If you were in a canoe and tried to pull someone in the water toward you with a rope, your canoe would move toward the person. The wing is like the boat, the air is the person, and the rope models the attractive force between the wing and the air. It is classic "action and reaction." You move a mass of air down and the wing moves up. That's the *lift* generated by the top of the wing.

As the diagram suggests, the wing has also spent some of its energy, necessarily, in moving the air forward. The imaginary rubber bands pull it back some. That's the *drag* that is caused by the lift you are generating. When lift is considered this way, it's immediately clear that it can't occur without the penalty of some drag.

The acceleration of the air around the sharper curvature near the front of the top of the wing also imparts a downward and forward component to the motion of the molecules of air (usually a slowing of their upward and backward motion, which is equivalent) and thus contributes to lift. The bottom of the wing is easier to understand, and an explanation is left to the reader.

The experiments with the miniature wind tunnel described earlier are readily understood in terms of the Coanda effect: the downward-curved wing entrained the airflow to move downward, and a force upward is developed in reaction. The upward-curved (concave) airfoil entrained the airflow to move upward, and a force downward was the result. The lumpy wing generates a lot of drag by moving air molecules up and down repeatedly. This eats up energy (by generating frictional heat), but doesn't create a net downward motion of the air and therefore doesn't create a net upward movement of the wing. The Coanda effect helps us visualize why angle of attack (the fore-and-aft tilt of the wing, as illustrated earlier) is crucially important; why planes can fly inverted; why flat and thin wings work; and why experiment 1 with its convex and concave strips of paper works as it does.

What has been presented so far is by no means a complete account of lift and drag, but it does tend to give a good picture of the phenomenon. We will now use this grasp to get a reasonable hold on the spinning ball problem.

Why the spinning ball curves

Let's take another look at the figure from James Trefil's book (fig. 2). The Coanda effect tells us that the air is pulled along with the surface of the ball. Consider Trefil's side A, which is rotating in the direction of flight. It is trying to entrain air with it as it spins. This action is opposed by the oncoming air. Thus, to entrain the air around the ball on this side, it must first decelerate it and then reaccelerate it in the opposite direction. On the B side, which is rotating opposite the direction of flight, the air is already moving (relative to the ball) in the

[7]Discovered independently by the German mathematician M. Wilhelm Kutta (1867–1944) and the Russian physicist Nikolay Zhukovsky (1847–1921).

same direction and is thus more easily entrained. The air more readily follows the curvature of the B side around and acquired a velocity toward the A side. The ball therefore moves toward the B side by reaction.

It's time again for a simple experiment. It's difficult to experiment with baseballs because their weight is large compared to the aerodynamic forces acting on them, and it's very hard to control the magnitude and direction of the spin. So let's look at a case where the ball is lighter and the aerodynamic effects easier to see. I use a cheap inflatable beach ball (expensive ones are make of heavier material and show aerodynamic effects less). Thrown with enough bottom spin (bottom moving forward), such a ball will actually *rise* in a curve as it travels forward. The lift due to spin can be so strong that it's greater than the downward force of gravity! Soon air resistance stops both the spin and the forward motion of the ball and it falls, but not before it has shown that Trefil's explanation of how spin affects the flight of a ball is wrong.

The lift due to spinning while moving through the air is usually called the *Magnus effect*.[8] Some books discuss the "Flettner rotor," which is a long-since abandoned attempt to use the Magnus effect to make an efficient boat sail. Many authors besides Trefil get the effect backwards, including the usually reliable S. F. Hoerner in his *Fluid-Dynamic Drag* (1965). College-level texts tend to get it right, but as noted in part I, Feynman's *Lectures on Physics* has the rotation backwards. I was relieved to see that the classic *Aerodynamics* by T. von Kármán gets the lift force on a spinning ball in the correct direction, though the reasoning seems a bit strained.

I wish I could send this article to that sixth-grade science teacher who wouldn't take the time to listen to my reasoning. Here's what happened. He sent me to the principal's office when I came in the next day with a balsa model plane with dead flat wings. It would fly with either side up, depending on how an aluminum foil elevator was adjusted. I used it to demonstrate that the explanation the class had been given must have been wrong, somehow. The principal, however, was informed that my offense was "flying paper airplanes in class," as though done with disruptive intent. After being warned that I was to improve my behavior, I went to my beloved math teacher, who suggested that I go to the library to find out how airplanes fly—only to discover that all the books agreed with my science teacher!

It's easier to think about lift and drag in terms of bending a stream of air than in terms of pressure differences.

It was a shock to realize that my teacher and even the library books could be wrong. And it was a revelation that I could trust by own thinking in the face of such concerted opposition. My playing with model airplanes had led me to take a major step toward intellectual independence—and a spirit of innovation that later led me to create the Macintosh computer project and other inventions as an adult.

Further reading

There are many fine books and articles on the subject of model airplane aerodynamics (and many more on aerodynamics in general). Commendably accurate and readable are books and articles for modelers by Professor Martin Simons (for example, *Model Aircraft Aerodynamics*, 2nd ed., Argus Books Ltd., London, 1987). Much can be learned from Frank Zaic's delightful, if not terribly technical, series of the *Model Aeronautic Yearbook* (1936–64, available from the Academy of Model Aeronautics in the United States). No treatments are more professional than those of Michael Selig (for example, Selig et al., *Airfoils at Low Speeds*, Soartech 8, 1989, available from Herb Stokely, 1504 Horseshoe Circle, Virginia Beach VA 23451). All of these authors are also well-known modelers. Graduate or upper-level undergraduate texts—for example, *Foundations of Aerodynamics* by Kuethe and Chow and *Aerodynamics for Engineering Students* by Houghton and Carruthers—require a knowledge of calculus, including partial differential equations. *Modern Subsonic Aerodynamics* by R. T. Jones (Aircraft Designs, Inc., 1988) is an informal treatment by a master, and *Fluid-Dynamic Drag* by S. F. Hoerner (Hoerner Fluid Dynamics, 1965) is a magnificent compendium of experimental results—it has little theory, but practical designers find his work invaluable. Finally, the epigraph came from *Do You Speak Model Airplane?* by Dave Thornburg (Pony X Press, 1992, 5 Monticello Dr., Albuquerque NM 87123).

I am very appreciative of the suggestions I have received from a number of careful readers, including Bill Aldridge, Dr. Vincent Panico, Professor Michael Selig, Professor Steve Berry, and Linda Blum. They have materially improved both the content and the exposition, but where I have not taken their advice my own errors probably shine through.

Jef Raskin *was a professor at the University of California at San Diego and originated the Macintosh computer at Apple Computer, Inc. He is a widely published writer, an avid model airplane builder and competitor, and an active musician and composer.*

[8]H. G. Magnus (1802–1870), a German physicist and chemist, demonstrated this effect in 1853.

Chebyshev's problem

Polynomials of least deviation from zero

by S. Tabachnikov and S. Gashkov

WITH THIS ARTICLE WE PAY TRIBUTE TO THE outstanding Russian mathematician Pafnuty Chebyshev, who died a hundred years ago, on November 26, 1894. (See also the Anthology department on page 35). His famous achievements include a solution of one of the most beautiful problems concerning polynomials—Chebyshev's problem on polynomials of least deviation from zero. We'll derive the properties of these polynomials in two ways: in the first section of the article (written by the first author), by exploring the geometry of their graphs; in the following sections (written by the second author), by undertaking a little investigation into trigonometric polynomials in general. Choose whichever you like better.

Fix a segment of the number axis, say, [–2, 2]. (Well, to be fair, it's not just "say"—the formulas for this segment come out the simplest.) Let

$$f(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_0$$

be a *reduced* polynomial of the nth degree (this means that its leading coefficient is 1). The range of $f(x)$ on the segment [–2, 2] is the segment $[m, M]$, where m is the minimum and M the maximum value of the polynomial. The *deviation of f(x) from zero* is the greatest of the numbers $|m|$ and $|M|$. If the deviation is c, then the graph of the polynomial (on [–2, 2]) lies entirely in the band $|y| \le c$ and is not contained in any narrower band with the same midline (the x-axis).

Chebyshev's problem consists in finding a reduced polynomial $f_n(x)$ of degree n whose deviation from zero is the smallest. (The condition that its leading coefficient is 1 does not allow the graph to be squeezed arbitrarily close to the x-axis.) On the face of it, this problem doesn't excite a lot of enthusiasm: to find the deviation, you have to take derivatives and solve equations of the nth degree . . . It's all the more surprising that it can be solved geometrically—and almost without calculations!

Art by Dmitry Krymov

Let's start small and look at the cases of small degrees. For $n = 1$ we deal simply with a linear function $f(x) = x + a$. Its range is the segment $[-2 + a, 2 + a]$ of length 4. So the least deviation from zero equals 2, and $f_1(x) = x$.

The case $n = 2$ (quadratic function) is only a little more complicated. Here the graph is a shifted segment of the parabola $y = x^2$, and it's not hard to see that its most economical position is the one in figure 1. That is, $f_2(x) = x^2 - 2$, and the deviation from zero is again equal to 2.

Exercise 1. Check your intuition by calculation: prove that the deviation of a reduced quadratic polynomial from zero is no less than 2.[1]

We could have offered you the investigation of the case of cubic polynomials to make sure that the least deviation from zero is 2 in this case as well. This problem can still be handled with our "bare hands." But we can't wait to tell you how the general problem is solved.

Suppose we've succeeded in detecting a reduced polynomial $f_n(x)$ of degree n whose graph lies in the band $|y| \le c$ and contains $n + 1$ points on the band's border such that the rightmost of these points lies on the line $y = c$, the next point on the left lies on $y = -c$, the next one again lies on $y = c$, and so on (see figure 2 on the next page for $n = 5$).

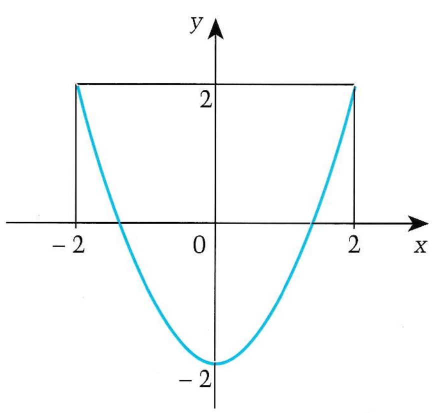


Figure 1

[1]See also challenge M120 in the July/August 1994 issue.—*Ed.*

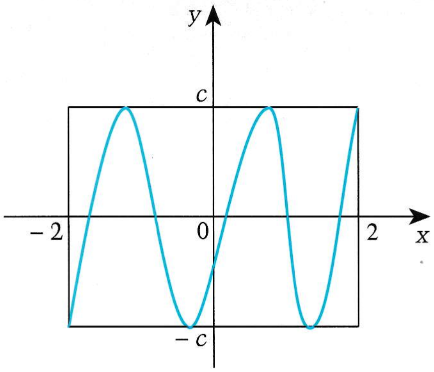


Figure 2

THEOREM. *The deviation from zero of any polynomial of the nth degree other than $f_n(x)$ is greater than c.*

Here's the geometric proof that makes Chebyshev's problem so attractive. Let $g(x)$ be any reduced polynomial of degree n whose deviation from zero does not exceed c. Then its graph also lies in the band $|y| \le c$. Cut this band with vertical segments into n rectangles as shown in figure 3. The blue curve in this figure—the graph of $f_n(x)$—joins diagonally two opposite vertices of each rectangle. Therefore the red line—the graph of $g(x)$—intersects the blue one inside each of the rectangles. So the equation $f_n(x) - g(x) = 0$ has no less than n roots. But the degree of the polynomial $f_n(x) - g(x)$ is no greater than $n - 1$ (the terms x^n cancel out). If such a polynomial has n or more roots, it has to be identically equal to zero—that is, $g(x) = f_n(x)$. The proof is complete.

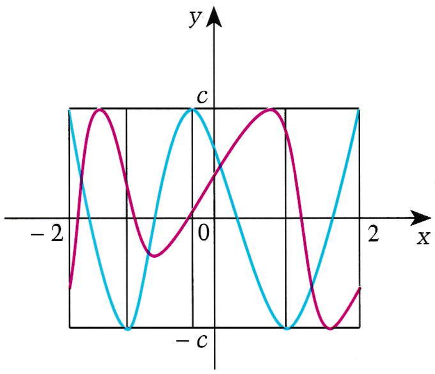


Figure 3

On second thought, we see that this proof gives us "free of charge" quite a lot of additional information: although we don't know yet the value of c, we do know that the polynomial of least deviation from zero is unique for a given degree, and we can visualize its graph.

Exercise 2. There's a gap in the proof above (this is the price of its beauty). What should be done if the two graphs touch one another (fig. 4)? (Hint: recall the definition of a multiple root of a polynomial.)

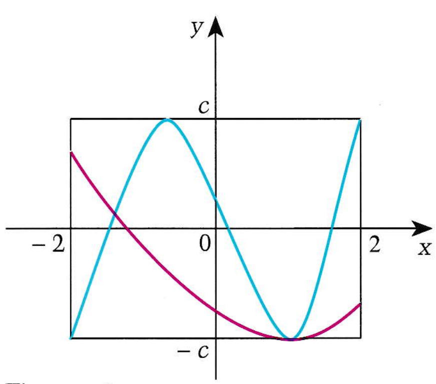


Figure 4

Thus, we have to bring forward a polynomial $f_n(x)$ with a graph like the one in figure 2. This figure will undoubtedly remind you of trigonometric functions. Sure enough, it's time for them to come into play!

It would be simplest to take as $f_n(x)$ something like the function cos $n\alpha$—except that the cosine function isn't a polynomial. The following lemma comes to the rescue.

LEMMA. *The function 2 cos nα can be written as a reduced polynomial $f_n(x)$ of degree n of the function 2 cos α: 2 cos nα = f_n(2 cos α).*

For instance,

$$2 \cos 2\alpha = 4 \cos^2\alpha - 2 = (2 \cos \alpha)^2 - 2$$

—that is, $f_2(x) = x^2 - 2$;

$$2 \cos 3\alpha = 8 \cos^3\alpha - 6 \cos \alpha = (2 \cos \alpha)^3 - 3(2 \cos \alpha)$$

—that is, $f_3(x) = x^3 - 3x$.

In the general case we use induction over n. Suppose the statement of the lemma is true for a certain n and $n - 1$ $(n \ge 2)$:

$$2 \cos (n-1)\alpha = f_{n-1}(2 \cos \alpha),$$
$$2 \cos n\alpha = f_n(2 \cos \alpha).$$

Then the formula

$$\cos A + \cos B = 2 \cos (A - B)/2 \cos (A + B)/2$$

implies

$$\cos (n+1)\alpha + (\cos n - 1)\alpha = 2 \cos \alpha \cos n\alpha,$$

or

$$2 \cos (n+1)\alpha = (2 \cos \alpha)(2 \cos n\alpha) - 2 \cos (n-1)\alpha$$
$$= (2 \cos \alpha) f_n(2 \cos \alpha) - f_{n-1}(2 \cos \alpha).$$

It follows that

$$f_{n+1}(x) = x f_n(x) - f_{n-1}(x).$$

This completes the proof and, in addition, gives a recursive formula for computing $f_n(x)$.

Perhaps you've guessed that the polynomials $f_n(x)$ are just what we need. Indeed, let α run through the segment [0, π]. Then $n\alpha$ varies from 0 to $n\pi$, and the functions $x = 2 \cos \alpha$ and $f_n(x) = 2 \cos n\alpha$ take values on [−2, 2]. In so doing, x sweeps this segment once, while $f_n(x)$ sweeps it n times, alternately taking values ±2 for $x = \arccos (\pi k/n)$, $k = 0, 1, \ldots, n$. This means that the graph of $f_n(x)$ lies in the band $|y| \le 2$ and passes alternately through $n + 1$ points on its upper and lower borders. That is, $f_n(x)$ is the polynomial of least deviation from zero on the segment [−2, 2] (and the deviation is 2). These polynomials are called *Chebyshev polynomials.*

The discussion above leads to an unexpected conclusion: *for any reduced polynomial g(x), there is a point in [−2, 2] such that the absolute value of g(x) at this point is no less than 2.* We think it would be impossible to foresee such a result.

After Chebyshev's problem has been solved for the segment [−2, 2], one can solve it on any other segment. It will suffice to change the variable in the Chebyshev polynomials.[2]

[2]In fact, it is more usual to define them as polynomials given on the segment [−1, 1] by the formula $T_n(x) = \frac{1}{2} f_n(2x)$.—*Ed.*

Exercise 3. Find the least deviation from zero for reduced polynomials of degree n on the segments (a) $[0, 4]$, (b) $[-1, 1]$.

Two kinds of polynomials

In addition to the common algebraic polynomials you've encountered at school, mathematicians also study the *trigonometric polynomials*

$$f(\alpha) = a_0 + a_1 \cos \alpha + a_2 \cos 2\alpha + \ldots + a_n \cos n\alpha, \qquad (1)$$

where $a_0, a_1, \ldots, a_n$ are numerical coefficients. The number n is called the *degree* and a_n the *leading coefficient* of this trigonometric polynomial.

You may be wondering, "Why 'polynomials?'" The reason is that $f(\alpha)$ is an *algebraic* polynomial of the function $\cos \alpha$. In the first section, it was proved that $2 \cos \alpha$ is a polynomial of degree n of $2 \cos \alpha$ with leading coefficient 1. For any $n = 0, 1, 2, \ldots,$

$$\cos n\alpha = p_n(\cos \alpha),$$

where $p_n(x)$ is a polynomial of degree n with leading coefficient 2^{n-1}:

$$\begin{aligned}
&\cos 0\alpha = 1 = p_0(x),\\
&\cos \alpha = x = p_1(x),\\
&\cos 2\alpha = 2\cos^2 \alpha - 1 = 2x^2 - 1 = p_2(x), \qquad (2)\\
&\cos 3\alpha = 4\cos^3 \alpha - 3\cos \alpha = 4x^3 - 3x = p_3(x),\\
&\cos n\alpha = 2^{n-1}x^n + \ldots = p_n(x),
\end{aligned}$$

where $x = \cos \alpha$.

Multiplying these equalities by $a_0, a_1, \ldots, a_n$, respectively, adding them up, we get a representation of a general trigonometric polynomial (1) as an algebraic polynomial of degree n of $\cos \alpha$.

Now let's read equalities (2) from right to left and from the bottom up. The last of them gives

$$x^n = \frac{1}{2^{n-1}} \cos n\alpha + \cdots, \qquad (3)$$

where the dots mean the terms containing x^k, $k < n$. The second expression from the bottom becomes

$$x^{n-1} = \frac{1}{2^{n-2}} \cos(n-1)\alpha + \cdots, \qquad (4)$$

where the dots stand for monomials with x^k, $k < n - 1$, and so on. Plugging equation (4) and the similar equations for x^k, $k < n - 1$, into equation (3), we come up with this: *if $x = \cos \alpha$, then x^n is a trigonometric polynomial of degree n with leading coefficient $1/2^{n-1}$.*

It follows that any algebraic polynomial $p_n(x)$ of the nth degree and with leading coefficient 1 for $x = \cos \alpha$ becomes a trigonometric polynomial of the nth degree with leading coefficient $1/2^{n-1}$.

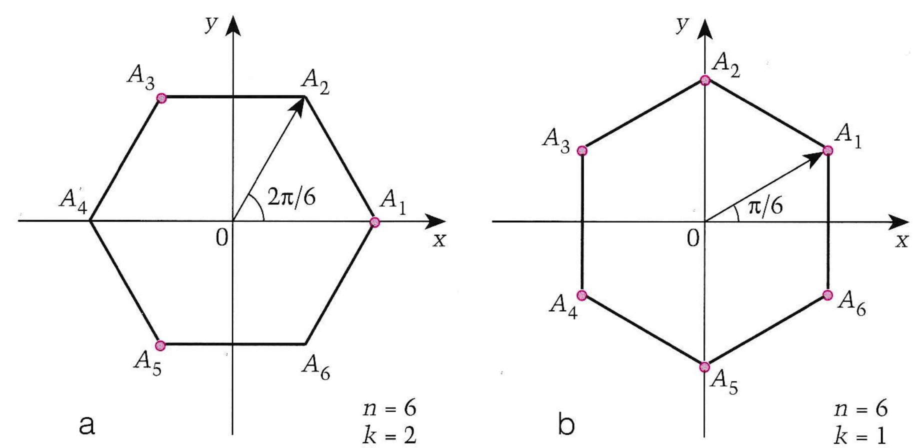


Figure 5

The mean value of a trigonometric polynomial

Consider a trigonometric polynomial with leading coefficient a_n. Divide its domain $[0, 2\pi]$ into $2n$ equal parts with the points $0, \pi/n, 2\pi/n, 3\pi/n, \ldots, (2n-1)\pi/n$, and calculate the value of

$$\frac{1}{2n}\left[f(0) - f\left(\frac{\pi}{n}\right) + f\left(\frac{2\pi}{n}\right) - f\left(\frac{3\pi}{n}\right) + \cdots - f\left(\frac{(2n-1)\pi}{n}\right)\right]. \qquad (5)$$

This will be called, for now, the *mean value* of the polynomial (but note the alternating signs of the terms).

LEMMA. *This number equals a_n.*

Since a trigonometric polynomial (equation (1)) is the sum of functions $g_k(\alpha) = \cos k\alpha$, $0 \le k \le n$, with number coefficients, its mean value (expression (5)) is a similar sum (with the same coefficients) of the mean values of functions g_k. Let's prove that these mean values for all $k < n$ are zeros.

For $k = 0$, we have $g_0(\alpha) = 1$, so

$$g_0 - g_0\left(\frac{\pi}{n}\right) + \cdots = 1 - 1 + 1 - \cdots - 1 = 0.$$

Now take any k, $0 < k < n$. The sum in expression (5) for $f = g_k$ is equal to the difference of two sums:

$$p = \cos 0 + \cos\left(\frac{2\pi k}{n}\right) + \cos\left(\frac{4\pi k}{n}\right) + \cdots + \cos\left(\frac{(2n-2)\pi k}{n}\right)$$

and

$$q = \cos\left(\frac{\pi k}{n}\right) + \cos\left(\frac{\pi k}{n} + \frac{2\pi k}{n}\right) + \cos\left(\frac{\pi k}{n} + \frac{4\pi k}{n}\right) + \cdots$$
$$+ \cos\left(\frac{\pi k}{n} + \frac{(2n-2)\pi k}{n}\right).$$

Each of these sums is zero. Indeed, consider a regular n-gon in figure 5a (in which the case $n = 6$, $k = 2$ is illustrated). The vertices numbered $A_1, A_{1+k}, A_{1+2k}, \ldots$ are a subset of its vertices (in the figure, an equilateral triangle), and after a number of steps one of these points coincides with A_1. This set of vertices has rotational symmetry, so the sum of the vectors drawn from the center of the polygon to the marked vertices is equal to zero. (Rotation by $2\pi k/n$ about the center takes the polygon and the marked vertices into themselves, so the sum in question stays unchanged under this rotation, which is impossible for a nonzero sum.) Therefore, the sum of the projections of these vectors onto the x-axis is also zero. But this is just the sum p! A similar argument using figure 5b (where $k = 1$) proves that $q = 0$.

So it remains to compute the sum in expression (5) for $f(\alpha) = a_n \cos n\alpha$. In this case, $f(\pi k/n) = a \cos(\pi k) = (-1)^k a$, so our sum equals

$$\frac{a_n}{2n}\left[1-(-1)+1-(-1)+\cdots\right]=\frac{a_n}{2n}\cdot 2n=a_n,$$

and we're done.

A solution of Chebyshev's problem

Any labor should be rewarded: we're ready to solve Chebyshev's problem—that is, to find the least possible deviation from zero of a reduced polynomial (with leading coefficient 1) on a certain segment. Now it will be more convenient to consider the segment $[-1, 1]$.

First, let's estimate from below the deviation from zero of the trigonometric polynomial $f(\alpha)$ from equation (1). Consider the *mean absolute value* of $f(\alpha)$ at the points $k\pi/n$, $0 \le k \le 2n - 1$:

$$\begin{aligned}&\frac{1}{2n}\left[\left|f(0)\right|+\left|f\left(\frac{\pi}{n}\right)\right|+\left|f\left(\frac{2\pi}{n}\right)\right|+\cdots+\left|f\frac{(2n-1)\pi}{n}\right|\right]\\&\ge\frac{1}{2n}\left|f(0)-f\left(\frac{\pi}{n}\right)+f\left(\frac{2\pi}{n}\right)-\cdots-f\left(\frac{(2n-1)\pi}{n}\right)\right|\qquad(6)\\&=\left|a_n\right|.\end{aligned}$$

The inequality here follows from the fact that the absolute value of a sum is not greater than the sum of the absolute values of all its terms, and the equality follows from the lemma in the previous section. Now, the estimate in equation (6) implies that at least for one k,

$$\left|f\left(\frac{\pi k}{n}\right)\right|\ge\left|a_n\right|.$$

Therefore, *the deviation of a trigonometric polynomial from zero is not less than the absolute value of its leading coefficient.*

Now Chebyshev's problem is solved "in one move." Since the substitution $x = \cos\alpha$ turns an arbitrary reduced polynomial $p_n(x)$ of degree n into a trigonometric polynomial $f(\alpha)$ with leading coefficient $1/2^{n-1}$ (as we saw above), and x runs through the segment $[-1, 1]$ (from 1 to -1), as α varies from 0 to 2π, the deviation of $p_n(x)$ on $[-1, 1]$ equals that of $f(\alpha)$ on $[0, 2\pi]$. Therefore, *the deviation on $[-1, 1]$ of a reduced polynomial of the nth degree from zero is no less than $1/2^{n-1}$.*

We can see that this deviation is actually achieved for the polynomial $p_n(x)$ defined by the formula $p_n(\cos\alpha) = (1/2^{n-1})\cos n\alpha$—that is, for

$$p_n(x)=\left(\frac{1}{2^{n-1}}\right)\cos(n\arccos x).$$

In conclusion, we leave you with several problems to work through on your own.

Problems

1. A trigonometric polynomial of degree n has no more that n roots on $[0, \pi]$ and no more than $2n$ roots on $[0, 2\pi]$.
2. If two trigonometric polynomials of degree n take equal values at $n + 1$ points in the segment $[0, \pi]$, then they are equal everywhere.
3. The only reduced polynomial of degree n that deviates from zero on the segment $[-1, 1]$ by $1/2^{n-1}$ is the Chebyshev polynomial $(1/2^{n-1})\cos(n \arccos x)$. (Hint: when does equation (6) become an exact equality?)
4. Prove the identity

$$\begin{aligned}&\cos\alpha+\cos(\alpha+x)+\cos(\alpha+2x)+\cdots+\cos(\alpha+nx)\\&=\frac{\sin\left[\alpha+(n+1/2)x\right]-\sin(\alpha-x/2)}{2\sin x/2}.\end{aligned}$$

(Hint: $\sin[\alpha + (k + 1/2)x] - \sin[\alpha + (k - 1/2)x] = 2\sin(x/2)\cos(\alpha + kx)$.)

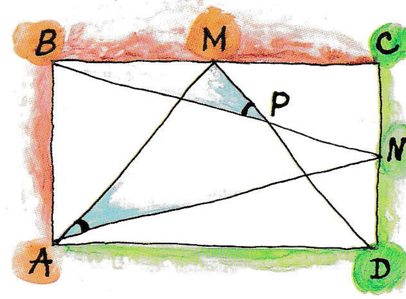
5. From problem 4, deduce the lemma on the mean value of a trigonometric polynomial.
6. Prove that a trigonometric polynomial without a constant term, $f(\alpha) = a_1\cos\alpha + a_2\cos 2\alpha + \ldots + a_n\cos n\alpha$, necessarily has a root. (Hint: what is the mean value of $f(\alpha)$)?
7. Recall the proof of the lemma, and make sure that the sequence of the vertices of the polygon returns to the start after $n/\mathrm{GCD}(n, k)$ steps. (GCD stands for "greatest common devisor.")

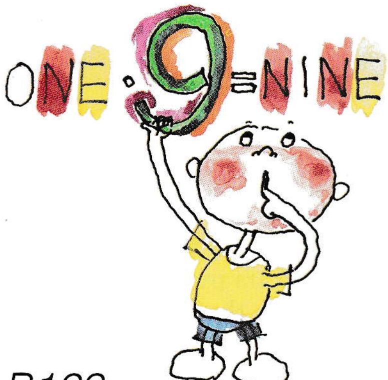
In this article we've touched upon the analytical properties of Chebyshev polynomials. Their combinatory properties are no less interesting—but that's a subject for another article.

Just for the fun of it!

B121

Equal angles appear. In a rectangle *ABCD*, *M* and *N* are the midpoints of *BC* and *CD* (see the figure), *P* is the intersection point of *DM* and *BN*. Prove that the angles *MAN* and *BPM* are congruent. (V. Proizvolov)



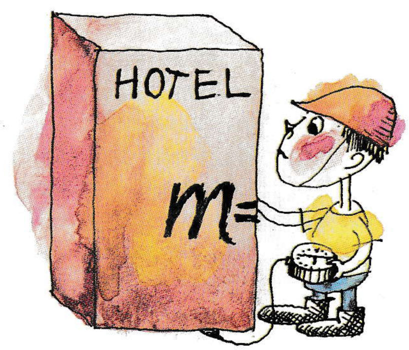


B122

Alphanumeric multiplication. Solve the number rebus ONE · 9 = NINE, where 9 is "nine" and each letter stands for one (and only one) of the other digits. (P. Filevich)

B123

Pascalmeter. John invented a pascalmeter—that is, a device to measure pressure. To show how it works, he came to a construction site where a new hotel was being built. He demonstrated that a brick laid flat exerted a pressure of 1,368 Pa; when set on its front side, it exerted a pressure of 2,581 Pa; and when set on its smallest face, it exerted a pressure of 5,404 Pa. A 4-m wall made of such bricks exerted a pressure of 88,200 Pa. What is the mass of one brick? (A. Pidora)



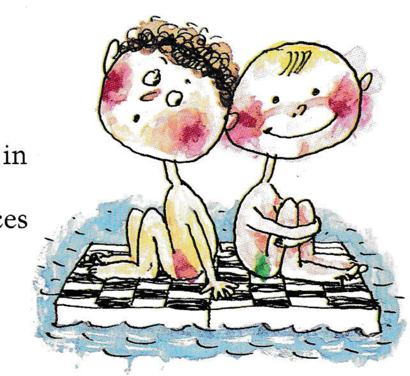
B124

Permutations in order. The language of the tribe of Robitecs consists of all possible permutations of the eight letters R, O, B, I, T, E, C, S, and no other words. When the chief of the tribe learned about such useful things as dictionaries, he instructed his court linguist to compile a dictionary of the tribe's language. The linguist wrote the name of the tribe as the first word in the dictionary, and, based on the order of the letters thus specified, began to order words in the usual way. What word did he write after *scretibo*? Before *biscrote*? After *iscetbor*? Can you offer a simple way to order the words in the dictionary? By the way, what's the last word?

B125

Unresignable position. The game of checkers often ends in a draw—even in Russia, where kings can move through any number of unobstructed squares (unlike American kings, who must plod square by square). Is a draw possible in the game of give-away checkers, played according to the usual Russian (or Continental) rules except that each player tries to get rid of all his or her pieces before the other? More exactly, does there exist a position in which neither player will lose unless a mistake is made? (A. Domashenko)

(See the box on page 61 for a full description of Continental Checkers.)



Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Optics for a stargazer

On seeing stars at noon: do the old tales hold water?

by Vladimir Surdin

THERE IS AN OLD AND RATHer widespread belief that the stars can be seen even in the daytime if the observer looks for them from the bottom of a deep well. Sometimes you can come across this assertion in authoritative sources. More than a thousand years ago Aristotle wrote that the stars can be seen during the day from inside a deep cave. Later the Roman historian Pliny repeated this belief, replacing the cave with a well. Many writers have mentioned this phenomenon in their works: Rudyard Kipling, for instance, wrote about the stars being visible at noon from the bottom of a deep ravine. In his book *Star-Land* (Boston, 1889) Sir Robert Ball gives a detailed account of how to see the stars in the daytime from the bottom of a tall chimney (fig. 1), ascribing this ability to the fact that human sight becomes more keen in a dark tube.

So, is there any truth to the matter? Has anyone researched the question of seeing stars during the day? I must admit that I have yet to be offered the opportunity to go down into a very deep well or crawl into a tall chimney. Nevertheless, at various times there have been inquisitive people trying to prove the "well phenomenon." The famous German naturalist and traveler Alexander von Humboldt tried to see the stars in the daytime from deep pits in America and Siberia, but to no avail. Adventurous individuals are still to be found among us. To name just one, let's see what a reporter from *Komsomolskaya Pravda*, L. Repin, wrote on May 24, 1978: "People say that it is possible to see the stars in the sky in broad daylight from the bottom of a deep well. Once I tried to verify it and went down into a 60-meter well, but I did not see any stars—only a little square of blindingly blue sky."

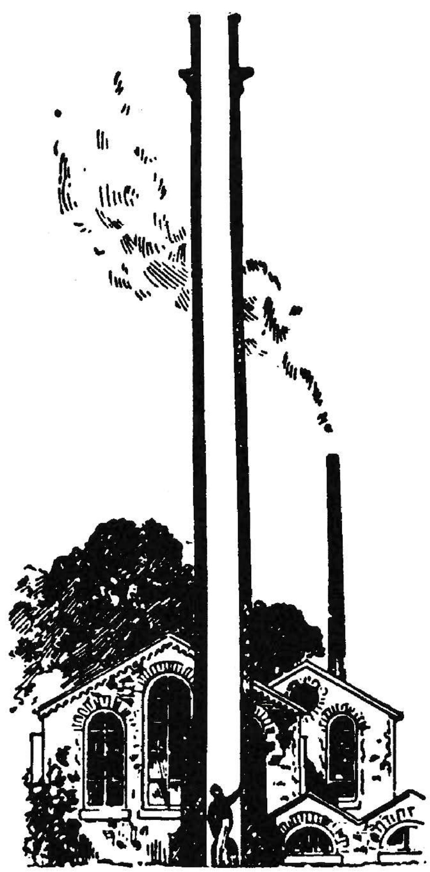


Figure 1

Illustration from the 1899 edition of Sir Robert Ball's 1889 book Star-Land. *The caption read: "How the stars are to be seen in broad daylight."*

Then there's the evidence of Richard Sanderson, an experienced amateur astronomer from the town of Springfield, Massachusetts, who described his observations in *The Skeptical Inquirer* (1992):

> About 20 years ago, when I worked as a trainee in the planetarium of the Springfield Museum of Science, I talked with some of my colleagues about this old popular belief. Our discussion was overheard by the director of the museum, Frank Corcosh, who proposed that we solve the problem by experiment. He led us to the basement of the museum, where the base of a tall, thin chimney was located. A small door led into the shaft, where we managed to insert our inquisitive heads. I still remember the feeling of excitement connected with the prospect of seeing the nocturnal luminaries in the daytime.
>
> Looking up along the flue I saw a bright circle standing out against a background of the pitch darkness of the chimney's interior. The surrounding darkness made my pupils widen, and the patch of sky shone even brighter.
>
> I immediately realized that I could never see the stars in broad daylight with such a "device." When we had climbed out of the basement, Corcosh

Art by Sergey Barkhin

C694

remarked that only one star could be seen in the daytime in good weather—the Sun.

Thus, the nocturnal luminaries can't be seen in the daytime from either a deep well or tall tube. Still, let's not be too hasty: there are some tubes that make it possible to see stars even in broad daylight. I'm talking about astronomical tubes—telescopes. What makes them so special? Why do tubes with lenses allow us to see the stars during the day, while empty ones don't?

First of all, let's think things through: why aren't the stars visible in the daytime? Simply because the sky is so bright due to the scattered sunlight. If the scattered light becomes weaker for any reason (for instance, during a total eclipse of the Sun), the bright stars and planets are stunningly visible in the daytime. They are also clearly seen in outer space and from the surface of the Moon. So why does the scattered sunlight in the atmosphere hide them from us? After all, the light from the stars doesn't become any weaker.

To understand this phenomenon, we need to know how we see. As you know, the eye's lens forms an image on the back of the eye—on the light-sensitive layer called the retina. The retina is composed of a large number of elementary light receivers—cones and rods. They differ in their sensitivity to color, but for our purposes this doesn't matter, so we'll just refer to them as "cones." The important thing is that each cone sends information to the brain about the flux of light landing on it, and the brain synthesizes an overall picture from these individual signals.

The eye is an extremely complex instrument for collecting information, but in a way it's similar to a "smart" electronic device like a radio. The eye has an automatic amplification system that decreases its sensitivity in bright light and increases it in dim light. It also has a system of noise suppression that smoothes the random fluctuations in the light flux both over time and over a number of adjacent cones on the retina. This system has certain threshold properties, so the brain doesn't see rapid changes in the image (this is the idea behind movies) or small fluctuations in luminosity.

When we look at a star at night, the light flux landing on a single cone, though small, is still much greater than the flux of the dark sky falling on the adjacent cones. So the brain interprets it as a meaningful signal. But in the daytime every cone is illuminated by so much light from the sky that a little extra light from a star landing on a cone is not perceived by the brain as a real difference in the light flux and the brain "writes it off" as a fluctuation. A star can become visible against the daytime sky only when its light flux is comparable to that coming from a patch of sky projected onto a single cone. The angular magnitude of this patch is known as the resolving power of the human eye and is equal to about one minute of arc.

Among the heavenly bodies only Venus is sometimes seen in the daytime sky. It's not an easy thing to pull off: the sky must be especially clear, and you need to know approximately where the planet is located in the sky. All the other planets and stars are far less bright than Venus, so it's absolutely impossible to see them in the daytime without a telescope. However, some astronomers assert that under ideal conditions they can see Jupiter in the daylight, and Jupiter is about a fifth as bright as Venus. As for the brightest star in our sky, Sirius, nobody has seen it in the daytime at sea level. They say you can see it in the mountains against the dark-violet sky.

It's easy to show how the bright background masks the luminous points. Here's what Yakov Perelman recommended in his book *Astronomy for Fun*:

> A simple experiment can demonstrate the disappearance of the stars in daylight. In the side wall of a cardboard box punch a dozen holes in the form of a constellation, then stick a sheet of white paper on it from the outside. Put the box in a darkened room and illuminate it from the inside. You can clearly see the illuminated points on the perforated side like stars in the night sky (fig. 2). But if you turn on another light in the room that's bright enough, the artificial stars on the paper disappear without a trace: "daylight" has extinguished the "stars."

So how does a telescope help us to see the heavenly bodies by day? Obviously its objective (set of lenses) collects much more light than the pupil of the eye can. But in this respect the images of a star and a patch of sky are equivalent, because their light fluxes are increased by the same

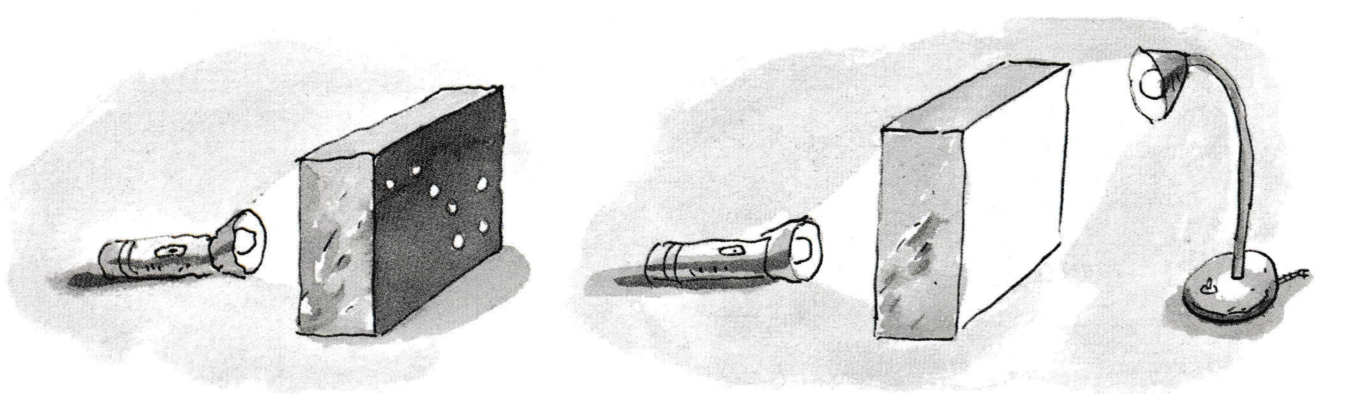


Figure 2

factor, which is approximately equal to the ratio of the objective's area to the area of the pupil. Something more important is at play here: the telescope improves the resolving power of the eye because it magnifies the angular size of the objects observed. This means that the telescope projects the same patch of sky on a larger number of cones, each of which absorbs proportionally less light. For example, if a telescope increases the angular size by a factor of A, the observed luminosity decreases by a factor of A^2. However, a star has a very small angular size, so its light still falls on a single cone even after passing through the telescope. But now the extra light from the star can be rather significant in comparison with the decreased luminosity (per cone) of the background sky. Thus, with a magnification of 45x, the effective luminosity of the sky is decreased by a factor of $45^2 \cong 2{,}000$, which makes it possible to see the brightest stars and planets against the daytime sky.

So does this mean we can just take a telescope with a high magnification and see the dimmest stars by day? No. The Earth's atmosphere isn't homogeneous, so a star's image is blurred and has a certain angular magnitude, though very small. At night and in good weather in the mountains, it's about 1 arc second, and in the daytime at sea level it's no less than 2–3″. Therefore, if a telescope has a magnification of more than 30–60x, the angular size of a star exceeds the eye's resolving power and its image is projected onto several cones. So there's no point in increasing the magnification: the star's brightness will decrease as much as that of the sky.

Let's figure out which stars can be seen during the day with a telescope. In good weather the daytime sky has a magnitude of –5 per square arc minute,[1] corresponding approximately to the light falling onto one cone. The magnitude of Venus is about –4. Therefore, we assume that a star will be visible if its luminosity differs from that of a square arc minute of sky by more than 1. As we found earlier, we can decrease the brightness of the sky with a telescope by a factor of no more than 2,000—that is, by a luminosity factor of about 8. So the sky's luminosity will decrease from –5 to $-5 + 8 = 3$ per square arc minute, which means we can see stars with a magnitude as low as 4. Astronomical observations confirm this estimate.

Now that we've settled the telescope question, let's return to the well. Can a well decrease the sky's luminosity for an observer at the bottom? In principle, yes—though not with lenses, but geometrically, by removing the entire field of vision except for a small patch of sky, which sends a light flux comparable to that of a star. To do this, the well's opening must appear to the observer to have an angular magnitude of no more than 1′. If the well's diameter is 1 m, its depth must be more than $1\text{ m}/\sin 1' = 3.4$ km! Even in this case the observer will see only a bright point whose brightness increases momentarily when a star passes directly overhead. Even by a stretch of the imagination it's difficult to consider this procedure as "an observation of the stellar sky," to say nothing of finding such a well. As for the probability that a bright star will pass directly overhead (±0.5′), I'll let you estimate that yourself, but I can tell you that you'll have to wait more than a thousand years for that marvelous moment!

Generally speaking, a tall tube can also play a role in the daytime observation of the stars. It creates a channel of air in which there is practically no scattering of sunlight. If such a tube passed all the way out of the atmosphere, we could observe the night sky at any time of day! Almost the entire mass of air is confined near the Earth's surface, reaching a height of about 20 km. The tube would have to be very tall indeed!

So we are led to conclude that the daytime sighting of stars is simply a myth. Where did it come from? We can only speculate. It's possible that someone at the bottom of a deep pit actually did see Venus in the sky. Such an occurrence has a very small probability and is possible in principle only in the tropics, where Venus passes directly overhead. It's more likely that people who went down into a deep well or entered a cave saw dust particles lit by sunlight against the dark walls. Maybe they mistook them for stars?

Still, it's too early to consider the case closed. We need to look more carefully at optical illusions, at unexpected combinations of natural conditions, at rare physical effects. It is you, our esteemed readers, who can provide valuable help in the matter.

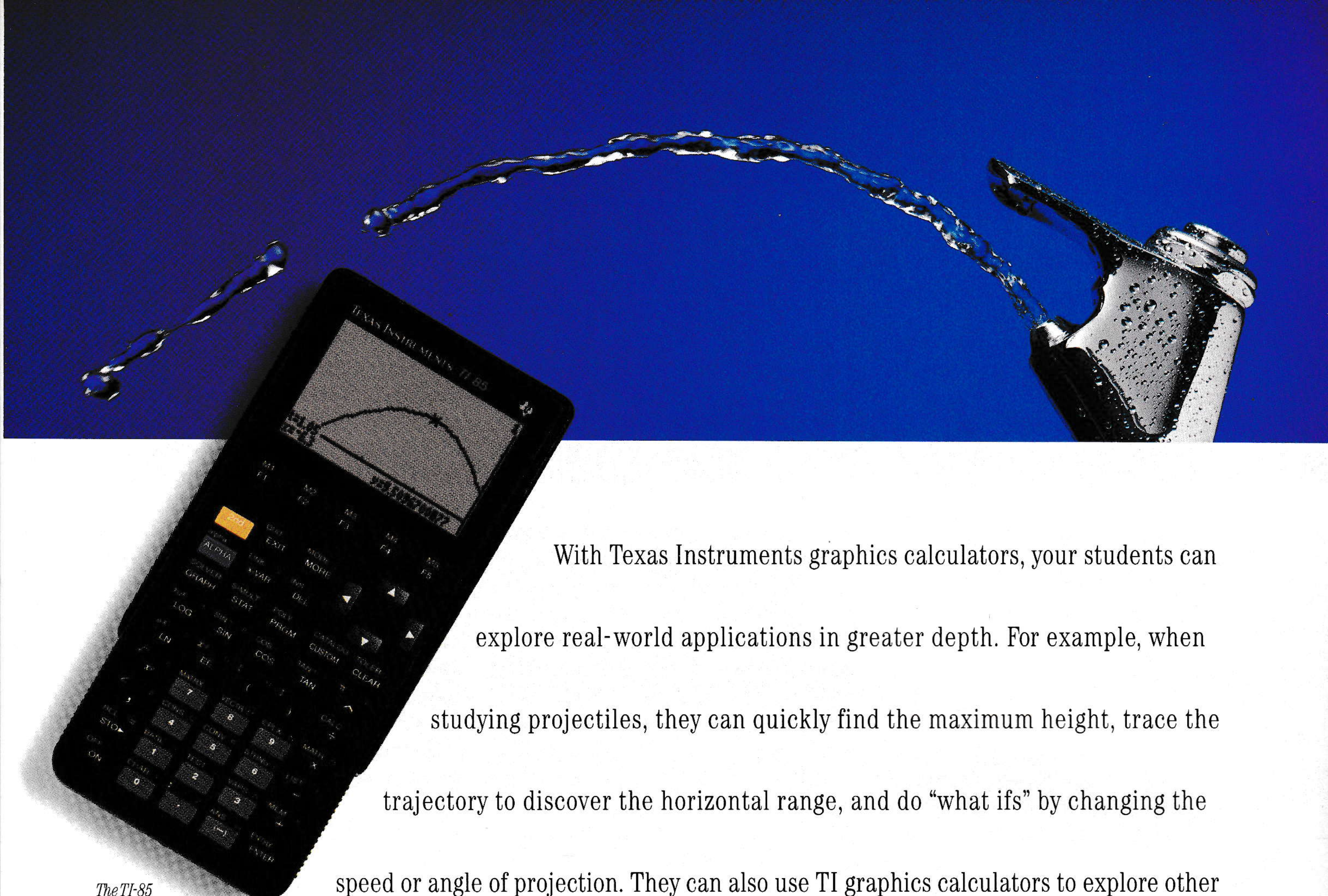
Indeed, Ramiro Cruz, an amateur astronomer from Houston, Texas, decided to clear up for himself the rumors that Sirius can be seen in the daytime. He looked for the star in the southwest part of the sky in April 1992 not long before sunset. Take note of the fact that he knew where to look! With the naked eye he managed to spot Sirius no earlier than 21 minutes before sunset. When he used binoculars with a magnification of 70×50, he found the star 43 minutes before sunset.[2] We have enough data to estimate the sky's luminosity at the moment he sighted the star.

Houston is situated at 30° N latitude, so the celestial equator crosses the horizon there at an angle of $90° - 30° = 60°$. As the observations were made just after the vernal equinox, the Sun was near the equator and approached the horizon at the same angle. It takes the Sun one minute to pass an arc of $360°/(24.60) = 0.25°$ in

[1]Astronomers measure the luminosity of heavenly bodies by their "stellar magnitude," which is denoted as (for example) –5. A decrease in stellar magnitude of 1 corresponds to an increase in brightness by a factor of 2.5. Most of the stars we can see on a clear, dark night range in magnitude from 6 (for the dimmest) to 1 (for the brightest.)

[2]See *Sky and Telescope*, Vol. 85, No. 2, Feb. 1993, p. 112.

CONTINUED ON PAGE 43

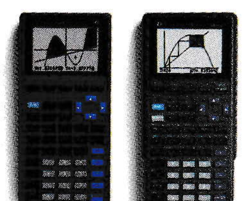


The TI-85 Graphics Calculator

With Texas Instruments graphics calculators, your students can explore real-world applications in greater depth. For example, when studying projectiles, they can quickly find the maximum height, trace the trajectory to discover the horizontal range, and do "what ifs" by changing the speed or angle of projection. They can also use TI graphics calculators to explore other physics applications, including kinematics, dynamics, work and conservation laws, and circuits. And TI

They'll Understand More Because They See More.

graphics calculators are more accessible than a computer lab and easier to use. The real-world applications you're teaching are right under their nose. Give your students the handheld power to solve these real-world problems and make physics just as accessible. For information on TI's full line of graphics calculators, call 1-800-TI-CARES. Ask for our TI-85 physics applications book!



The TI-81 and TI-82 Graphics Calculators

In Canada, call 1-800-661-2007.

TEXAS INSTRUMENTS

Circle No. 2 on Reader Service Card

Challenges in physics and math

Math

M121

Symmetric inequality. For any two integers m and n greater than 1, prove the inequality

$$\frac{1}{\sqrt[n]{m+1}} + \frac{1}{\sqrt[m]{n+1}} > 1.$$

You may want to use the following modification of Bernoulli's inequality: for any $x > 0$ and $0 < \alpha < 1$, we have $(1 + x)^{\alpha} < 1 + \alpha x$. (L. Kurlyandchik)

M122

Combining ones and twos. Prove that for any positive integer n there is a number composed only of the digits 1 and 2 that is divisible by 2^n. (B. Ivlev)

M123

Black out. Several unit squares of an infinite square grid are colored black. Prove that it is possible to cut out a number of (non-unit) squares such that (1) they will cover all the black unit squares and (2) in each of them the black squares will cover no less than 1/5 and no greater than 4/5 of its total area. (G. Rozenblume)

M124

Watching a snail. A group of naturalists observed a crawling snail for $t > 1$ minute. Each of them watched the snail during exactly 1 minute and noticed that it crawled a distance of exactly 1 m during this period. The observation was never interrupted. What longest and shortest path could the snail crawl during these t minutes? You may start with small values of t—say, $t = 2.5$ minutes. (N. Konstantinov)

M125

Remarkable line in a quadrilateral. A quadrilateral has both circumscribed and inscribed circles. Prove that the intersection point of its diagonals and the centers of the circles lie on the same straight line. (V. Protasov)

Physics

P121

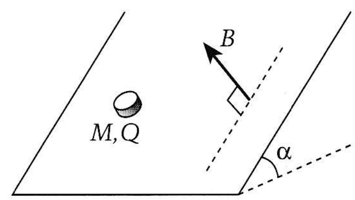
Taking off. A small airplane with its engine shut off can glide down with a minimum speed $v = 150$ km/h at an angle $\alpha = 5°$ with the horizon. (If the pilot reduces this speed or this angle, the plane goes into a spin.) What minimum thrust must the engine produce for the plane to take off from a horizontal runway? In each case the airplane's velocity is directed along the fuselage. The mass of the plane $m = 2{,}000$ kg. (A. Andrianov)

P122

Gas and piston. One mole of an ideal monatomic gas is contained under a massive piston at a temperature T_0 in a thermally isolated vertical cylinder. The gas is compressed by pressing down on the piston. After performing work W, the piston is released and assumes a new equilibrium position. Find the temperature T in this state. (V. Uzdin)

P123

Puck in a magnetic field. A small puck of mass M carrying a charge Q lies on a plane inclined at an angle α. The coefficient of friction is μ. There is a magnetic field B perpendicular to the plane (see the figure below). The puck is released with



zero initial velocity. Find the steady-state velocity of the puck, including its direction. (A. Alexeyev)

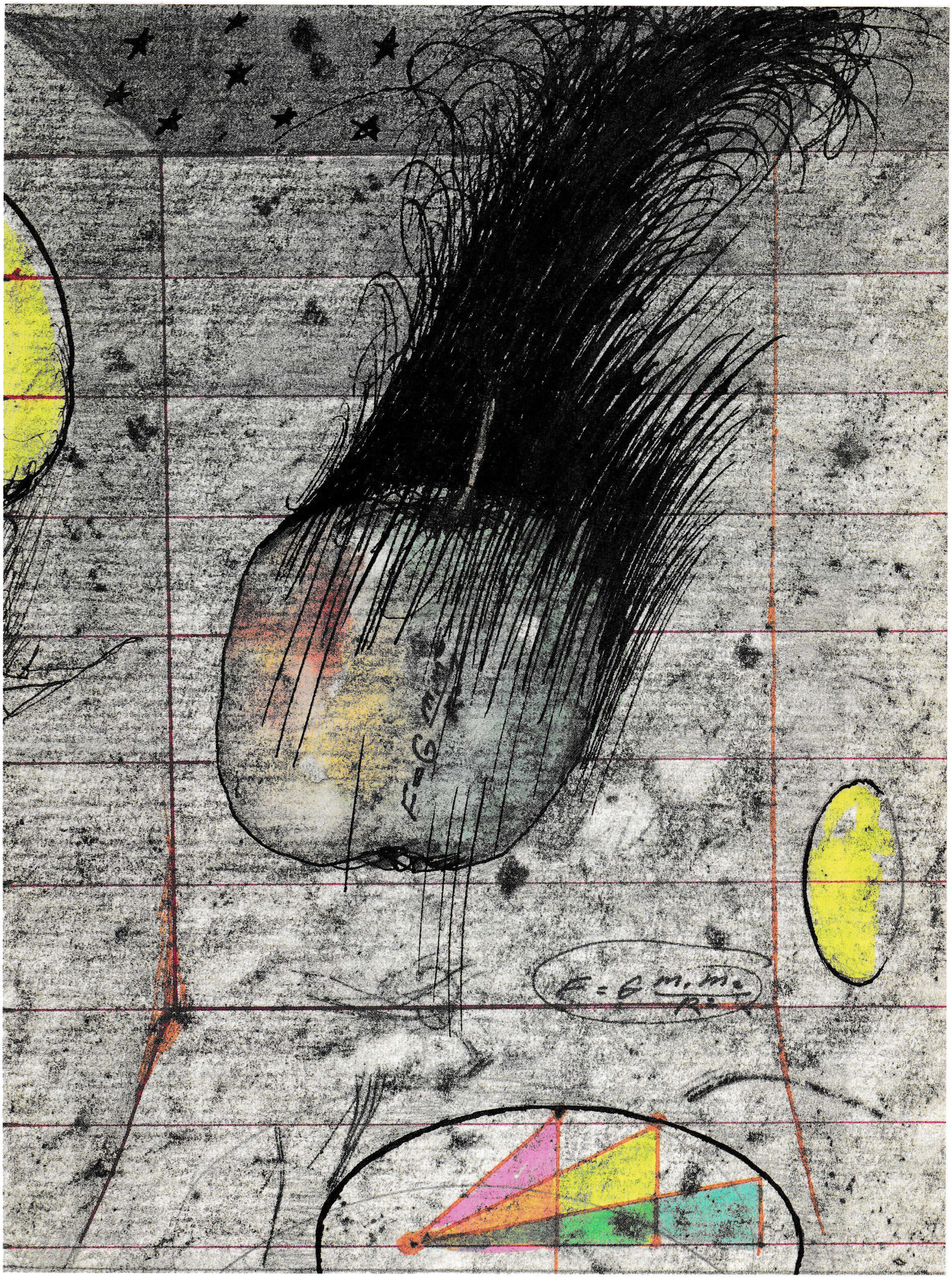
P124

Most powerful heater. You have a 1-ohm, a 2-ohm, and a 3-ohm resistor. Each of them can release no more than 1 W of power. How should they be connected and what voltage should be supplied to obtain a heater with the most total power? (A. Zilberman)

P125

Conical light conductor. A light conductor in the form of a truncated cone is made of glass coated on the inside with silver (for better reflection of the incident rays). The planes of the bases of the cone are perpendicular to its axis, their diameters are D and d, and the height of the cone is H ($H >> D >> d$). A beam of light parallel to the cone's axis falls on the large base. Will all the rays emerge from the small base after undergoing multiple reflections? (S. Pankov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 57



Why study mathematics?

What mathematicians think about it

by Vladimir Arnold[1]

WHY MUST WE STUDY mathematics? In 1267 this question was answered by the English philosopher Roger Bacon: "[H]e who ignores it can not know the other sciences nor the things of this world. . . . And what is worse, men who are ignorant of it do not perceive their own ignorance, and therefore seek no remedy for it." At this point I could have ended my lecture, but people think that maybe something has changed over the last seven centuries . . .

Let's look at more recent evidence. One of the creators of quantum mechanics, Paul Dirac, says that when you build a physical theory, no physical conceptions should be trusted. So what *is* to be trusted? According to the famous physicist, only a mathematical scheme—even if, at first glance, it isn't connected with physics.

Indeed, physics has cast aside all the purely physical conceptions prevalent at the beginning of this century, while the mathematical models incorporated into the physicists' arsenal gradually acquire physical meaning. Here the stability of mathematics clearly manifests itself.

Thus, mathematical modeling is a fruitful method of understanding in the natural sciences. We'll approach mathematical models from another angle, examining the problems of mathematical education.

Three approaches to teaching math

In Russian mathematical education (both secondary and higher) we follow the European system based on the "Bourbakization" of mathematics. (Nicolas Bourbaki is the pseudonym of a group of French mathematicians who, since 1939, have been publishing a series of books in which the principal divisions of modern mathematics were presented formally—that is, by means of the axiomatic method based on set theory.)

The formalization of mathematics leads to a certain formalization of its teaching, demonstrating the costs of Bourbakization in mathematical education. Here's a dramatic example. Second-year pupils at a French school are asked, "How much is two plus three?" The answer follows, "Since addition is commutative, it's three plus two."

A truly remarkable answer! It's absolutely correct, but the pupils didn't even think about simply adding these two numbers, because their instruction laid stress on the properties of operations. In Europe, educators had already become aware of the shortcomings of this approach and began to back off from Bourbakization.

In the last few years, Russian mathematical education had undergone Americanization. It's based on the principle: teach what's needed for practical application. So someone who doesn't think he or she will need mathematics need not study it at all. Mathematics is optional for high school juniors and seniors—for instance, one third of high school seniors don't take algebra. The effect is illustrated by the following example. A test for 14-year-old American students asked them to estimate (not compute, but just estimate) what happens to the number 120 if we take 80 percent of it. Three versions of an answer were offered: it will increase, remain the same, or decrease. The right answer was chosen by about 30% of the tested students. This means that they checked answers haphazardly. The conclusion: nobody knows anything.

The second particular feature of the American approach to teaching math is computerization. The fascination with computers by itself doesn't contribute to the development of thinking ability. Take another example from an American test.

There are 26 students in a class. They are going on a car trip. A car can take one parent and four students. How many parents should be invited to help?

[1] These are slightly abridged notes from a lecture given by Arnold, one of the leading mathematicians of our time, at the Republic Institute for Advancing the Qualifications of Educators in Moscow on April 16, 1992. The notes were submitted by Prof. Y. Fominykh of Perm.—*Ed.*

Art by Yury Vashchenko

A typical answer is 65 parents. A computer gives out $26 \div 4 = 6.5$. And a student already knows that if the solution must be an integer, one should do something with the decimal point—for instance, just throw it away.

And now look at an example from an official examination in 1992 for students:

Which of the following pairs most closely resembles the relation between angle and degree—

(a) time and hour;

(b) milk and quart;

(c) area and square inch

(and so on)?

The answer is area and square inch, because the degree is the minimal unit of angle measure, and the square inch is the minimal unit of area, whereas an hour can be divided into minutes.

The authors of this problem were obviously taught according to the American system. I'm afraid we'll soon get to this level, too.[2] It's only surprising that there are so many outstanding mathematicians and physicists in the United States.

Today our mathematical education slowly turns from the European system to the American. As always, we are late, lagging behind Europe by about 30 years, so 30 years later we should be ready to set things aright and get out of the dead-end where we'll be driven by the American educational system with its pragmatism, optionality, and mass computerization.

Our traditional mathematical education used to be at a higher level and was based on the culture of arithmetic problems. Even as recently as twenty years ago some families still had copies of the old "merchant" problems. Now it's all gone. The algebraization of the last reform of mathematical education turns students into robots. It is the arithmetic approach that demonstrates the "meatiness" of the math we teach.

Consider, for example, these problems:

(1) There are three apples. One is taken away. How many are left?

(2) How many cuts with the saw are needed to divide a log into three pieces?

(3) Boris has three more sisters than he has brothers. How many more girls than boys are in his family?

From the point of view of arithmetic these are all different problems—their content is different. And the intellectual effort needed to solve the problems is quite different, too, although the algebraic model is always the same: $3 - 1 = 2$. What strikes you first of all in mathematics is the surprising universality of its models and their incomprehensible effectiveness in applications.

As the great Russian poet Vladimir Mayakovsky said: "The man who first formulated that 'two and two is four' was a great mathematician, even if he obtained this truth from adding two cigarette butts to two cigarette butts. All his followers, even though they may have added immeasurably greater things—say, locomotives to locomotives—are not mathematicians." To "count locomotives" is the American way of teaching math. It is disastrous. The example of the development of physics at the beginning of this century shows that locomotive mathematics turned out to be worse than cigarette-butt mathematics: applied mathematics couldn't keep pace with physics, while theoretical mathematics supplied everything physicists needed for further development of their science. Locomotive mathematics lags behind practice: while we teach how to calculate with abacuses, computers appear. We must teach how to think, not how to push buttons.

Admittedly, a mathematical model doesn't always give immediate practical returns. Sometimes it may prove useful only after two thousand years. That's what happened with *conic sections*.

[2]A New York professor, Joe Birman, explained to me that for him as an American the "correct" solution to this problem was absolutely clear. "The point is," he said, "I can imagine precisely the level of idiocy of the author of these problems."

Conic sections and the law of gravity

Conic sections were discovered in ancient Greece and described by Apollonius of Perga (265–170 B.C.) in an eight-volume treatise. But the need for this theory arose only in the 16th century, when Johannes Kepler was deriving his laws of planetary motion.[3] His teacher, Tycho Brahe, had scrupulously measured the positions of the planets of the solar system in the observatory at Uraniborg over the course of 20 years. After his teacher's death, Kepler got down to the mathematical processing of the results of these observations and found that the trajectory of Mars, for instance, is an ellipse.

An ellipse is the locus of points such that the sum of the distances from these points to two given points, called the foci, is constant. A remarkable theorem—which, unfortunately, is not proved at school—says that the section of a cone by a plane tilted at a large enough angle to its axis is an ellipse. Its proof is pretty simple (see figure 1). The two spheres inscribed in the cone and

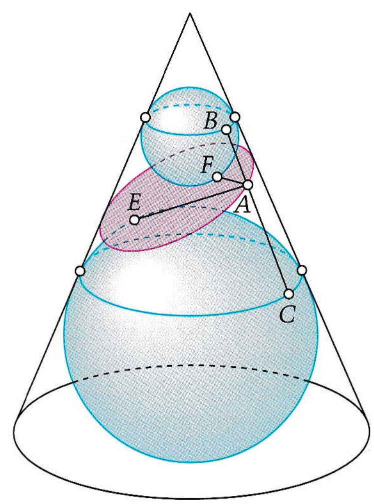


Figure 1

Ellipse with foci F and E and Dandelin spheres. Two tangents drawn to the same sphere from the same point are congruent, so FA + EA = BA + AC = BC, a constant.

[3]See also "An Act of Divine Providence" in the May/June 1993 issue of *Quantum* and "The Fruits of Kepler's Struggle" in the January/February 1992 issue.—*Ed.*

touching the plane (at the foci E and F of the ellipse in the section) used in this proof are called Dandelin spheres.

To understand the chain of Kepler's reasoning, we'll need a few simple facts about the geometry of the ellipse. It can be shown that the length of the *major semiaxis* of an ellipse OK (fig. 2), usually denoted a,

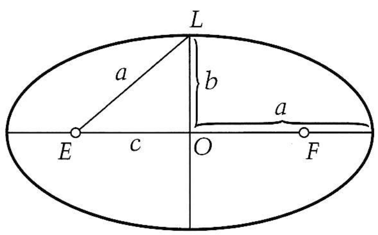


Figure 2

Foci, semiaxes, and eccentricity of an ellipse.

is equal to the length of the hypotenuse EL of the triangle with legs $b = OL$ (the *minor semiaxis*) and $c = EO$. The ratio c/a characterizes the ellipse's shape and is called its *eccentricity,* because it's proportional to the displacement of the foci from the center of the ellipse. The eccentricity is usually denoted by e.

By the Pythagorean Theorem, the ratio of the semiaxes equals $b/a = \sqrt{1-e^2} \cong 1 - e^2/2$ for small e. It follows that an ellipse with a small eccentricity is virtually indistinguishable from a circle. For instance, if $e = 0.1$, then the minor axis is shorter than the major axis by only 1/2,000 of the length of the latter. For such an ellipse with major axis 1 meter long, the minor axis will be shorter by only half a centimeter, yielding an unnoticeable difference between such an ellipse and a circle. However, the foci are 5 cm apart from the center, which is quite noticeable.

The formula $b/a = \sqrt{1-e^2} \cong 1-e^2/2$ (which means that the longer leg of a stretched right triangle is practically as long as the hypotenuse and gives a good approximation of the difference between their lengths) is one of the most remarkable facts in all of mathematics. (Unfortunately this isn't taught in school.)

For example, suppose you're

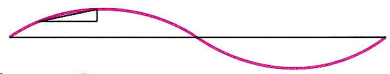


Figure 3

How much longer is a sine curve than a straight line?

coming back home along a sine curve. How much will your path be longer than the straight one (fig. 3)? The first impression (twice as long) exaggerates the length, of course. Yet it seems that the curved path will be about half again as long as the straight length. In actual fact, it's only about 20% longer. The reason is that the greater part of the sinusoid is only slightly tilted toward the axis, so the corresponding hypotenuses are barely longer than the legs.

Here is another application of this formula. The engines of the first jet planes were attached to the wings near the fuselage, so the air rushing from the engines was harmful to the tail assembly. The designers, who knew and felt the formula we've been examining, turned the engines by a small angle α (fig. 4).

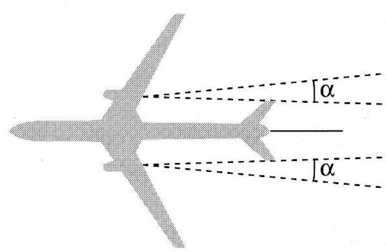


Figure 4

Saving the tail.

The tail assembly was saved (the deviation of the stream of air is proportional to α), while the net thrust remained practically the same (the loss is approximately $\alpha^2/2$, where α is measured in radians—for an angle of 3° only about 1/800 of the power is lost).

Let's turn back to Kepler. First he thought that the orbit of Mars is a circle. But the Sun happened to be offset by about 0.1 of the orbit's radius from the orbit's center. Kepler didn't stop at this result (remarkable in itself)—because he knew the theory of conic sections. Kepler knew that an ellipse with a small eccentricity looks very much like a circle, and he examined how the small remaining deviation of the orbit from a circle behaved. It's interesting that this verification was made possible only by the exceptional precision of Tycho Brahe's observations, performed with the naked eye. At the time, astronomers did not trust telescopes much, and even later, at the end of the 17th century, it still had to be proved that telescopic observations could attain as high a precision as those with the naked eye.

New physics often begins with refinement of the last significant digit of the previous theory. If Kepler had been satisfied with an eccentric circular orbit, or if Tycho Brahe's observations were less exact, the development of celestial mechanics (and, perhaps, of all of theoretical physics) could have been delayed, perhaps even for centuries.

The orbit of Mars turned out to be slightly oblate in the direction perpendicular to its diameter passing through the Sun by approximately half a percent—that is, by $e^2/2$. This led Kepler to the idea of elliptical planetary orbits.

If the theory of conic sections had not been worked out by mathematicians in advance, certain fundamental laws of nature wouldn't have been discovered in time, modern science and technology would not have arisen, and our civilization would have remained at the medieval level. Or, at the very least, the paths of history would have been totally different.

Kepler discovered the laws of planetary motion, but the fact that planets actually move in ellipses was proved by Sir Isaac Newton in his book *Mathematical Principles of Natural Philosophy* (1687), which laid the basis for all of modern theoretical physics. He derived from his law of universal gravitation the fact that planetary orbits are ellipses. Note that before Newton this problem was examined by his contemporary Robert Hooke. He studied the law of a body's motion in a gravitational field assuming that the force of gravity is in inverse proportion to the square of distance. Having approximately integrated the equation of the motion, Hooke drew orbits and discovered that they look like ellipses. His scientific honesty didn't allow him to call them ellipses, and he couldn't prove they were ellipses. So Hooke called them ellipsoids and proposed that Newton prove that Kepler's first law (that planets move in ellipses) follows from the law of inverse squares. Newton, who knew the ancient theory of conic sections very well, met this challenge by means of intricate constructions based on elementary geometry.

Later, second-order curves began to appear in scientific research more and more often. Why did this model prove so fruitful in application? Why, in particular, does the conic-section model describe planetary motion? It's a mystery. An enigma. There's no answer to this question. We believe in the power of rational science. Newton saw here a proof of God's existence: "This most graceful combination of the Sun, planets, and comets could not happen other than by the intention and the power of the mightiest and wisest creature. . . . It governs all not as the world's soul, but as master of the Universe, and by its supremacy should bear the name of God Almighty."

Modern space explorers also use the properties of conic sections when they plan the launching of satellites. So the foundation for modern physics and the scientific and technical revolution was established by a classical work by Apollonius. Yet the renowned Greek was thinking only about the beauty of the mathematical model when he investigated conic sections.

Computers, quantum mechanics, and Riemann surfaces

Another example is the story of the creation of computers. Long before the first computer came to being, its two principal mathematical components were lying in wait in mathematics: mathematical logic (specifically, Boolean algebra, created by George Boole, 1815–1864) and the schematic design of a computing machine. The first adding machine was designed by the French mathematician Blaise Pascal in 1641.

A third example is the development of wave mechanics by Erwin Schrödinger. By the time Schrödinger took up the oscillation problem, Werner Heisenberg's matrix version of quantum mechanics was already known. It wasn't clear how a discrete rather than a continuous spectrum could be obtained from the theory of waves in all of space. Schrödinger got help from the well-known German mathematician Hermann Weyl. Without Weyl's results in spectral theory on an infinite interval, we would never have heard about the famous Schrödinger equation. It's the same story again: a mathematician turns up with a theory ready to apply—that of boundary conditions at infinity—and all that was left to do was use it.

The next example is Riemann surfaces. They were introduced by the German mathematician Bernhard Riemann in the middle of the last century. These are the surfaces obtained by correspondingly cutting and gluing a number (or even an infinite number) of planes of a complex variable. Topologically, such a surface can be a sphere, a sphere with handles (fig. 5), and so on. The theory of Riemann surfaces was developed as part of the theory of functions of a complex variable. Later they proved useful in completely different problems. For instance, elliptic integrals get a simple geometric treatment on Riemann surfaces.

Karl Gustav Jacobi proved that Riemann surfaces "govern" two other problems:

1. Determining the number of ways of representing a given integer as the sum of four squares

$$N = x^2 + y^2 + z^2 + u^2;$$

2. Investigating oscillations of a pendulum, which leads to the differential equation

$$x'' = -\sin x.$$

Trig functions and counting snakes

As a fifth example, consider the so-called Bernoulli–Euler triangle:

$$\begin{array}{ccccccccccccccc}
&&&&&&&1&&&&&&& \\
&&&&&&1&&0&&&&&& \\
&&&&&0&&1&&1&&&&& \\
&&&&2&&2&&1&&0&&&& \\
&&&0&&2&&4&&5&&5&&& \\
&&16&&16&&14&&10&&5&&0&& \\
&0&&16&&32&&46&&56&&61&&61& \\
272&&272&&256&&224&&178&&122&&61&&0
\end{array}$$

It's filled in as follows. In line number 0, we write 1. Every odd line (first, third, and so on) is filled out from the right by writing in each position the sum of all the numbers in the previous line to the right of this position. Even lines are filled out in the same way but from the left.

The miracle concealed in this triangle was discovered a hundred years ago. The clue to it is given by this "simple" theorem (mathematicians

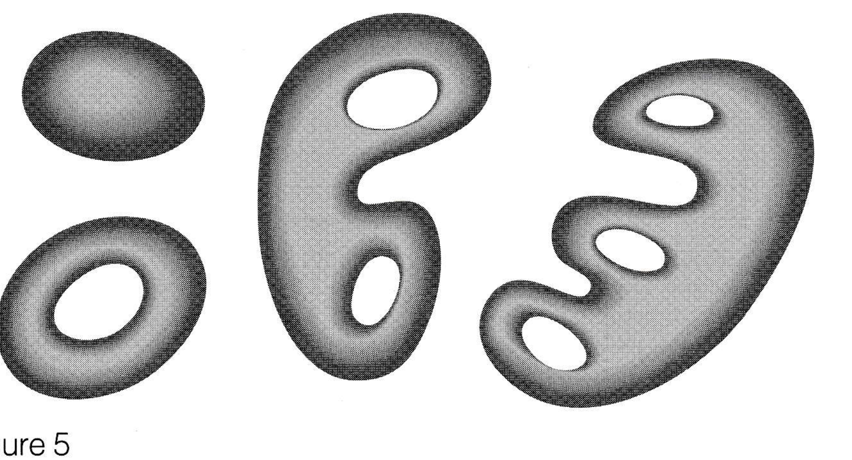


Figure 5
Riemann surfaces.

often disguise the fact that everything is quite simple):

$$\sec t + \tan t = \sum_{n=0}^{\infty} k_n \frac{t^n}{n!},$$

where the coefficients k_n are the numbers on the sloping sides of the Bernoulli–Euler triangle (k_n is the nonzero extreme number in the nth line).

It follows that the coefficients along the left slope yield the expansion into a power series of the function $\tan x$ (it's odd, and so its expansion contains only odd terms):

$$a_1 = \frac{k_1}{1!} = 1, \quad a_3 = \frac{k_3}{3!} = \frac{2}{6} = \frac{1}{3},$$

$$a_5 = \frac{k_5}{5!} = \frac{16}{5!} = \frac{2}{15}, \quad \ldots$$

and

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \cdots.$$

Similarly, the right slope yields the expansion of the secant function.

The Bernoulli–Euler triangle delivers the topological classification of the real polynomials $x^{n+1}\, a_1 x^n + \ldots + a_{n+1}$ all of whose n critical values (local extrema) are real and different.

The graph of such a polynomial resembles a snake, so I'll call it a snake. All possible types of snakes for $n \le 4$ are shown in figure 6. Two snakes are assigned to the same type it they can be transformed into each other by a smooth change of independent and dependent variables preserving orientation.[4]

Consider, for instance, the snakes of the polynomials of degree 4 ($n = 3$). The three critical points will necessarily go in the order minimum-maximum-minimum. The topological

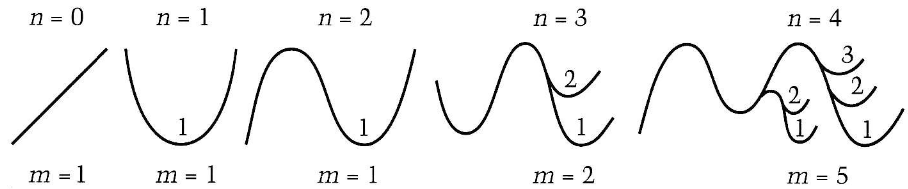


Figure 6

Classification of snakes: n is the number of bends, m is the number of snake types.

type of a snake is determined by whether the last minimum is lower or higher than the first. So the number of types for $n = 3$ is 2.

For snakes with $n = 4$ (max-min-max-min), the second maximum can be lower or higher than the first. In the first case there are two positions for the second minimum (below or above the first minimum, but always below both maxima); and in the second case three positions are possible. In all, there are five types.

If you have enough patience to draw all snakes with $n = 5$, you'll convince yourself that there are 16 of them, and for $n = 6$ there are 61 (these still lend themselves to drawing). Whenever Euler's number 61 crops up in some classification, we should look around and see if the rest of these numbers are nearby.

Now let's classify snakes by their "tails"—that is, by their rightmost critical points. Number as $1, 2, \ldots, n$ all n critical values in ascending order (from the bottom up). The number of a critical point will be called its height. The heights of the snakes' tails are indicated in figure 6. For instance, the five snakes with $n = 4$ are distributed by tail height as follows:

Height of tail	1	2	3	4
Number of snakes	2	2	1	0

Comparing the bottom line in our table with the numbers in the Bernoulli–Euler triangle (namely, the third line, counting from line 0), we see that they are the same. And, naturally, the sum $2 + 2 + 1 + 0 = 5$ (which is equal to k_4) is the total number of types of snakes with four bends.

You can make sure that the numbers of snakes with different tail heights for a certain number n of bends will always coincide with the corresponding line in the triangle (in some order). After this fact, unexpected enough in itself, has been noted (which necessarily requires some experimental work—sketching snakes), we can prove that the distribution of snakes by tail height satisfies the recursive law that defines the Bernoulli–Euler triangle.[5]

As to the analytic formula

$$K(t) = \sec t + \tan t$$

for

$$K(t) = \sum_{n=0}^{\infty} k_n \frac{t^n}{n!},$$

it can be proved as follows.

Take any snake with $n + 1$ bends, $n \ge 1$. Choose its highest local maximum and pull it upward to infinity. This will tear our snake into two shorter snakes and decrease the total

[4]That is, if the polynomials $p(x)$ and $q(x)$ that define the snakes are related by the equation $p(x) = f(q(g(x)))$, where the functions $f(x)$ and $g(x)$ have positive derivatives. You can think of these transformations as irregular shrinking and stretching of the coordinate plane along the axes without folding it or turning it over.—*Ed.*

[5]Take a snake with n bends, unbend its tail, and turn it over to get a new snake with $n - 1$ bends. Let the tail height of the first snake be h. Since the last critical point of a snake is always a minimum, the height of the next-to-last bend (a maximum) of the first snake is greater than h. Our operation turns this point into the last critical point of the second snake, so the tail height of this new snake is any number from 1 to $n - h$. This leads to the following relation for the numbers $s(n, h)$ of snakes with n bends whose tail height is h: $s(n, h) = s(n-1, 1) + s(n-1, 2) + \ldots + s(n-1, n-h)$. If we build a triangle from these numbers so that line 0 consists of the single number $s(1, 1) = 1$, and its $(n-1)$st line is $s(n, 1), s(n, 2), \ldots, s(n, n)$ for any even n, and $s(n, n), s(n, n-1), \ldots, s(n, 1)$ for any odd n, $n \ge 2$, we'll see that the above relation is just the one used to fill out the Bernoulli–Euler triangle. In particular, we see that the total number of snakes with n bends is equal to $s(n, 1) + \ldots + s(n, n) = s(n+1, 1) = k_n$.—*Ed.*

number of critical points by one. Similarly, we can pull the lowest local minimum to $-\infty$ (and turn over the left piece to make it a regular snake).

This reduction leads to the following recursive relation for $n \geq 1$:

$$2k_{n+1} = \sum_{i=0}^{n} \binom{n}{i} k_i k_{n-1}.$$

Here the product $k_i k_{n-1}$ counts the pairs of all possible "subsnakes"; the binomial coefficient $\binom{n}{i}$ takes care of possible different mutual arrangements of the critical values in the two pieces (any i numbers chosen from 1, 2, ..., n can be the heights of the bends in, say, the left piece with respect to all the n bends in the entire snake); and the factor 2 on the left accounts for the two ways in which every $(n + 1)$th snake can be torn and so enters the sum on the right. For $n = 0$ the factor 2 must drop out (why?).

In terms of the function K, this relation can be written as the differential equation

$$2\frac{dK}{dt} = 1 + K^2$$

(see if you can verify it directly). It follows that $K(t) = \sec t + \tan t$ (this function satisfies the equation and the initial condition $K(0) = k_0 = 1$).

From optimal fishing to optimal reforms

And the last example. Let's consider a model of the variation in the number of a certain population of animals (say, the number of fish in a pond or ocean).[6] In the simplest case the situation is described by the model $x' = kx$, where $x = x(t)$ is the size of the population at time t—that is, the rate of change of the population is proportional (with the coefficient k) to the population itself. The solution to this equation is the exponential function $x(t) = x(0)e^{kt}$.

However, in actual practice the living conditions of the population

[6]See also "The World according to Malthus and Volterra" in the July/August 1992 issue.—*Ed.*

Figures 7–11 present solutions to the model equation of the population change $x' = x - x^2 - c$. The red curves at left in all the figures graph the rate of change $v(x) = x - x^2 - c$ (in coordinates rotated for convenience); the blue curves at right are the solutions. The population grows at points with $v(x) > 0$, decreases for $v(x) < 0$, and is steady at points with $v(x) = 0$.

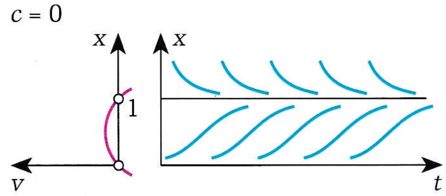


Figure 7

Stabilization of the population. All trajectories tend to the stable solution x = 1.

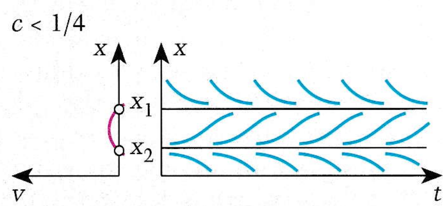


Figure 8

Model of fishing with a "moderate" catch. If the initial population $x(0) < x_2$, the population dies out; trajectories with $x(0) > x_2$ all approach $x = x_1$.

worsen as x increases, and the coefficient k decreases. Taking, for instance, $k = a - bx$, we get the so-called logistic equation. In the case $a = b = 1$, its solutions tend to a stable population level $x = 1$ (fig. 7).[7]

If, in addition, a certain quota c for the capture and consumption of a part of the population is introduced, the equation becomes only a little more complicated: $x' = x - x^2 - c$. This is the simplest model of fishing.

For a quota $c < 1/4$, a stationary solution is established again (fig. 8). For $c > 1/4$, we get rapid extinction (fig. 9).

[7]Don't be surprised at this strange population size: it's just a model, but a model that gives a more or less correct qualitative picture of the situation.—*Ed.*

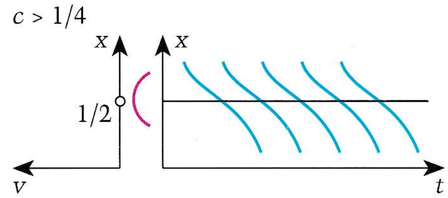


Figure 9

Overcatch. The population always dies out.

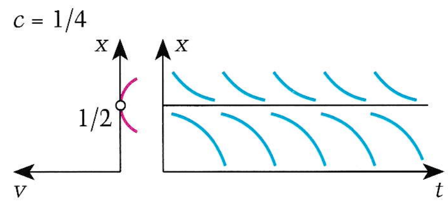


Figure 10

Destruction of the population by the optimal plan. The trajectories in the domain $x < 1/2$ lead to extinction; those with $x \geq 1/2$ tend to the steady state $x = 1/2$, but in the long run, due to small random perturbations, necessarily hit the dangerous domain as well.

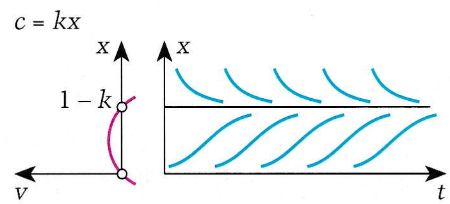


Figure 11

Stabilization of the system through feedback. (Compare with figure 7.)

For $c = 1/4$ the stationary regime is at the level $1/2$ (fig. 10). But this state is unstable: small random deviations lead to a catastrophe—annihilation of the population. How can an optimal catch be achieved while keeping the population at a certain stable level? The answer is not to assign a fixed plan for the catch, but rather to introduce feedback—that is, a catch quota proportional to actual resources. In a model with feedback $x' = x - x^2 - kx$, the optimal value of the coefficient is $k = 1/2$. With this choice of k all solutions stabilize at the level $x_0 = 1/2$, which means that the average catch over a long period will be $kx_0 = 1/4$ (fig. 11). This is the same catch as with the highest permissible fixed plan for the catch.

Greater productivity in this case is impossible. But with a rigid plan the system loses its stability and is guaranteed to self-destruct, while feedback stabilizes it, and small variations in the coefficient k do not lead to disaster.

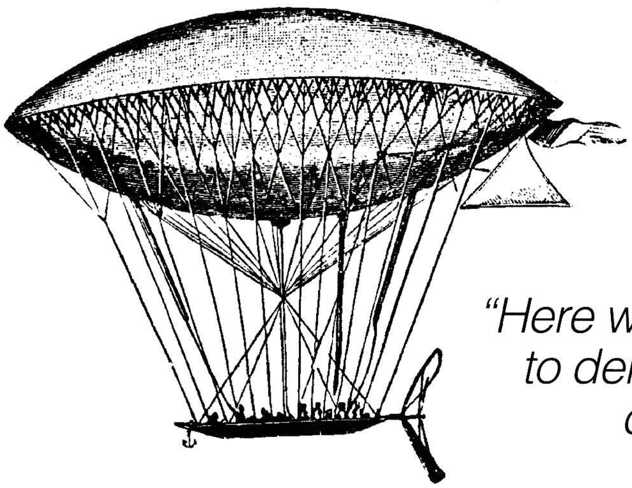
It wouldn't be bad if the persons who make crucial decisions were familiar with similar models and other rules for choosing strategic social options.

Simpler mathematical considerations—the fact that laws of nature are described by differential equations—allow us to understand certain seemingly paradoxical phenomena in our life. For several decades the state of the Russian economy has been a matter of concern for specialists: militarization, monopolism, and the general incompetence of the leadership caused the second derivative to become stably negative (that is, the rate of development steadily slowed). This didn't really scare those who didn't understand mathematics, because the first derivative was still positive (social well-being kept increasing). But mathematicians know that a permanently negative derivative, even of a higher order, ultimately makes the first derivative negative—that is, leads to a decrease in production and in the welfare of society, and this process of deterioration, when it becomes noticeable, will accelerate. Because of the inertia of the system, there are no means at all for instantaneously changing its state at this point, since changes of any kind affect only the sign of a higher derivative (for our *perestroika*, the third or even fourth derivative). Thus, the economic degradation we observe is caused by old mistakes made at the time of production growth rather than by wrong new decisions. Unfortunately, these elementary mathematical facts are too difficult to explain to a pillaged people inclined to ascribe all its hardships to failed reforms. Any reforms must certainly lead to a worsening, even if they are absolutely the correct steps to take.

Plans in this country were usually devised so as to optimize production for 20 years ("long enough for our lifetime"). It's clear to a mathematician that optimal planning of this sort must result in the complete destruction of all resources by the end of that period (otherwise, the remaining resources could have been used and, therefore, the plan wouldn't be optimal). Fortunately, the plans used to be "corrected" and were never fulfilled. However, the basic tendencies were kept, so roughly speaking, we had eaten all we had by the beginning of *perestroika*.

Attempts to create detailed "day-by-day" programs of economic reform are similar to attempts to plan the entire economy and are like trying to give minute-by-minute instructions to someone driving from Moscow to St. Petersburg: "At such-and-such a minute, turn right; at such-and-such a minute, turn left..." Success can be achieved only through feedback. That is, what's needed is not a program (trajectory) but, in mathematical terms, a vector field in the space of the system's states, a mechanism for making decisions as required by the attained state rather than the calendar date.

Some of these points should be kept in mind as well when it comes to reforming the educational system. Our examples show that "there is nothing more practical than a good theory." It's essential that educators not chase after the practical need of the moment, but rather have the long-term goals of society constantly in view.



Models of efficiency

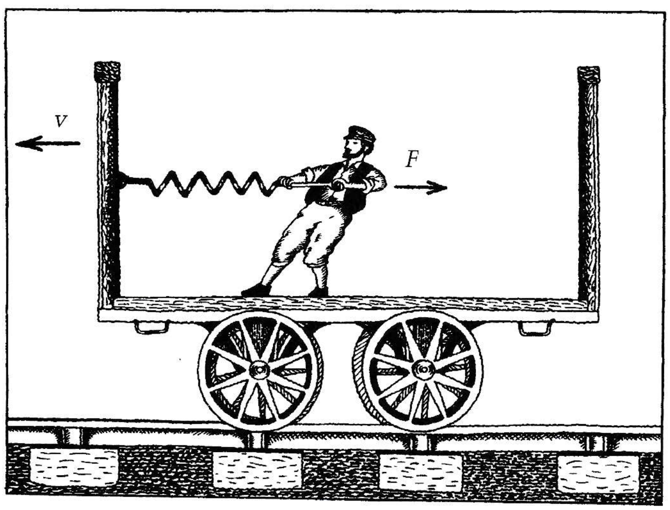
"Here we use the expression 'motive force' to denote the efficiency that an engine can develop."—Sadi Carnot

THE RESTLESS ACTIVITY OF mankind, in the last three centuries in particular, have prompted us to think constantly about how to create engines that will "spend less and produce more." Many fruitless attempts were made before science set limits to our unfettered inventiveness and showed us how to improve the engines and motors we already had. However, as a first step it was necessary to investigate such notions as work, power, and efficiency. After establishing themselves as fundamental notions in the mental toolbox of scientists and engineers, these concepts now demonstrate a wonderful versatility. Whatever processes we're working with—mechanical, thermal, or electrical—we can apply them with confidence.

Problems

1. Is mechanical work performed on a mass that is carried horizontally with a uniform velocity along a straight line?

2. In a uniformly moving railway car a man stands and stretches a spring with force F, as shown in the figure at the right. The car travels a distance D. What work has been performed by the man in the reference frame associated with the Earth?



3. Can the force of static friction perform mechanical work?

4. A bubble of gas rises from the bottom of a pond. Does the gas perform work?

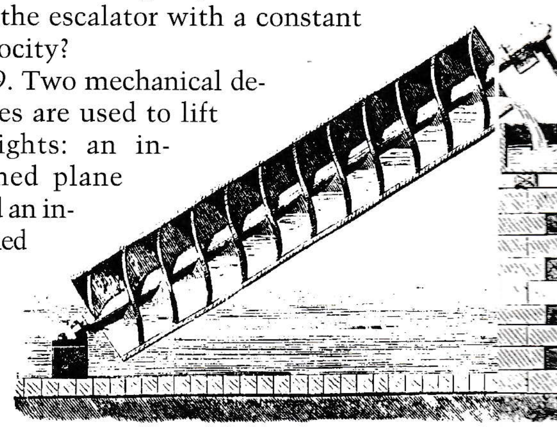
5. Why are the engines of racing cars so much more powerful than those of ordinary automobiles?

6. Assume that the drag forces of air and water increase proportionally to the square of a ship's velocity. (We neglect so-called *wave losses*.) How much less power does the ship need when its velocity decreases by a factor of 3?

7. A rocket hovers over the Earth's surface. What is the power of its engine expended on?

8. Would there be a change in the power developed by the motors of an escalator if a person standing on the upward-moving steps starts to walk up the escalator with a constant velocity?

9. Two mechanical devices are used to lift weights: an inclined plane and an inclined conveyer belt sliding over rollers. Which of them is more efficient?



10. Would a hydraulic press work if its cylinder were filled with gas instead of fluid?

11. The temperature of the air, which serves as a heat sink for a car engine, becomes appreciably lower in winter than in summer. Does this lead to an increase in the engine's efficiency in winter?

12. What is the heat source and what is the heat sink in a rocket engine?

13. Two electrical loads are connected to a battery, first in series, then in parallel. When will the efficiency be greater?

14. Can the efficiency of a battery equal 1?

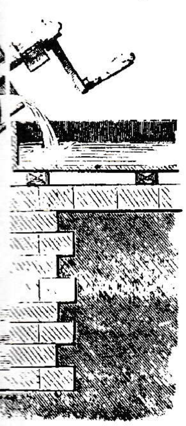
Microexperiment

Turn on an electric stove and observe it for an extended period of time. Why, despite the continuous expenditure of electrical energy, does the temperature of the coil not increase without limit?

It's interesting that . . .

. . . the French scientist V. Poncelet came up with a not exactly scientific but nonetheless extremely practical definition: "Mechanical work is what you pay money for."

. . . when a person tries to maintain a constant muscular force, even in the absence of movement the muscles contract and relax continuously, causing microscopic movements. So the muscles are performing a significant amount of work, in accordance with the standard definition of that term.

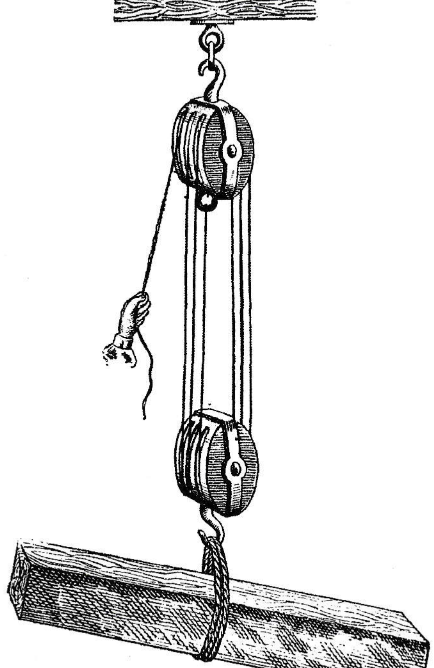


. . . the unit of an engine's power, "horsepower," introduced at the end of the 18th century by James Watt, is still in use. It was defined as the average work that a strong English draft horse could perform per second working uniformly for an entire day.

. . . the efficiency of a theoretical heat engine that would exploit the temperature difference between surface and deep ocean water does not exceed a few percent.

. . . the power developed when a click beetle lying on its back pushes off is approximately 100 times that developed by any of its muscles individually.

. . . a great amount of power does not necessarily mean a large thrust. For example, in proposed photon rockets the force of jet propulsion is assumed to equal tens or hundreds of newtons. Such a rocket could not even take



off from the Earth without assistance.

. . . a very tempting possibility exists of obtaining electrical energy directly from the chemical energy of a fuel and an oxidizer without combustion in a so-called electrochemical generator, which has a very high efficiency.

. . . when burning 1 milligram of gasoline in a car engine, we obtain about 40 joules of heat energy, a small fraction of which winds up as kinetic energy of the car. One milligram of sugar provides an organism with the same 40 joules of energy, but the energy is utilized much more effectively to maintain its body temperature and for other biological functions. ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 62

The most profit with the least effort

In memoriam P. L. Chebyshev (1821–1894)

by Yuly Danilov

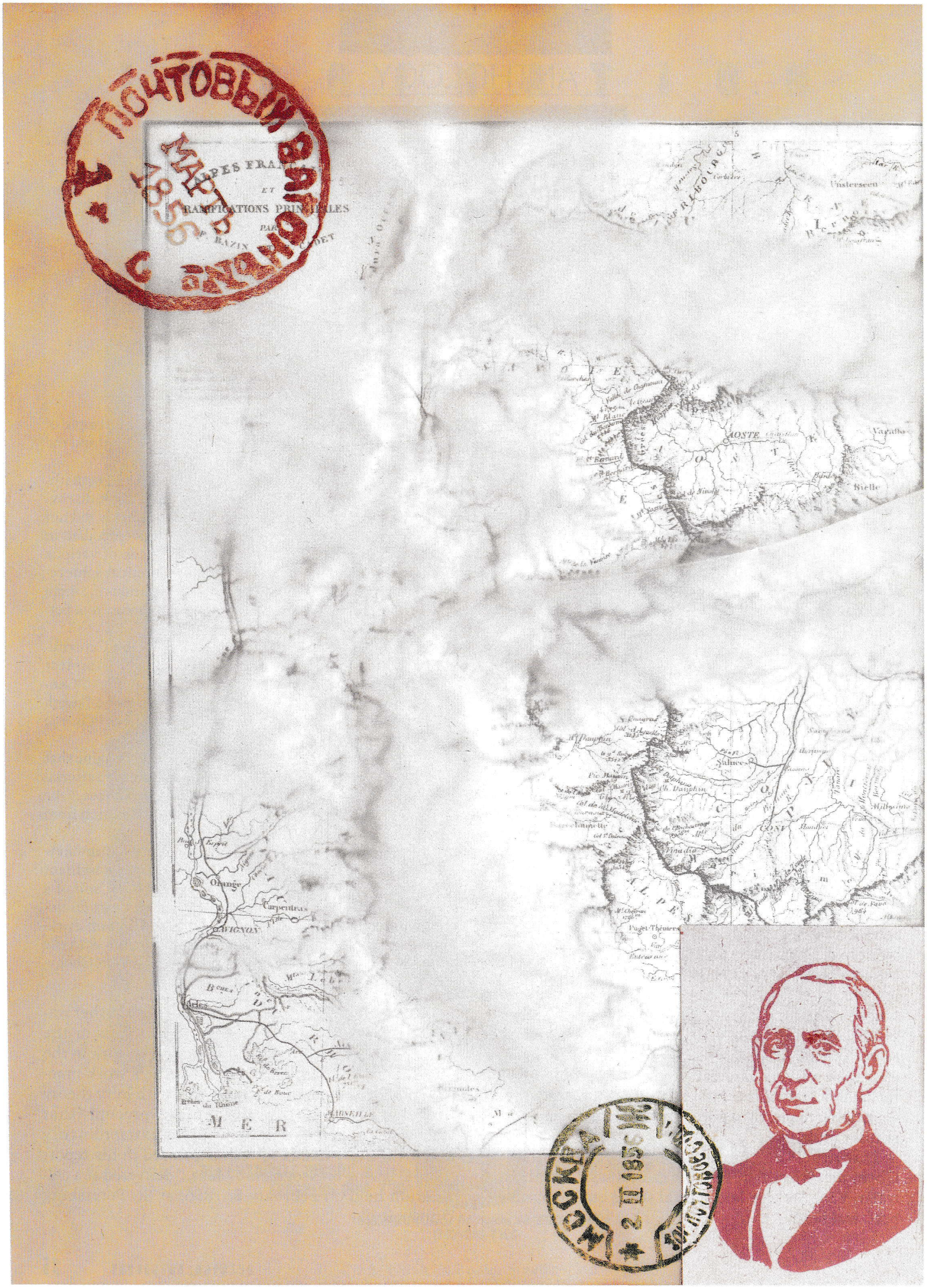
THE LIST OF HONORIFIC TITLES AND DEGREES of P. L. Chebyshev is long and impressive: Doctor of Mathematics and Astronomy (1849); ordinary member (academician) of the Saint Petersburg Academy of Sciences (1859); ordinary (after 1872, emeritus) professor of St. Petersburg University (1860); corresponding member of the Academic Board of the Ministry of State Property (1854), the Liege Royal Society (1856), the Philomathematical Society in Paris (1856), and the Cherbourg Society of the Natural Sciences (1866); founding member of the Moscow Mathematical Society (1867); member of the Berlin (1871), Bologna (1873), and Royal Italian academies of sciences; member of the Royal Society of London (1877) and the Mathematical Society of France (1882); foreign associate member of Academy of Sciences in Paris (1874); foreign member of Swedish Academy of Sciences (1893); honorary member of the Moscow (1858), Kiev (1869), Novorossiisk (1878), St. Petersburg (1882), and Kazan (1893) universities, of the Artillery Academy (1870), the Academic Board of the Ministry of Education (1873), the Moscow Society of Natural Scientists (1889), the St. Petersburg Mineralogical Society (1890), and the St. Petersburg Mathematical Society (1893).

Chebyshev founded a St. Petersburg mathematical school and achieved worldwide fame for his brilliant work in probability, integration of algebraic functions, number theory and other branches of pure mathematics. He nurtured many outstanding mathematicians whose names signify the glory of the mathematical sciences: G. F. Voronoy, D. A. Grave, E. I. Zolotarev, A. N. Korkin, A. M. Lyapunov, and others. Chebyshev also took a lively interest in practical problems: the rational design of mechanical devices, the drawing of geographic maps, the optimal cut of clothing, and other similar problems.

During a celebration of the centennial of Chebyshev's birth in 1921, the great Russian mathematician V. A. Steklov laid special stress on Chebyshev's practical bent:

> The genius of Chebyshev . . . was a striking example of the union of practice in the highest sense of the word and the creative, generalizing force of an abstract thinker.
>
> He turned practical problems into a corresponding mathematical theory, which would turn out to be a new discovery in the domain of pure science; but the discovery did not remain in the sphere of pure thought, but was embodied in reality—in machines and mechanical devices of various kinds that served as a realization of his theoretical achievements. Along with purely theoretical investigations . . . there is a series of papers whose titles may seem strange to a person not in the field. . . .
>
> For example: "On a Mechanical Device"; "On a Centrifugal Equalizer"; "On Gear Wheels"; "On the Simplest Joints"; and so on. Then we find "On the Drawing of Geographic Maps" and, finally, his paper delivered to the French Association for the Improvement of the Sciences "On the Cut of Clothing."
>
> Coming upon such a title, could a lay person imagine that the investigation belongs not to a specialist in the sartorial arts but to the author of "The Theory of Congruences," the creator of "The Theory of Functions of Least Deviation from Zero"? . . .
>
> The almost boundless domain of new problems and new methods of solving them arise out of Chebyshev's brilliant ideas, which emerged and were developed on the soil of a single philosophical idea: *to take Nature as it is, as an unavoidable, real observational fact, and to derive from available observational data as much profit as possible with the least effort.* (A. V. Steklov, "Theory and Practice in Chebyshev's Works")

Chebyshev's practical works, with their rather unusual titles, were not the whims of a genius or the fruits of his leisure away from his arduous work in the domain of pure mathematics. In his life's work practical problems were indissolubly tied with lofty theory and flowed from a philosophical precept that Chebyshev adhered to through all his extraordinarily fruitful activity and is so well expressed in his report "The Drawing of Geographic Maps," delivered at a celebration held on February 8, 1856, at St. Petersburg University.


ПОЧТОВЫЙ ВАГОНЪ
МАРТЪ
1856
ET
RAMIFICATIONS PRINCIPALES
AOSTE
Saluces
Orange
Carpentras
MER
МОСКВА
2 III 1856

The drawing of geographic maps

A paper written for a celebration at the Imperial St. Petersburg University on the 8th of February, 1856

Gentlemen!

The mathematical sciences, from the most ancient times, have attracted special attention; at present they have garnered even more interest due to their impact on the arts and industry. The convergence of theory and practice gives the most fruitful results, and not only practice gains; the sciences themselves develop under its influence: practice opens new subjects for them to investigate or new aspects of subjects known long ago. In spite of the higher degree of development to which the mathematical sciences have been brought by the work of the great geometers of the last three centuries, practice clearly reveals their incompleteness in many respects; it offers essentially new problems for science and thus induces a search for completely new methods. If theory gains much from new applications of an old method or new developments of it, then it gains all the more from the discovery of new methods, and in this case the sciences find in practice a reliable guide.

The practical activity of humankind is extraordinarily diverse, and science of course lacks many and various methods for satisfying all its requirements. But of these the most important are the methods needed to solve different variants of one and the same problem common to all practical activity of humankind: *how to dispose of one's means in order to achieve the greatest gain.*

Solving problems of this kind constitutes the subject of the so-called *theory of maximum and minimum values.* These problems, purely practical in nature, are particularly important for theory as well: all laws governing the motion of matter, ponderable and imponderable, are solutions to problems of this sort. It is impossible not to notice their especially fruitful influence on the development of the mathematical sciences.

Before the discovery of the infinitesimal calculus only particular examples of the solutions of such problems were known; but already in these solutions there was the origin of a new, very important branch of the mathematical sciences, known by the name of *differential calculus.* In order to show the influence of these problems, I shall cite the passage from Newton's famous treatise *Philosophiae naturalis principia mathematica*, where he speaks of the origin of this discovery, whose applications and results are now numberless.

> About ten years ago (in 1667), when I had been corresponding with Leibniz, I wrote to him that I knew a method for determining *maximum and minimum values,* drawing tangents, and solving other similar questions, and that my method with the same convenience can be applied to equations containing radicals as well as rational numbers. At that time I concealed my method in an anagram that had the following meaning: "Given an equation containing any number of things that flow, find the current, and vice versa." To this, the celebrated Leibniz answered that, for his part, he had also found such a method and reported it to me in the same letter. His method differed from mine only in its name and notation. (Note to statement VII of book 2, 1713 edition)

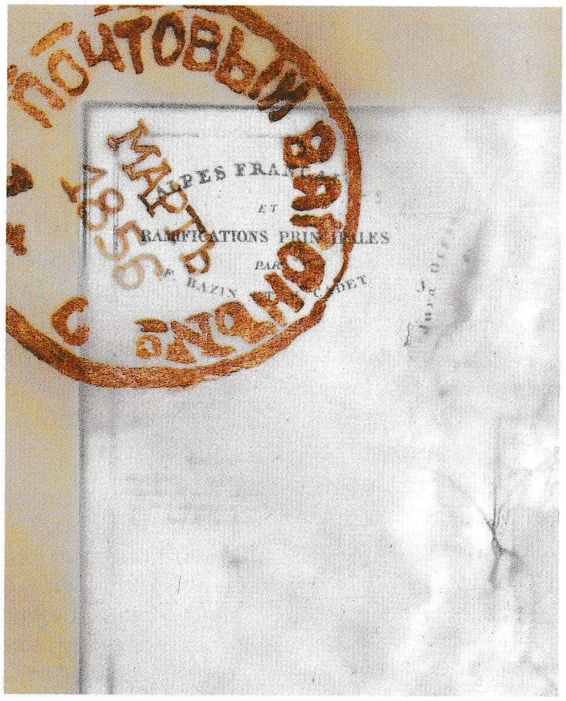
But the subject was not exhausted with the discovery of differential calculus and the solution of problems similar to those that led to its discovery, and this was found in the research of Newton himself: the question of finding a shape that enables a body moving in a fluid to encounter the minimum resistance, which he solved, posed a problem of *maximum and minimum values* that were fundamentally different from similar problems that can be solved by differential calculus. A general method of solving problems of this type, especially important for theoretical mechanics, brought to light another new calculus, known by the name of *variational calculus.*

In spite of this development of mathematics with regard to the theory of *maximum and minimum values,* it is not hard to see that practice goes further and requires the solution of problems of *maximum and minimum values* of a new kind, fundamentally

Art by Yury Vashchenko

different from those solved by differential and integral calculus.

As an example of problems of this kind and their solution, we can present our research on Watt's parallelogram, published in *Memoirs des Savants Étrangers* of our Academy for 1854. From the results we achieved, examining the method needed to find the optimal design for this kind mechanism, we can see that questions of practice in this case as well led to numerous theoretical results that are of interest for science; that the methods called forth initially by practice are the means of solving new theoretical problems that are of interest independently of their practical significance.



Another example of this type of problem, and an especially striking one, is the *drawing of geographic maps*. Given the modern state of the theory of geographic maps, one can show countless methods of drawing them such that very small elements of the land will retain their true form in the image. However, in doing so, because of the spheroidal surface of the Earth, the scale of the image of different land elements will necessarily be different, so that equal elements, taken at different places depicted on the map, will be different sizes. The greater these changes in scale, the more irregular the geographic map. And because the magnitude of these changes in scale across the area of the same portion of the surface can be larger or smaller, depending on the method of projection used in the map, the following question naturally arises:

For what projection will these changes in scale be smallest?

In a paper I delivered at a conference of the Academy of Sciences on January 18 [1856], I showed that this problem, translated into the language of calculus, can be reduced to a particular problem of *maximum and minimum values* that is fundamentally different from problems solved by differential and variational calculus. This problem is similar to those that were the subject of the aforementioned paper "On Watt's Parallelogram" but belongs to a higher rank of such problems: there, several constant values had to be found; here, two unknown functions have to be found, which is tantamount to determining an infinite number of constants. This presupposes the same difference between these problems as that between problems in differential and variational calculus. From the theoretical point of view, this subject is all the more interesting in that it leads to an investigation of a partial differential equation that is particularly remarkable and, among other things, expresses the heat equilibrium in infinitely thin plates. Thus, a question about the most advantageous map projection is related to this remarkable property of heat: given a state of thermal equilibrium in a round, infinitely thin plate, the temperature in the center is the *mean* of the temperatures of all the points on the circle; the same is true of a sphere: the temperature in the center is the *mean* of the temperatures on the surface.

The final solution regarding the most advantageous map projection is very simple: the optimal projection for depicting some portion of the Earth's surface on a map is that in which the scale on the boundary of the image preserves one and the same value, easily determined by the accepted, normal value for the scale. As for finding the projection that possesses such a property, it can be reduced to solving an ordinary problem in the *integration of partial differential equations* when the integral is given on the boundaries and must remain finite and continuous inside.

Thus, there is one projection for depicting each country on a map that is the most advantageous. This projection is determined by the position of the country relative to the equator and the shape of its boundaries; the parallels of latitude and meridians of longitude will be different curved lines, but in general close to circles and straight lines, if an insignificant portion of the Earth's surface is projected. These lines can be drawn point by point without any difficulty.

Especially remarkable are those cases when the parallels and meridians turn out to be perfect circles or straight lines; this makes it much easier to draw maps of insignificant dimensions. In his paper "Sur la construction des cartes geographiques" (*Noveaux Memoirs de l'Academie de Berlin*, 1779), Lagrange found all projections for which this holds true. On the basis of the property of the most advantageous projection, it can easily be shown in general which countries can be depicted optimally by using such projections: the boundaries of these countries are determined by points in which the scale preserves one and the same value under this kind of projection. The boundaries of countries determined in this way are in general rather complicated curves. But as the area to be depicted on the map is decreased, they become simpler and quickly approach ellipses, so that they differ insignificantly from ellipses even for maps of such extensive countries as, for example, the European part of Russia. These ellipses have certain defined positions: their center is located at the center of the projection; one of the axes goes along the meridian. The ratio of the axes of these ellipses is defined by the position of their center relative to the equator and, in particular, by a magnitude that Lagrange called the *indicator* of the projection.

Conversely, in order to depict any portion of the Earth's surface that is not too large and is bounded by such an ellipse, it is possible to find a method of projection in which parallels and meridians are circles and straight lines and which gives us an image that is close to being perfect. But to achieve this, according to what was said above, the center of the projection and its *indicator* must be chosen properly in accordance with the position of the country and the shape of its boundaries. Therefore, the particular methods of projection that preserve similarity in infinitesimal elements—that is, stereographic projections (polar and horizontal) and the Gauss and Mercator projections, which can all be deduced from a general method involving a particular conjecture about the position of the center of the projection or the value of the indicator—can give an image on the map that is close to perfect only in certain particular cases.

Thus, if the aforementioned ellipse turns into a circle, the indicator becomes equal to 1, and the most advantageous projection reduces to a stereographic horizontal projection, which turns into a polar projection when the center of the circle coincides with the Earth's pole. As the axis of the ellipse, directed along the meridian, becomes smaller, the most advantageous projection approaches the Gauss projection. When the center approaches the equator, the Mercator projection becomes the most advantageous.

It is clear from this that, in attempting to obtain the optimal image of different countries on a map, one must not restrict oneself to one or several methods, but it is necessary to use a general method, each time appropriately selecting both the center of the projection and the value of the indicator.

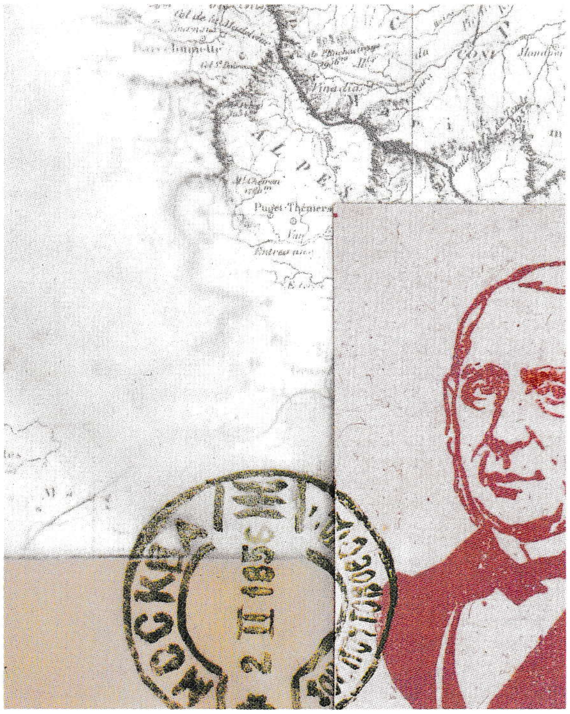
According to what was said above, this can easily be done when one is depicting on a map a portion of the Earth's surface whose boundary is an ellipse with an axis directed along a meridian. But such simple cases do not occur in practice; the boundaries of different countries are always extraordinarily irregular curves. In spite of this, for the best image of a country that is not too large, one can determine both the position of the center of the projection and the value of the indicator, comparing the shape of the boundaries with an ellipse or other conic sections. To this end, it is sufficient to have only an approximate image of the country for which the most advantageous position of the center of the projection and the value of the indicator is being sought, and because of this a map drawn according to any method can be used here.

Strictly speaking, here one can make three different assumptions, which give rise to three different solutions; but comparing them, one can easily find the most advantageous one. First, the country to be projected can be considered part of an area bounded by an ellipse with an axis along a meridian; for countries in which the greatest propagation along meridians and parallels is almost opposite the center, this always corresponds to the most advantageous solution. This case occurs in practice most frequently. Second, the land to be projected can be considered part of an area between two ellipses, hyperbolas, or parabolas situated identically. This can give the most advantageous solution only for depicting countries that are bent in the shape of a sickle or are a narrow band slanting toward the meridians and parallels. Third and last, the country can be compared with an area confined between the branches of two reciprocal hyperbolas; this corresponds to countries whose boundaries are significantly concave opposite the center.

Turning our attention to the first assumption, which applies to the majority of cases occurring in practice, we note that, of the set of ellipses that can be circumscribed about the country to be projected, the most advantageous projection is determined by the smallest of them, if in order to compare the different ellipses we take the length of the mean diameter slanted equally toward both axes.

From the appearance of the country to be projected, it is not difficult to find the points upon which the ellipse will rest and use them to determine its axis and center. The center of this ellipse will be the most advantageous position of the center of the projection, and the ratio of the axes determines the most advantageous indicator. All this refers to drawing maps of extremely small countries; but for larger countries, in accordance with the method of successive approximation, one can easily find corrections for both the position of the center of the projection and the value of the indicator. Thus, the most advantageous method of drawing a map of the given country will be found, in which the parallels and meridians remains circles.

It can be seen from this that the drawing of geographic maps is one of those practical problems that are solved differently for different countries, and that the method of drawing that is advantageous for France, Germany, and England may turn out to be disadvantageous for Russia. In addition, Russia, due to its great size, presents a special challenge in drawing its map, and because of this fact the choice of the projection that most suits its area, the shape of its boundaries, and its position relative to the equator is especially important. Maps of its different parts present very sensitive changes in scale, to say nothing of maps that encompass all of Russia. Thus, depicting on a map everything that belongs to Russia on this side of Ural Mountains according to Gauss's method, one allows changes in scale of more than 1/20, and this, when one is measuring surfaces, gives a difference of *one* square mile per *ten*, which is a very significant error. The error of a map decreases in a stereographic horizontal projection with a properly selected center, but in this case the differences in scale are as much as 1/34; this corresponds to a difference of *one* square mile per *seventeen*, when the surface is being measured. These errors are not so small as to be unworthy of our attention; the way to diminish them is to determine which projection corresponds best to the shape and position of the land to be projected.



Looking at this part of Russia on a map, we note that in the general outline of its boundaries it is far from approaching ellipses with axes directed along a meridian, and in this case, as we have seen, it is impossible to obtain an optimal image on the map, preserving the meridians and parallels as circles or straight lines. Such a simplification in drawing its map implies a considerable reduction in the degree of regularity of the image. To obtain a truer image, it is necessary, according to what was said above, to determine the method of projection by integrating a special equation. Since this integration must be carried out under a condition that depends on the shape of the boundaries, and these boundaries are very complicated curves, exact integration is, of course, impossible. But practice does not require this. For practical purposes it is sufficient to restrict oneself to a change in scale of *one ten-thousandth*, and in this case everything can be reduced to finding several coefficients, which can easily be calculated with accuracy sufficient for practice according to the shape of the boundaries, no matter now curved they are. As for the parallels and meridians, they can be drawn point by point without difficulty.

Turning now to the simplest methods for drawing maps, where the parallels and meridians are circles and straight lines, we note that Russia's possessions on this side of the Urals, together with the Caucasus and Georgia, spread more from north to south than from east to west, and that is why this territory cannot be compared with a circle and even less so with an ellipse, whose axis from north to south is very small compared to the axis from east to west. Therefore, according to what has been said, neither a Gauss projection nor a stereographic projection corresponds in this case to the land to be projected. Applying to this case the method presented here for determining the center and indicator of the projection, we note that the center of the smallest ellipse that, with an axis along a meridian, embraces all the possessions of Russia to the Urals, along with the Caucasus and Georgia, is located between Yaroslavl and Uglich at 57°36′ latitude; the ratio of its axes is 17/10. Taking this ellipse as a basis, we find that the most advantageous projection has an indicator of 1.0788. This value differs from 1, the indicator of a stereographic projection, by less than by one *tenth*. But even this difference has a considerable influence on the degree of accuracy of the image. As we have seen, a stereographic projection, with its center positioned most advantageously, covering the area of Russia under consideration gives a change in scale of 1/34. Taking the value we obtained, 1.0788, for the indicator of the projection and choosing its center between Yaroslavl and Uglich (at the 57° longitude, 57°42′30″ latitude), we obtained a map of this part of Russia where changes in scale do not exceed 1/50, and this is the highest degree of accuracy that can be attained, preserving the parallels and meridians as circles or straight lines.

In much the same way, gentlemen, most practical problems can be reduced to problems of maximum and minimum values—problems quite new for science; and only by solving these problems can we satisfy the requirements of practice, which everywhere seeks the best, the most profitable. ◘

Turning algebraic identities into geometric inequalities

You do it using complex numbers

by Zalman Skopets

MANY INEQUALITIES CONCERNING absolute values turn into statements of identity when the absolute values are removed. We can sometimes go backwards, too: an identity involving polynomials can turn into a valid inequality if we take the absolute value of each term on one side, while taking the absolute value of the whole polynomial on the other side.

We can then think of the arguments in such a relation as complex numbers and give them a geometric interpretation. This simple idea allows us to obtain a number of interesting geometric inequalities that are rather difficult to prove directly.

Before we look at some examples, let's recall the primary definitions, notations, and facts about complex numbers. This will give us all the information we'll need, so you don't need any prior knowledge about complex numbers to understand the arguments below (at least, theoretically).

Complex numbers, which will be denoted by a, b, c, ..., are expressions of the form $x + iy$, where x and y are real numbers and i is a so-called *imaginary unit* defined by the property $i^2 = -1$. A complex number $a = x + iy$ can be represented by the point $A(x, y)$ in the coordinate plane

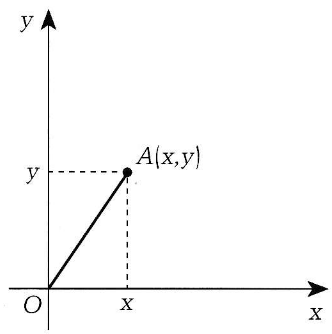


Figure 1

(fig. 1). Then the distance from this point to the origin O is called the absolute value of a and is denoted by $|a|$:

$$|a| = |x + iy| = \sqrt{x^2 + y^2} = OA. \quad (1)$$

The sum of two complex numbers $a_1 = x_1 + iy_1$ and $a_2 = x_2 + iy_2$ is defined as $a = a_1 + a_2 = (x_1 + x_2) + i(y_1 + y_2)$. The geometric construction of the point A from A_1 and A_2 is shown in figure 2 (in terms of vectors, $\overrightarrow{OA} = \overrightarrow{OA_1} + \overrightarrow{OA_2}$). By the Triangle Inequality, $OA \le OA_1 + A_1A = OA_1 + OA_2$, so according to equation (1), we have

$$|a_1 + a_2| \le |a_1| + |a_2|.$$

The equality here is achieved if (and only if) the points O, A_1, and A_2 are collinear and O doesn't belong to the segment A_1A_2.

Thus, *the absolute value of the sum of two complex numbers does not exceed the sum of their absolute values.*

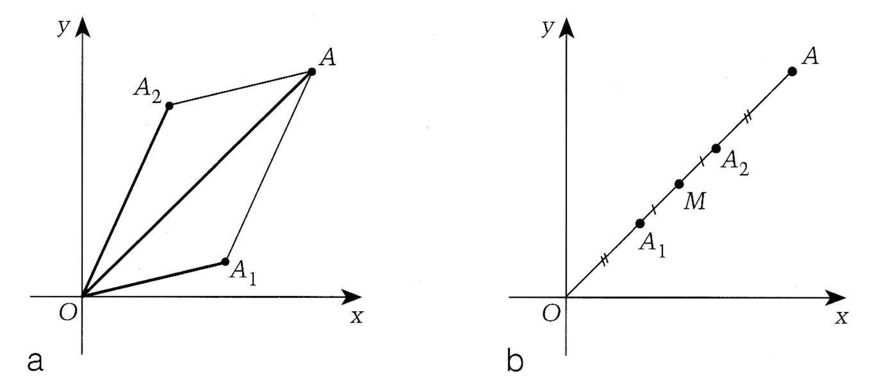


Figure 2

(a) Points O, A_1, A_2 are not collinear; (b) points O, A_1, A_2 are collinear (M is the common midpoint of A_1A_2 and OA).

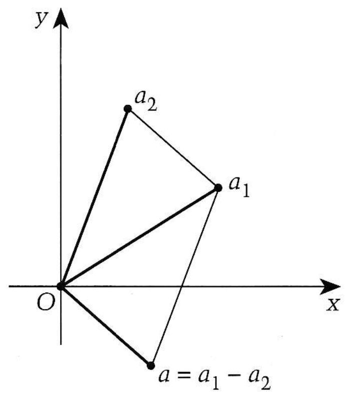


Figure 3
In this figure $|a_1 - a_2| = |a|$.

The difference $a_1 - a_2$ of two complex numbers a_1 and a_2 is the number a such that $a + a_2 = a_1$. The construction in figure 3 makes it clear that *the absolute value of the difference between two complex numbers is equal to the distance between the points that represent them.*

Further, from $OA_1 - OA_2 \le A_1A_2 \le OA_1 + OA_2$ we get

$$|a_1| - |a_2| \le |a_1 - a_2| \le |a_1| + |a_2|.$$

If points O, A_1, A_2 are collinear, one of these two inequalities becomes an exact equality.

So, *the absolute value of the difference between two complex numbers is no greater than the sum and no less than the difference of their absolute values.*

Also, we can see (fig. 4) that the following inequality, for instance, is true:

$$|a + b + c| \le |a| + |b| + |c|.$$

Notice that two opposite numbers a and $-a$ are represented by two

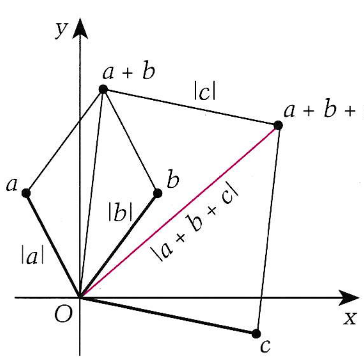


Figure 4

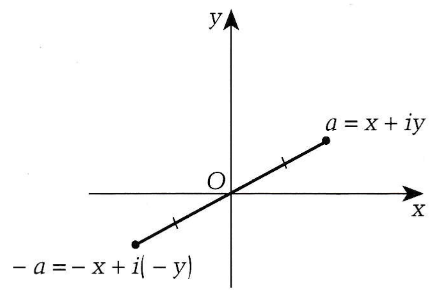


Figure 5

points, (x, y) and $(-x, -y)$, symmetric about the origin (fig. 5), so $|-a| = |a|$.

In fact, all the properties above could also be formulated in terms of vectors. Essentially new opportunities for applications arise only when we make use of another operation on complex numbers—multiplication. It is defined by the formula

$$(x_1 + iy_1)(x_2 + iy_2) = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1),$$

which is obtained by multiplying out the two factors—

$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 + iy_1y_2 + ix_1y_2 + i^2y_1y_2,$$

replacing i^2 with -1, and collecting like terms. One can verify by direct computation that the product of complex numbers obeys the following properties:

$$ab = ba,$$
$$a(b + c) = ab + ac,$$
$$|ab| = |a| \cdot |b|.$$

So algebraically we can deal with the addition, multiplication, and absolute value of complex numbers as we usually do in the case of real numbers. More details about complex numbers can be found in any precalculus textbook. Here we'll confine ourselves to this brief introduction, which will suffice to demonstrate our method of deriving geometric inequalities from algebraic identities.

Example 1. Consider the well-known identity

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

Taking the absolute value of both sides, we get

$$|a + b| \cdot |a - b| = |a^2 - b^2 \le |a^2| + |b^2|,$$

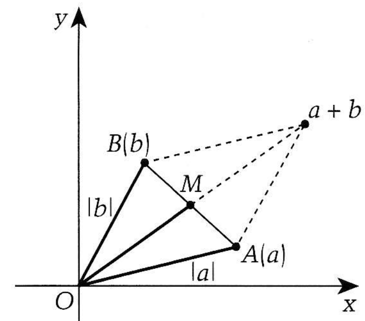


Figure 6

or

$$2OM \cdot AB \le OA^2 + OB^2,$$

where M is the midpoint of segment AB and O is the origin (fig. 6). In other words, *the sum of the squares of two side lengths of a triangle is not less than twice the product of the third side length and the length of the median drawn to the third side.*

Example 2. We can verify the identity

$$-(b - c)(c - a)(a - b) = a^2(b - c) + b^2(c - a) + c^2(a - b).$$

It follows that

$$|b - c| \cdot |c - a| \cdot |a - b| \le |a|^2 \cdot |b - c| + |b|^2 \cdot |c - a| + |c|^2 \cdot |a - b|,$$

or (see figure 7)

$$BC \cdot CA \cdot AB \le OA^2 \cdot BC + OB^2 \cdot CA + OC^2 \cdot AB$$

for any four points O, A, B, C in the plane. This inequality is especially interesting in these two particular cases.

(1) For an equilateral triangle ABC we have

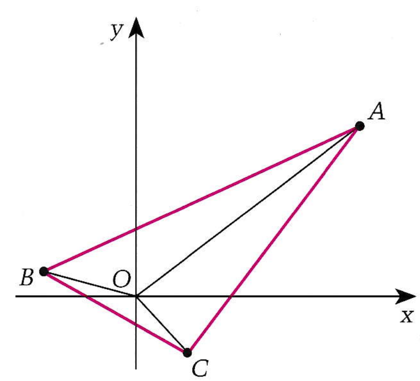


Figure 7

$$OA^2 + OB^2 + OC^2 \geq d^2,$$

where d is its side length. Thus, *the sum of the squares of the distances from an arbitrary point O to the vertices of an equilateral triangle is not less than the square of its side length.*

(2) If O is the circumcenter of triangle ABC and R is its circumradius, then

$$AB \cdot BC \cdot CA \leq R^2(AB + BC + CA),$$

or

$$\frac{AB \cdot BC \cdot CA}{AB + BC + CA} \leq R^2$$

—*the ratio of the product of all the triangle's side lengths to its perimeter is not greater than the square of its circumradius.*

Recalling that the area of the triangle ABC can be expressed as $AB \cdot BC \cdot CA/4R$, or $(AB + BC + CA)r/2$, where r is the inradius, we can rewrite the last inequality in this neat form: $R \geq 2r$.

Example 3. We can verify the identity

$$(a - b)(a - c)(b - c) = (b - c)(b + c)^2 + (c - a)(c + a)^2 + (a - b)(a + b)^2.$$

Denote by d_a, d_b, d_c the distances from the circumcenter of a triangle ABC to its sides BC, CA, AB, respectively (fig. 8). Let the circumcenter be the origin. Then

$$|b + c| = 2d_a,$$
$$|c + a| = 2d_b,$$
$$|a + b| = 2d_c,$$

and so the identity yields

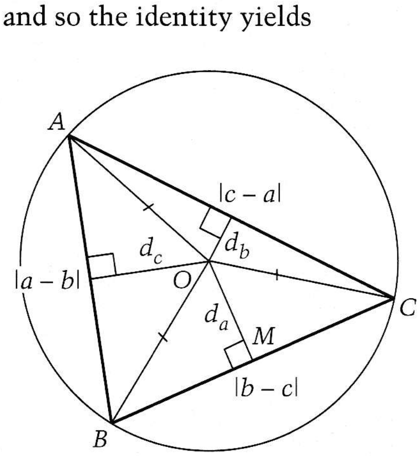


Figure 8

$$AB \cdot BC \cdot CA \leq 4(d_a^2 \cdot BC + d_b^2 \cdot CA + d_c^2 \cdot AB).$$

Notice that $d_a \cdot BC = R^2 \sin \angle BOC$, because both these expressions equal twice the area of triangle BOC. On the other hand, $d_a = OB \cos \angle BOM$. If angle A in triangle ABC is not obtuse, then $\angle BOC = 2A$ and $\angle BOM = \frac{1}{2}\angle BOC = A$; otherwise, $\angle BOC = 360° - 2A$ and $\angle BOM = 180° - A$. But in either case $\sin \angle BOC \cdot \cos \angle BOM = \sin 2A \cdot \cos A$, so $d_a^2 \cdot BC = R^3 \cos A \cdot \sin 2A$. Similar formulas are true for $d_b^2 \cdot CA$ and $d_c^2 \cdot AB$, while $AB \cdot BC \cdot CA = 4R \cdot S_{ABC}$, where S_{ABC} is the area of ABC (this fact had been mentioned above). Putting all of these together, we come up with the inequality

$$\frac{S_{ABC}}{R^2} \leq \cos A \sin 2A + \cos B \sin 2B + \cos C \sin 2C,$$

which is true for any triangle.

Now prove these geometric inequalities yourself.

Exercises

1. Derive from the identity $a^2 - b^2 = (a + b)(a - b)$ the inequality $2OM \cdot AB \geq OA^2 - OB^2$, where M is the midpoint of AB. When does it turn into an equality?

2. Derive from the identity

$$a^3(b - c) + b^3(c - a) + c^3(a - b) = -(a + b + c)(a - b)(b - c)(c - a)$$

the inequality $OH \leq R^2/2r$, where O is the circumcenter, H the orthocenter, R the circumradius, and r the inradius of a triangle ABC. (The orthocenter is the point where a triangle's heights intersect.)

3. Prove the inequality $AB \cdot CD + BC \cdot DA \geq AC \cdot BD$ for any four points A, B, C, D in the plane. An additional problem (which may require some additional facts about complex numbers): prove that this inequality becomes an equality if $ABCD$ is a quadrilateral inscribed in a circle (Ptolemy's theorem).

4. Prove that $AB \cdot AM \cdot BM + BC \cdot BM \cdot CM + CA \cdot CM \cdot AM \geq AB \cdot BC \cdot CA$ for any four points A, B, C, M. ◘

HINTS ON PAGE 62

"OPTICS FOR A STARGAZER" CONTINUED FROM PAGE 21

the sky. Therefore, the altitude a of the Sun above the horizon t minutes before sunset was

$$a = 0.25° \cdot \sin 60° \cdot t \cong 0.2t.$$

Thus, you can see Sirius with the naked eye when the Sun's altitude does not exceed $a_n \cong 0.2° \cdot 21 \cong 4.5°$, and with binoculars when the corresponding value is $a_b \cong 0.2° \cdot 43 \cong 9°$. Under these conditions the luminosity of the sky directly overhead is 7% and 13%, respectively, of its midday value.[3] Remember that the magnitude of Sirius is one fifteenth that of Venus. When the brightness of the sky decreases by a factor of 15 before sunset, Sirius can be seen with the naked eye. Binoculars help one see it in a brighter sky, because it increases the brightness of the star while changing the surface brightness of the sky only negligibly. So an instructive experiment was carried out by an amateur astronomer from Houston!

Now it's easy for us to believe that Sirius can be seen in the daytime in the mountains or from an airplane, because the sky is 15 to 20 times darker at altitudes of 5–7 km than at sea level. The next time you're in an airplane, take a look around: can you see Sirius, or Jupiter, or Venus? ◘

[3] D. Y. Martynov, *A Textbook of Practical Astrophysics*, Moscow and Leningrad, Nauka, 1977, p. 300.

PHYSICS CONTEST

Rising star

"What is the sound of one hand clapping?"—Zen koan
"Can two sounds make silence?"—Physics challenge

by Arthur Eisenkraft and Larry D. Kirkpatrick

WE SUPPOSE THAT YOUNG people are first introduced to waves while attending or watching sports events. These stadium waves can provide some useful insights into the most counterintuitive property of waves: the wave moves, but the medium does not. In a stadium wave, a group of spectators at one end of the stadium stands and then sits. This triggers the adjoining section of fans to stand and sit, followed by the next section, and so on. While the wave of people standing and sitting moves around the stadium, no person moves in that direction—that is, the people remain at their seats. Leonardo da Vinci noticed this wave property in water and remarked that the wave flees the place of creation while the water does not.

An interesting wave phenomenon that is not easily demonstrated in a stadium wave is interference. What happens when two waves meet? A first step in our understanding will be to look at two pulses passing each other on a spring. What one notices is that as the peaks of the pulses meet, a momentary superpeak is created (fig. 1a). What is perhaps more surprising is that when a peak pulse meets a valley pulse, there may be a point on the string that doesn't move. For this point, it's as if no pulse passed by (fig. 1b).

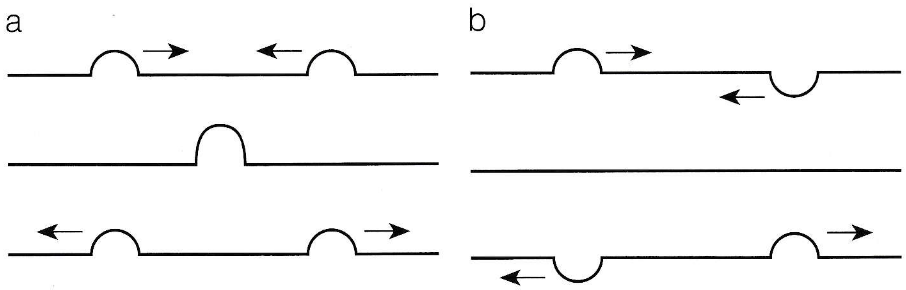


Figure 1

A periodic wave is a continuous series of pulses (fig. 2). This representation can be assigned to any wave phenomenon. A sound wave, which propagates by disturbing the air in compressions and rarefactions, can use figure 2 as a graphical representation, where the peaks are compressions and the valleys are the rarefactions. A series of sketches in which two waves pass each other will reveal that the sum of the waves produces points on the string that undergo large displacements and other points that undergo *no* displacement whatsoever (see figure 3 on page 46). The points of maximum disturbance are called *antinodes*, while the points of no disturbance are referred to as *nodes*.

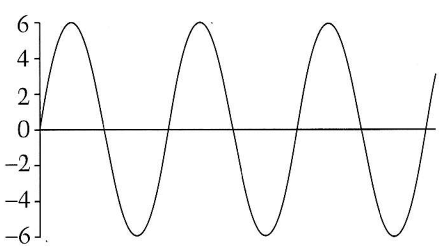
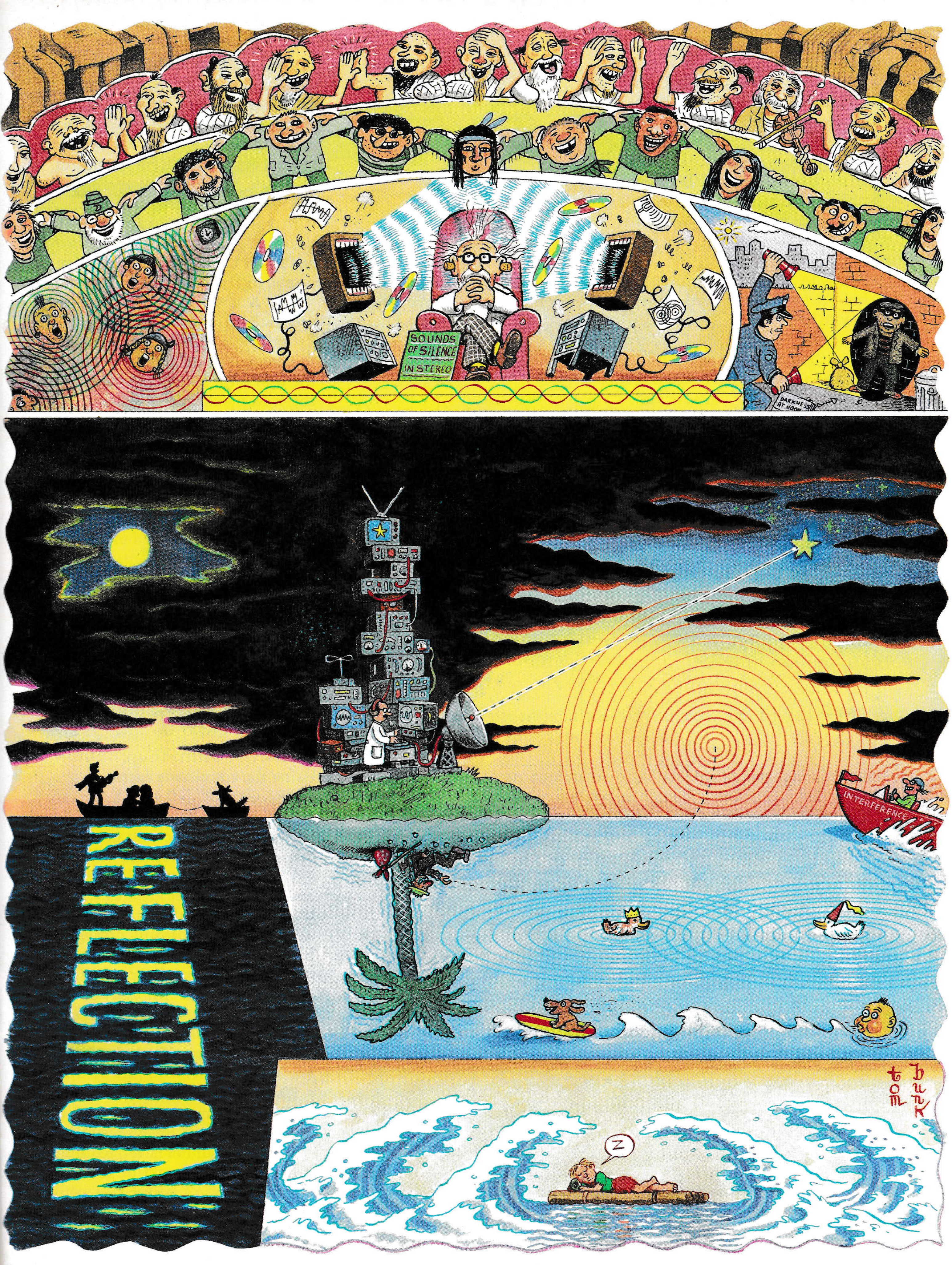


Figure 2

The interference of sound waves can create these nodal points, and one of the jobs of the acoustical engineer is to ensure that a new concert hall does not have places (due to reflections) where aspects of the music cannot be heard. Acoustical engineering is both an art and a science. It is part good fortune and part mystery why Carnegie Hall and La Scala have such exceptionally fine acoustics.

And so, in answer to the physics challenge posed above, two sounds *can* create silence. Two light sources can also create darkness, as in Young's double-slit experiment or a Michelson interferometer. And two

Art by Tomas Bunk


SOUNDS OF SILENCE
IN STEREO
DARKNESS AT NOON
INTERFERENCE
REFLECTION!
Z
tom bunk

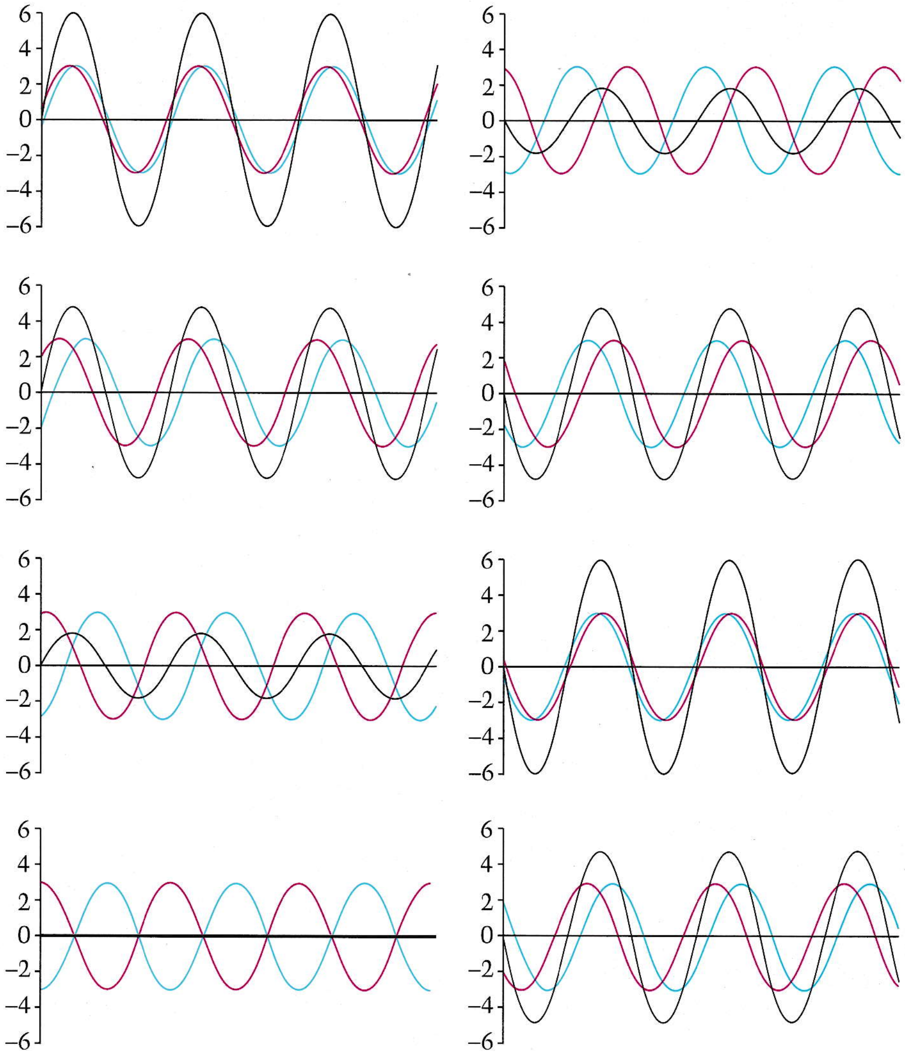


Figure 3

beams of electrons can produce locations where no electrons will reach in what is arguably the most profound discovery of the 20th century. The Zen koan about one hand clapping must remain a mystery. We're not sure what insights physics can offer to this puzzle.

This month's contest problem is from the XII International Physics Olympiad, held in Bulgaria in 1981. Readers of *Quantum* are urged to send in a solution and to provide a brief autobiography. Other readers may just wish to comment on the problem in general by sending messages to the authors via e-mail (quantum@nsta.org).

The receiver of a radio observatory is placed on an island, near the shore, at a height of 2 m above sea level. It detects only the horizontal components of the electric field. When a radiostar radiating waves with a wavelength of 21 cm rises above the horizon, the receiver records maxima and minima.

A. Determine the altitude of the star when maximum and minimum are observed.

B. Does the intensity decrease or increase after the star first appears above the horizon?

C. Investigate the ratio of the intensity of the successive maxima and minima.

(Note: The ratio of the amplitudes of the incident and reflected waves is $(n - \sin\theta)/(n + \sin\theta)$, where θ is the angle of the incident wave measured from the horizontal and n is the index of refraction. For radio waves and water, $n = 9$.)

Please send your solutions to *Quantum*, 1840 Wilson Blvd., Arlington VA 22201-3000 within a month of receipt of this issue. The best solutions will be noted in this space and their authors will receive special certificates from *Quantum*.

Fun with liquid nitrogen

We hope our *Quantum* readers enjoyed "playing" with liquid nitrogen in the Contest Problem that appeared in the March/April issue.

Part A asked you to calculate the specific heat of aluminum. We begin by writing down an expression for the conservation of energy where the first term is the heat gained by the water (subscript "w") and the second term is the heat lost by the piece of aluminum (subscript "Al"):

$$c_w m_w \Delta T_w + c_{Al} m_{Al} \Delta T_{Al} = 0,$$

where c is the specific heat, m is the mass, and ΔT is the change in temperature. Therefore,

$$c_{Al} = -c_w \frac{m_w}{m_{Al}} \frac{\Delta T_w}{\Delta T_{Al}}$$

$$= \left(-1 \frac{\text{cal}}{\text{g} \cdot {}^\circ\text{C}}\right)\left(\frac{100\ \text{g}}{36.2\ \text{g}}\right)\left(\frac{6^\circ\text{C}}{-77^\circ\text{C}}\right)$$

$$= 0.215 \frac{\text{cal}}{\text{g} \cdot {}^\circ\text{C}}.$$

Using the conversion factor 1 cal = 4.186 J, we get $c_{Al} = 0.9$ J/g · K at room temperature, in agreement with the graph of the specific heat of aluminum given in the problem.

Part B used data taken at the XXIV International Physics Olympiad to calculate the latent heat of vaporization of liquid nitrogen. There are two complications involved in this experiment: (1) the change in the specific heat of the aluminum as a function of temperature and (2) the loss of heat to the surroundings, since the liquid nitrogen boils at 77 K.

The specific heat of aluminum as a function of temperature is shown in the graph given in the problem. Because the amount of heat q required to change the temperature of 1 g of aluminum by ΔT is just $c\Delta T$, this

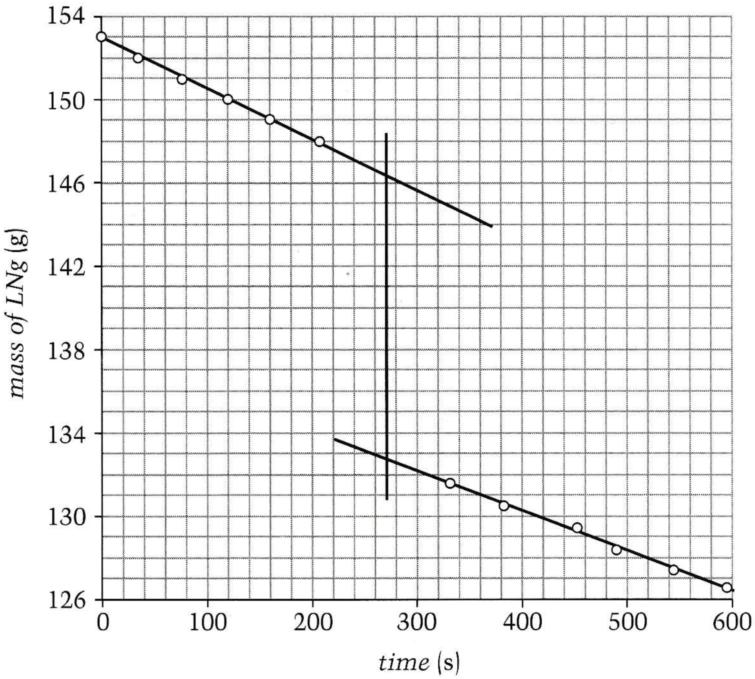


Figure 4

heat can be obtained graphically by computing the area under the curve between the two temperatures. The easiest way of doing this is to count boxes (estimating fractional boxes) under the curve between T = 77 K and T = 293 K. We obtain 300 ± 6 boxes with the area of each box being (0.05 J/g · K)(10 K) = 0.5 J/g. (Don't forget the 130 boxes below the x-axis.) This gives us a total q = 150 ± 3 J/g. Therefore, the total heat given up by the aluminum is

$$Q = m_{Al}q = (19.4\text{ g})(150\text{ J/g}) = 2{,}910\text{ J}.$$

We now need to obtain the mass of liquid nitrogen that was evaporated with this heat. To do this we plot the mass of the liquid nitrogen in the calorimeter as a function of time as shown in figure 4. (Don't forget to subtract the mass of the aluminum after it is put into the liquid nitrogen.) This graph shows that the loss of mass due to heat from the surroundings is quite important and occurs at a different rate before and after the aluminum is put into the liquid nitrogen. We can obtain a very good estimate of the mass of liquid nitrogen vaporized by looking at the difference in the two lines at the middle of the time interval—that is, around 270 s. This yields m_N = 14.4 ± 0.3 g. We can now calculate the latent heat of vaporization:

$$L = \frac{Q}{m_N} = \frac{2{,}910\text{ J}}{14.4\text{ g}} = 202\text{ J/g}.$$

In part C you were asked to calculate the error in your value for the latent heat. You may remember the simple rule that when experimental values are multiplied or divided, the percentage errors in the measured quantities add. This gives an overestimate and can be refined by adding the errors in quadrature—that is, add the squares of the errors and then take the square root.

The error in q is 2.0%, while the error in m_{Al} is 0.5%. Adding these in quadrature yields a 2.1% error in Q, or Q = 2,910 ± 60 J. The error in m_N is 2.1%. Adding this in quadrature to that in Q yields an error in L of 3%. Therefore, our experimental value is

$$L = 202 \pm 6\text{ J/g}.$$

Erratum

During typesetting, an error crept into the last contest problem and went undetected during proofing. The display equation in the third column on page 32 should read

$$\frac{1}{s'} = \frac{1}{f} - \frac{1}{s} = \frac{1}{f} - \frac{1}{3f} = \frac{2}{3f}.$$

We hope all our readers caught this typo.

Spinning in a jet stream

Vacuuming up Bernoulli's law and the Magnus force

by Stanislav Kuzmin

PERHAPS SOME *QUANTUM* readers have had a chance to observe how a Ping-Pong ball hangs in the jet of air created by a hair drier or a vacuum cleaner. Let's modify the experiment a little bit. Take a wooden ball from, say, a children's wooden construction set and drill a hole along its axis about 1 cm in diameter. If this ball is placed in the stream of air, it first just hovers, then it begins to revolve with increasing speed such that the hole has a horizontal axis. The frequency of revolution may be as high as 100 per second, and the height the ball attains is five times higher than at first (compare figures 1a and 1b). Here's how the frequency of revolution was measured in our experiment: a small magnet was embedded in the ball's surface, the ball was placed in a coil of wire, and the induced voltage was fed to the lead of an oscillograph.

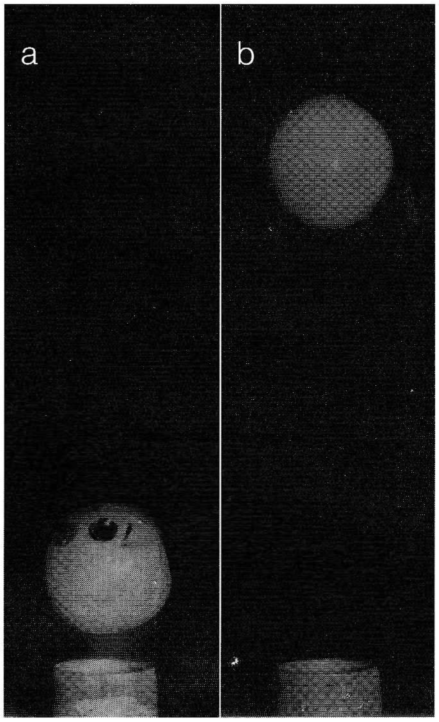


Figure 1

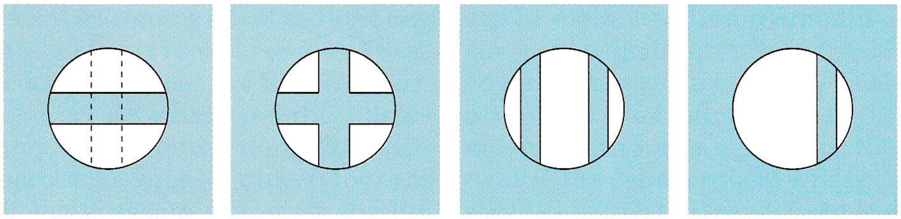


Figure 2

A similar experiment can be carried out with other kinds of balls—also solid but drilled in different ways (fig. 2). In every case both lift and rotation are observed, but the rotation of the ball with a shifted hole doesn't occur about a single axis (in scientific language, it *precesses*). This phenomenon seems to be connected with a change in the position of the center of mass relative to its geometric center.

For these experiments we can also use hollow balls (for instance, Ping-Pong balls) with holes in them. True, they rotate more slowly and don't rise as high. However, if a paper tube is inserted into the hole, the rate of rotation increases.

All these experiments show that the flow of air through the hole is of primary importance. If the hole is closed with, say, modeling clay, the rotation stops. To gain a better understanding of the shape of the air flow around the rotating ball, we can attach long threads to the vacuum cleaner tube and then photograph the apparatus using a flash. Figure 3 shows that the flow presses against the ball on the side that moves with the flow and moves away from the ball on the opposite side. As this

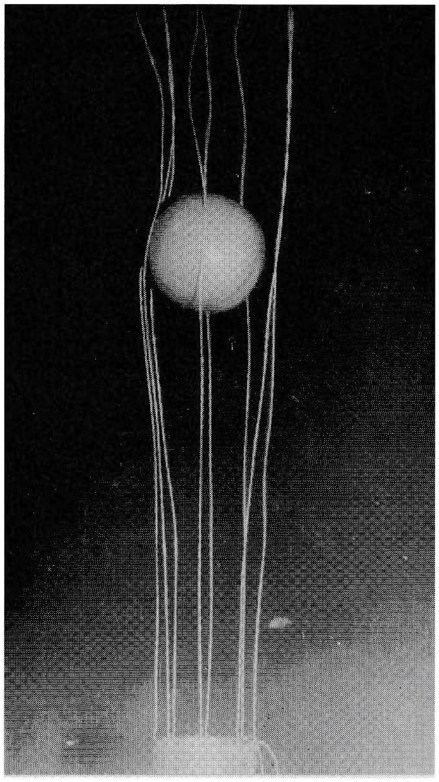


Figure 3

Art by Dmitry Krymov

occurs, the ball shifts relative to the center of the jet.

Let's try to come up with a possible explanation of the experiment. First we consider the case of an absolutely smooth ball without a hole. It is suspended in the jet in a stable way even if the jet is slightly inclined. This is explained by Bernoulli's law, because the pressure within the jet is lower than that in the surrounding air, so when the ball shifts a little bit, the side emerging from the jet will be pushed back in.[1]

Now let's look at a rotating ball. If it is rotating at the center of the jet, the velocity and, therefore, the pressure are different on opposite sides of the ball, because the rotating ball slows down the flow on one side and accelerates it on the other. Then a force arises that shifts the ball aside. However, as our ball hangs at the same place, the average pressure on both sides should be the same. It follows that the rotating ball must be displaced from the center of the jet, because the equality of the velocities on opposite sides is possible only in such a position. (The velocity in a jet decreases with the distance from its center, which means that the side of the ball near the jet's center must move opposite to the flow, while the other side will move in the direction of the jet.)

So why does the ball rotate? Again we begin by looking at a ball without a hole, which hangs exactly on the axis of the jet and doesn't rotate. If we move the ball to one side, the air on its outer side begins to rub on a larger surface area compared to the ball's other side, which results in the ball's rotation about a horizontal axis. When it is released, the ball returns to its equilibrium position on the jet's axis and the rotation stops.

In a similar experiment the ball with a hole plays another game—the hole makes the rotation stable even without an external force. This is because the hole changes the structure of the flow around the ball. During a quarter-turn (fig. 4a) the small jet coming out of the hole creates a cloud of small vortices. These vortices prevent the flow from becoming detached, and this allows the flow to move along the ball's surface for a longer time. It looks as if the viscosity of the air and the corresponding friction were increased. During the phase of rotation shown in figure 4b this effect is absent, since the flow was able to detach itself by this time. Thus, a frictional force arises that turns the ball. As it turns, a transverse Magnus force arises that moves the stable position of the ball relative to the jet's axis.

[1]However, see "Foiled by the Coanda Effect" on page 4.—*Ed.*

Art by Dmitry Krymov

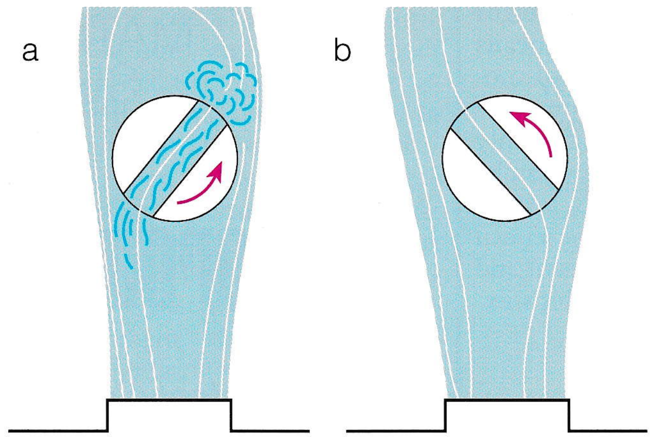


Figure 4

The extra lift of the perforated ball might be connected with the existence (in some positions) of a vertical component of the force displacing the ball from the jet's axis (fig. 5a). The air passing through the hole sucks in the rest of the stream, which causes the jet to incline. This doesn't happen in the opposite position (fig. 5b) because nothing here prevents the stream from detaching from the ball.

Now let's try to make some estimates. Along the vertical direction the ball is affected by three forces: gravitation, the drag of the air, and the vertical component of the Magnus force. The drag force can be written as $F_d = k\rho v^2 S$, where k is the coefficient of drag (which depends on the shape of the body), ρ is the density of air, v is the velocity of the airflow, and S is the cross-sectional area of the ball. The order of magnitude of the Magnus force can be estimated from Bernoulli's law. Let a cylinder of length l and diameter d turn clockwise with a linear speed u in an infinite flow of gas moving with speed v. Then to the right of the cylinder there will be a speed $v - u$, and the speed to the left will be $v + u$. According to Bernoulli's law, the pressure difference on a cylinder is

$$\Delta p = \rho\frac{(v+u)^2}{2} - \rho\frac{(v-u)^2}{2},$$

and the Magnus force is $F_M \cong \rho vudl$.

For our rough estimate the ball can be approximated as a cylinder having equal length and diameter, while the linear speed of rotation can be taken as equal to the speed of the stream. To obtain the vertical component of the Magnus force, we multiply it by the sine of the jet's angle of inclination β (judging from figure 3, β is about 4–5°). Then for a ball we get $F_M \cong \rho d^2 v^2 \sin\beta$.

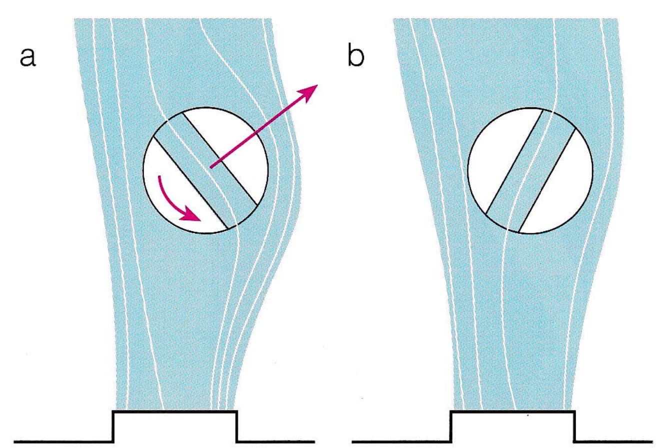


Figure 5

The change in speed with height in the jet can be estimated by means of the law of conservation of momentum. Since the static pressure and air density in the jet change only slightly in our case, this law can be written approximately as $v_1^2S_1 = v_2^2S_2$. The cross section of the jet can be expressed by the altitudes h_1 and h_2 and by the angle of the jet's expansion. Thus,

$$\frac{v_2}{v_1} = \frac{h_1}{h_2}.$$

It should be noted that the jet begins not where it exits from the tube but at the vertex of the angle of the jet's expansion, so its altitude is measured from the vertex and not the edge of the tube.

From the condition of equilibrium of the ball in the vertical direction $mg = F_d + F_M$, we can express the height to which the ball rises:

$$h_2 \cong h_1 d v_1 \sqrt{\frac{\rho(k + \sin\beta)}{mg}}.$$

Here h_1 is the height of the ball and v_1 is the velocity of the air as it exits the tube, both of which can be measured easily. Our estimates gave a height for the ball of about 30 cm, which is in good agreement with our experimental observations. ◘

When he wrote this article, **Stanislav Kuzmin** *was an eleventh grader at School No 130 in Novosibirsk, Russia.*

MATHEMATICAL SURPRISES

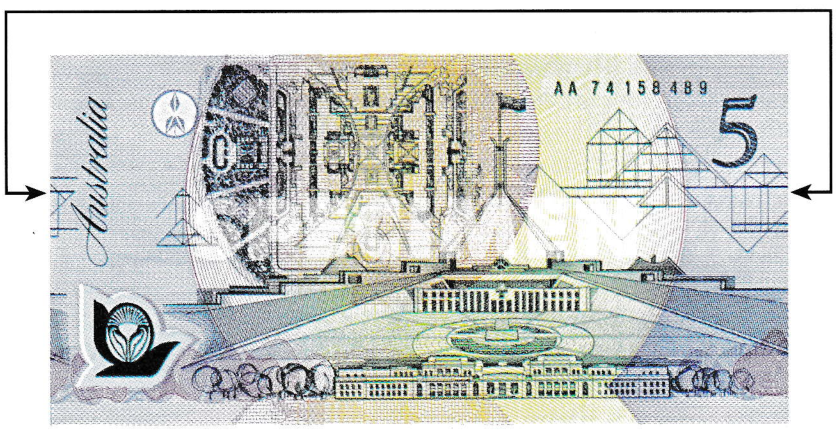
Toroidal currency

Australia's polymer money has a curious twist to it

by Martin Gardner

I DOUBT IF MANY MATHematicians outside Australia are aware that the Reserve Bank of that nation has recently issued two bank notes that are toroidal. They are the $5 bill printed in 1992 and the $10 bill printed in 1993.

The front and back of each of these handsome notes are shown. Observe that on both sides of each bill the patterns at top and bottom "wrap around," as well as the patterns on the left and right edges. As all topologists know, if the right and left sides of a rectangle, and its top and bottom, are joined, the result is the familiar



torus or doughnut shape. If just one pair of sides is connected by reversing one of the edges, the structure is a Klein bottle. If the reversal applies to both pairs of edges, the structure is a projective plane. Perhaps some day a nation will print projective plane or Klein bottle notes!

The purpose of the toroidal wraparounds is to make it harder to counterfeit the bills. Sophisticated color copiers have made counterfeiting much easier all over the globe.

An additional anticounterfeiting feature of each note is a transparent "window" at a lower corner (it appears black as printed here). Still another such device is the little circle showing four points of a star on one side and three points on the other. Hold either bill up to a strong light and the points fuse to form a perfect seven-pointed star, symbolizing Australia's seven original states. The slightest variation in register would distort the star.

The face on the $5 note obviously is that of Queen Elizabeth. The woman on the $10 note is Dame Mary Gilmore, an Australian poet who worked tirelessly to battle injustices in the nation, especially in the treatment of the native aborigines. The man on the opposite side is A. B. "Banjo" Paterson, a ballad singer and journalist best known for having written the words of "Waltzing Matilda," Australia's unofficial national anthem. (Matilda, by the way, is not a woman but a knapsack.) "The Man from Snowy River" is another of Paterson's popular songs.

The $10 note has been made 7 mm longer than the $5 bill to help sight-impaired persons distinguish between the two. Both notes are made of locally produced polymer rather than imported paper. The polymer lasts longer, stays cleaner, and can be recycled for plastic products.

Melbourne artist Max Robinson designed the new $10 note. Behind Paterson's profile, in microprinting, are lines from Paterson's verse, making the bill even more difficult to counterfeit. ◘

MATH INVESTIGATIONS

Constructing triangles from three located points

Of the 139 problems, 20 are still looking for a solution!

by George Berzsenyi

As promised in my last column, in this issue I'll share with my readers yet another set of unsolved construction problems, called to my attention by my friend Leroy (Roy) F. Meyers. These problems are based on an article by William (Bill) Wernick in the September 1982 issue of *Mathematics Magazine* and on subsequent work by Bill and Roy, which is yet to be published. Bill Wernick is a retired mathematician (from CCNY), whose *Advanced Geometric Constructions* (written with Alfred S. Posamentier and first published in 1973) is strongly recommended to my readers.

Following the work of Bill Wernick, the sixteen most important points of a triangle will be denoted as follows:

vertices	A, B, C
circumcenter	O
feet of medians	M_a, M_b, M_c
centroid	G
feet of altitudes	H_a, H_b, H_c
orthocenter	H
feet of angle bisectors	T_a, T_b, T_c
incenter	I

(For the sake of brevity, the term "feet" designates the point of intersection of the given lines (medians, altitudes, and angle bisectors) with the opposite sides of the triangle. The angle bisectors are of the internal angles.)

Bill Wernick's 139 problems consist of a listing of significantly distinct triples of these "located points" and ask for the (re)construction of triangle ABC from them. In a general setting, one may assume that three points are given in the plane, they are labeled by three of the sixteen symbols listed above, and one is to construct a triangle whose located points are the ones given. In a more restrictive setting, one could start with triangle ABC, label all sixteen of its points, erase all but three of them (along with the various lines connecting them), and attempt to reconstruct triangle ABC from them. Clearly, some arrangements of three located points in the plane will not be obtainable in the latter manner.

My first challenge to my readers is to **reconstruct the 139 significantly different problems mentioned above**. As a partial aid, in the table below I have retained the original numbering given to the list of problems by Bill.

The 139 problems fall into five categories:

1. *Redundant* triples, in which any two of the three given points will determine the third. Of the 119 problems resolved, only (A, B, M_c), (A, M_a, G), and (O, G, H) fall into this group.
2. *Locus-restricted* problems. These yield either an infinity of solutions or none at all, depending on the chice of one of the points. Twenty-three of the 119 unresolved problems fall into this category.
3. *Unsolvable* problems, which do not allow for the construction of a triangle by Euclidean tools (that is, compass and straight edge). Thus far, 20 such triples have been identified.
4. *Solvable* problems. In these one can construct a (basically) unique triangle by Euclidean tools.
5. *Unresolved* problems. These 20 triples are listed below:

77. O, H_a, T_b	122. G, T_a, T_b
78. O, H_a, I	123. G, T_a, I
81. O, T_a, T_b	127. H_a, H_b, T_c
90. M_a, M_b, I	128. H_a, H_b, I
109. M_a, H, T_b	132. H_a, T_a, T_b
110. M_a, H, I	134. H_a, T_b, T_c
111. M_a, T_a, T_b	135. H_a, T_b, I
113. M_a, T_b, T_c	136. H, T_a, T_b
118. G, H_a, T_b	137. H, T_a, I
119. G, H_a, I	138. T_a, T_b, T_c

It's highly probable that many of these problems are of the "unsolvable" variety; in that case, in addition to the tools mentioned in my previous column, the following result, found in G. Chrystal's *Algebra, an Elementary Textbook* (reprinted by Chelsay Publishing Company in 1964), may also be of use.

Theorem. *The monic quartic equation* $x^4 + ax^3 + bx^2 + cx + d = 0$ *with rational coefficients* a, b, c, d *has a constructible root if and only if it or its Lagrange resolvent,* $y^3 - by^2 + (ac - 4d)y + (4bd - c^3 - b^2d) = 0$, *has a rational root.*

Surely there are several other useful tools hidden in the literature which may also be of value as you **resolve these 20 problems**!

Bulletin Board

Duracell/NSTA Scholarship Competition

Over $90,000 (face value) in US Series EE Savings Bonds will be awarded in the 13th Annual Duracell/NSTA Scholarship Competition. The competition is open to full-time ninth through twelfth grade students in the United States and its territories. Entrants design and build a device powered by Duracell batteries. A good way to begin is to come up with an idea for a device that will help people or serve some useful purpose. Many successful inventors simply thought of something that will make life easier. Most of this year's top devices either helped the handicapped, were safety-related, or were an improvement on an existing item. (See Happenings in the May/June issue for a descriptions of winning inventions).

One first-place winner will take home a $20,000 bond, five second-place winners receive $10,000 bonds, ten third-place students get $1,000 bonds, 25 fourth-place winners receive $200 bonds, and 59 fifth-place students get $100 bonds. First- and second-place finishers, their parents or guardians, and their teacher-sponsors will also win an all-expenses paid trip to the NSTA National Convention in March of 1995 in Philadelphia. The top six winners will receive their awards at a special banquet to be held in their honor at the convention. The 100 winners will each receive a personalized certificate, suitable for framing. Everyone is a winner in the competition, because all entrants will receive a certificate of appreciation and an entry gift.

For the second year, First Step gives you an opportunity to send in an idea for a device you might want to enter in the formal competition. Simply give a 100–150-word description of a device you might like to design and build for the formal competition. All eligible entrants will receive a coupon for Duracell batteries and a general critique letter.

The 1994 top six winners received national media attention when they were featured in a segment that aired numerous times on the Cable News Network. This was in addition to the coverage they received in their local areas and from press in the Los Angeles area covering the NSTA National Convention in Anaheim, California.

You can receive an entry kit by writing to Duracell/NSTA Scholarship Competition, 1840 Wilson Blvd., Arlington VA 22201-3000. All entries must be received at NSTA Headquarters by January 13, 1995.

Toshiba/NSTA ExploraVision Awards competition

A voice-activated toilet . . . a machine that uses lasers to construct tunnels . . . a holographic traffic control system . . . a medical alert system implanted in the wrist . . . These are some of the future innovations envisioned by national finalist teams in the 7–9 and 10–12 grade level entry categories in the second annual Toshiba/NSTA ExploraVision Awards competition.

More than 17,000 students in grades K–12 from the United States and provinces of Canada entered the 1994 competition. Working in teams of three or four with a teacher-advisor, the students chose a present technology and envisioned how it might be used 20 years from now. Of the nearly 5,000 team entries received this year, almost 2,000 teams were from the 7–9 grade level, and over 1,000 teams were from the 10–12 grade level.

The twelve national finalist teams—four first-place and eight second-place teams—received a trip to Washington, D.C., for a weekend of activities, including a press conference and visits with congressional representatives. The highlight of the weekend was an awards dinner featuring George Takei, Star Trek's "Captain Sulu," and Rep. George E. Brown, Jr., chair of the Science, Space, and Technology Committee, who served as honorary chairperson for the awards ceremony. Broadcast coverage about the weekend included a story about the national finalist teams on CNN's "Future Watch" and "Real News For Kids" and a live interview with George Takei on a Washington, D.C., Fox television affiliate.

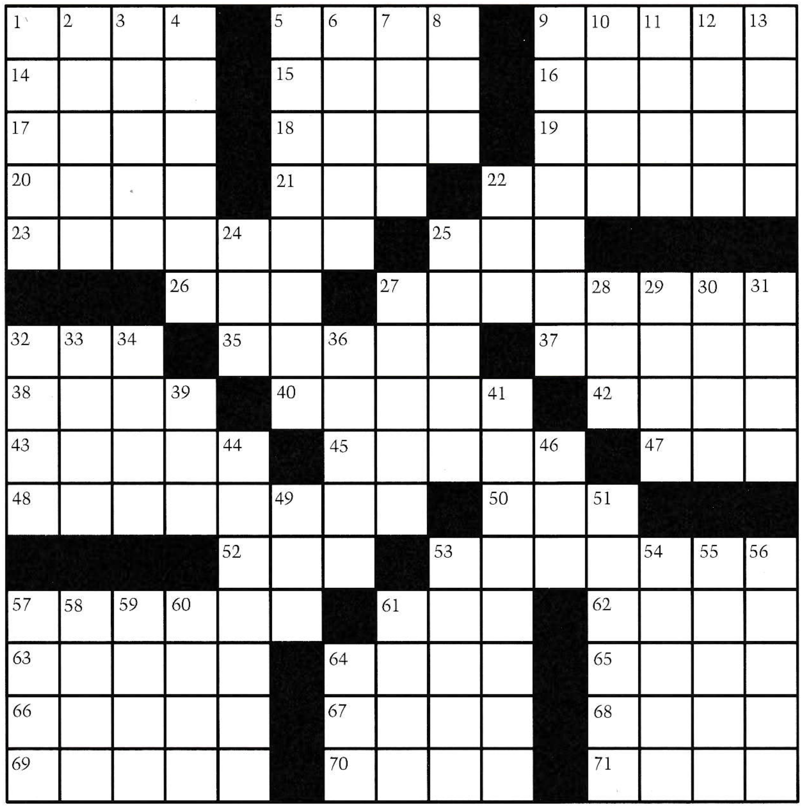
In addition to a trip to the nation's capital, students on the first-place teams each received a $10,000 US savings bond, and students on the second-place teams each received a $5,000 US savings bond. The teacher-advisors and schools of the national finalist teams were awarded their choice of Toshiba products, such as laptop computers, copiers, TVs, and VCRs.

The 1995 competition will expand to include all of Canada. Entry kits for the 1995 competition will be mailed to teachers in September.

CONTINUED ON PAGE 62

Criss × cross science

by David R. Martin



Across

1 Galatea's lover
5 Units of energy
9 Diver's equipment
14 Rod
15 Asian country
16 Fool
17 ___ acid
18 Tails
19 1967 Chem. Nobelist Manfred ___
20 Roman garment
21 Measurement system: abbr.
22 Eye's coat
23 Hard rubber
25 Floor covering
26 Distilled coal
27 It's high in the Dead Sea
32 Bank account
35 Units of substance
37 Hot drink
38 Danube tributary
40 Umbilicus
42 Mister in Munich
43 Does as commanded
45 Blood
47 ___ giant star
48 Atmospheric gas
50 ___ rod (brace)
52 Roman spirit
53 Degrees ___
57 Hipbones
61 Tap
62 Synthesized chromosomal material
63 Expect
64 Square root of 2809
65 Landed
66 Chalk and gelatin
67 Beige
68 Million billion (pref.)
69 German city
70 Unit of matter
71 First garden

Down

1 Like some angles
2 Leguminous tree
3 "___ and out I come"
4 Trig. function
5 Fermion
6 Extent
7 "... no other ___ before me."
8 Draft agcy.
9 ___ phase (of liquid crystals)
10 Inductor
11 Desire
12 South African
13 Woman's suffrage leader ___ Shaw
22 ___ soda (sodium carbonate)
24 "___ the greatest"
25 Microwave source
27 343's cubed root
28 Japanese drama
29 Bakery worker
30 Ripped
31 Unit of length
32 Element in steel
33 1944 Physics Nobelist
34 Aid a criminal
36 Monochromatic radiation source
39 County in SW Scotland
41 Trivalent rare earth
44 Nondispersive wave
46 Unit of length
49 State of matter
51 ___ velocity
53 Egypt's capital
54 Behaved like a disengaged engine
55 Combine
56 Snake in 71A
57 Turn the ___
58 Female sheep
59 Girl
60 Clamp
61 Ancient inhabitant of Britain
64 Hawaiian wreath

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JULY/AUGUST PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M121

Using the hint, we can estimate the denominators of the terms in the left side of the inequality in question:

$$(1+m)^{1/n} < 1 + \frac{m}{n},$$

$$(1+m)^{1/m} < 1 + \frac{n}{m}.$$

It follows that

$$\frac{1}{\sqrt[n]{m+1}} + \frac{1}{\sqrt[m]{n+1}} > \frac{n}{m+n} + \frac{m}{m+n} = 1,$$

and we're done.

To prove the version of Bernoulli's inequality given in the statement of the problem, we can consider the function $f(x) = (1 + x)^\alpha - \alpha x - 1$. Since $f'(x) = \alpha(1 + x)^{\alpha - 1} - \alpha$ for $x > 0$, this function decreases, so $f(x) < f(0) = 0$ for $x > 0$. (V. Senderov)

M122

Consider all n-digit numbers composed of ones and twos. There are exactly 2^n such numbers and as many possible remainders upon division by 2^n. So it will suffice to prove that all these 2^n numbers yield different remainders when divided by 2^n: then one (and only one) of these remainders will have to be zero.

We can do this by induction. We leave the case $n = 2$ to the reader, and assume that all $(n - 1)$-digit numbers consisting of ones and twos left different remainders when divided by 2^{n-1}. We will show that this induction hypothesis does not allow the existence of two n-digit numbers (of our form) with the same remainder upon division by 2^n.

Let a_n and b_n be any two of these n-digit numbers. Suppose the remainders of a_n and b_n upon division by 2^n are the same. Then a_n and b_n are both even or both odd—that is, they end with the same digit r ($r = 1$ or 2), so we can write them as $a_n = 10a_{n-1} + r$, $b_n = 10b_{n-1} + r$, where a_{n-1} and b_{n-1} are numbers composed of $n - 1$ ones and twos (a_n and b_n without their last digits). Since $a_n - b_n = 10(a_{n-1} - b_{n-1})$ is divisible by 2^n, $5(a_{n-1} - b_{n-1})$ is divisible by 2^{n-1}. This means that a_{n-1} and b_{n-1} give the same remainders when divided by 2^{n-1}. This contradicts the induction hypothesis.

M123

Let's call a square *suitable* if the portion of its area colored black is no greater than 4/5 and no less than 1/5. Since our grid is infinite, we can find an integer n big enough so that a certain square Q bounded by grid lines, and with a side of length 2^n, contains all black squares and their area is less than 1/5 of the area of Q. Divide this square into four congruent squares with the side length 2^{n-1}. In each of them, at most 4/5 of the area will be colored black. Those that have at least 1/5 of the area colored are suitable. All the rest have less than 1/5

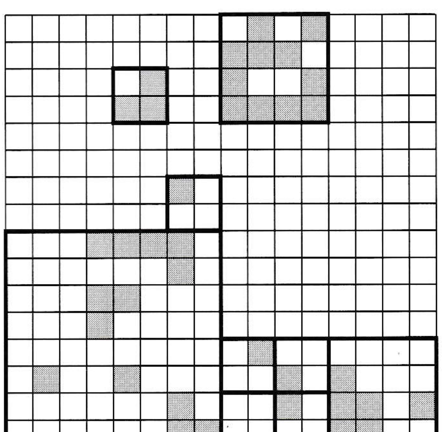


Figure 1

of the area colored black, so we can apply our subdivision process to them again, and so on (fig. 1).

After the $(n - 2)$nd subdivision we'll get a number of suitable squares and a number of 2×2 squares, each of which is at most 4/5 black. Those of the latter that contain at least one black square are suitable (they have at least $1/4 > 1/5$ of the area colored). None of the rest contains any black squares at all, so at this step all black squares will be covered by suitable squares, which was the goal.

The problem can be extended to space by replacing 1/5 and 4/5 by 1/9 and 8/9, respectively. The proof is modified in a natural way.

M124

This problem emerged from a mistake made by a student at Moscow University. Once he used the following lemma, which seemed obvious to him: *if a function is defined on a segment* $[a, b]$ *covered by a finite system of segments, and the variation of the function on each segment of this system is no greater than its length, then the variation of the function on the entire segment* $[a, b]$ *is no greater than* $b - a$. (The variation of a function in this case is the difference between its maximum and minimum values.) In our case we could take for the function the distance $s(x)$ that the snail crawls in a period of time x, $0 \le x \le t$, and for the segments that cover the interval $[0, t]$ we can use the intervals of observation of all the naturalists. Then, if the lemma were true, the entire path of the snail would not exceed t meters. This is what most people think about this situation—many even think that the snail must crawl at a constant speed of 1 meter per minute and travel exactly t meters during the entire given time interval. In fact, the path of the snail

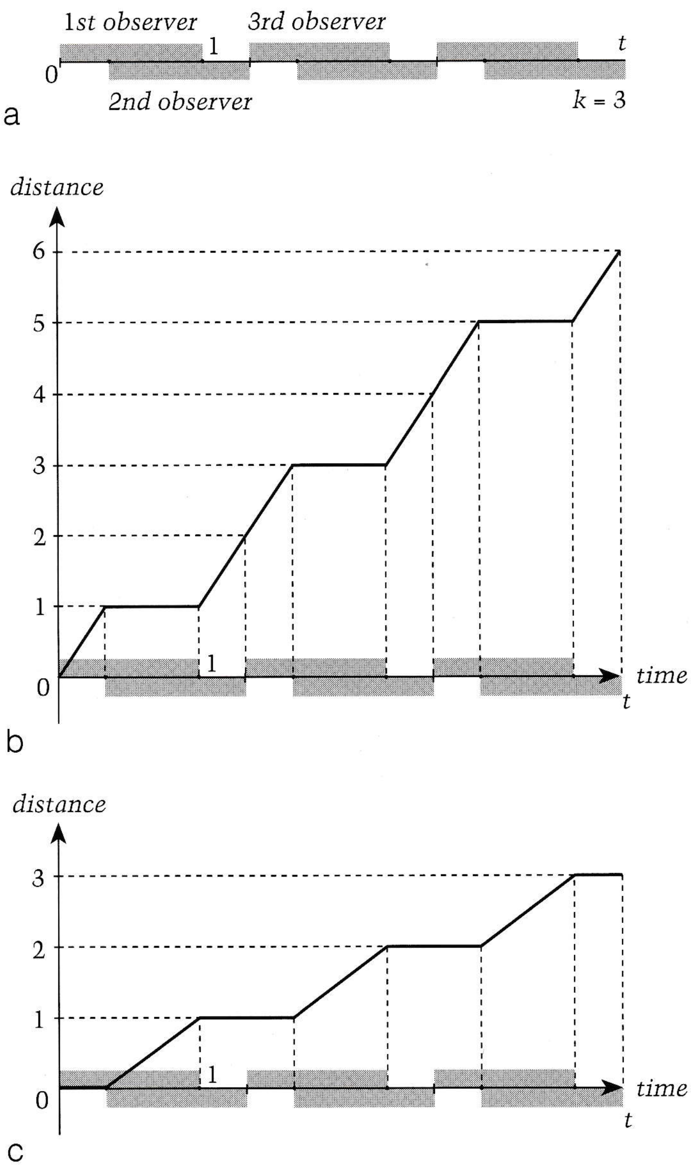


Figure 2

can range, roughly speaking, between $t/2$ and $2t$ meters. This apparent contradiction between common sense and actuality makes the problem particularly attractive.

More exactly, if m and M are the smallest and largest integers in the interval $(t/2, t)$, then the shortest and longest possible distances crawled by the snail are m and $2M$, respectively. (For $t > 1$ the numbers m and M are well defined.)

Let's explain how to arrange the observation and the movement of the snail to obtain these extreme values.

Take any integer k, $t/2 < k < t$ (we are particularly interested in $k = m$ and $k = M$). Divide the interval $[0, t]$ into k equal segments. The length t/k of each segment is greater than 1 and less than 2, so it can be covered with two overlapping unit intervals as shown in figure 2a. We arrange the intervals of observations of the naturalists to be these segments. The snail can decide to move only when it is observed by a single naturalist (fig. 2b), in which case it will crawl $2k$ meters. It can also decide to move only when it is observed by two naturalists (fig. 2c), in which case its total traveled distance will be k meters. Taking $k = 2M$ and m gives $2M$ meters and m meters for the length of the snail's path. (By the way, modifying the law of motion slightly, we can make the snail crawl *any* distance from m to $2M$. Check this yourself.)

It remains to show that the distance traveled cannot be greater than $2M$ or less than m.

For any order of observation, let $I_1 = [0, 1]$ be the first interval of observation, let I_2 be the rightmost interval intersecting I_1 (I_2 possibly has only a common endpoint with I_1, but the intersection of I_1 and I_2 cannot be empty, since the observation is never interrupted), and, in general, let I_{n+1} be the rightmost interval intersecting I_n. If there are l intervals in this sequence, then the last interval in this sequence, I_l, is $[t - 1, t]$. Notice that for $n = 2, \ldots, l - 1$ the intervals I_{n-1} and I_{n+1} are disjoint (otherwise, I_n would not be the *rightmost* interval intersecting I_{n-1}). Let k be the number of "odd" intervals $I_1, I_3, I_5, \ldots$ (so that $l = 2k$ or $l = 2k - 1$); then the entire interval $[0, t]$ contains k disjoint "odd" intervals and is completely covered by all the $l \le 2k$ intervals I_n. It follows that $k < t \le 2k$. Similarly, considering the l segments of the snail's path corresponding to the time intervals I_n, we can see that $k \le s \le 2k$. Since k is an integer, $k < t$ and $k > t/2$ (except for the case $t = 2k$), we get $m \le k \le M$, and so $m \le s \le 2M$ whenever $t \ne 2k$.

As to the exceptional case $t = 2k$, one can easily understand that here $l = t$, and $I_n = [n - 1, n]$ for all $n = 1, 2, \ldots, l$. Then the only possible value for s is $2k$, so the inequality $s \ge m$ is true again (here $m = k + 1$).

And here's an additional problem: correct the lemma formulated at the beginning of the solution. (N. Konstantinov, V. Dubrovsky)

M125

Our solution will be based on two useful ideas: the fact that the ratio of the distances to the sides of an angle is constant for all points on a line drawn inside the angle through its vertex, and the so-called "area method," which consists of using areas to express various geometric values.

Denote by $ABCD$ the given quadrilateral; 2α, 2β, 2γ, and 2δ are its angles at the vertices A, B, C, and D; O and I are the centers and R and r the radii of its circumcircle and incircle, respectively; and P is the intersection point of its diagonals (fig. 3). Since the quadrilateral has a

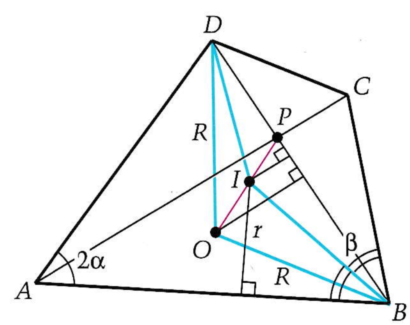


Figure 3

circumcircle, its opposite angles are supplementary. It follows that two of its consecutive angles are both acute (or right). Assume for definiteness that the angles 2α and 2β are not obtuse (and so 2γ and 2δ are not acute).

It will suffice to show that in this case points O and I lie in the angle APB and the ratios of the distances from these points to the sides of this angle are equal:

$$\frac{d(O, DB)}{d(O, CA)} = \frac{d(I, DB)}{d(I, CA)}$$

(the letter d here denotes distance).

Since the angle BCD is not acute, the circumcenter O lies on the same side of BD as point A (or on BD). The incenter I (where the angle bisectors of the quadrilateral intersect) also lies on this side of BD, because the angle at the vertex I of the quadrilateral $BCDI$ equals

$$2\pi - \beta - 2\gamma - \delta = \pi - (\beta + \delta) + \pi - 2\gamma$$

$$= \frac{\pi}{2} + 2\alpha \le \pi$$

($\beta + \delta = \pi/2$, since $ABCD$ is inscribed in a circle).

Similarly, both points O and I lie on the same side of AC as point B. So they lie in the angle APB.

Now the ratio of the distances from O and I to BD can be written as the ratio of the areas of triangles OBD and IBD with a common base BD—that is, it's equal to

$$\frac{OB \cdot OD \cdot \sin\angle BOD}{IB \cdot ID \cdot \sin\angle BID}.$$

But $OB = OD = R$, $\angle BOD = 2\angle BAD = 4\alpha$ (by the Inscribed Angle Theorem), $IB = r/\sin\beta$, $ID = r/\sin\delta = r/\cos\beta$ (see figure 3), and, as we've shown, $\angle BID = \pi/2 + 2\alpha$. It follows that

$$\frac{d(O, BD)}{d(I, BD)} = \frac{R^2 \sin 4\alpha \sin\beta \cos\beta}{r^2 \sin(\pi/2 + 2\alpha)}$$

$$= \frac{R^2 \sin 2\alpha \cos 2\alpha \sin 2\beta}{r^2 \cos 2\alpha}$$

$$= \frac{R^2}{r^2} \sin 2\alpha \sin 2\beta.$$

By the symmetry of the problem and of the obtained expression, the ratio $d(O, AC)/d(I, AC)$ is the same, which means that the equality of ratios we intended to prove is indeed true. (V. Dubrovsky)

Physics

P121

From the statement of the problem, the motion of the gliding airplane is uniform, so its weight $m\mathbf{g}$ and the drag force $\mathbf{F}_d$ are counterbalanced by the forces $\mathbf{F}_1$ and $\mathbf{F}_2$ (fig. 4).

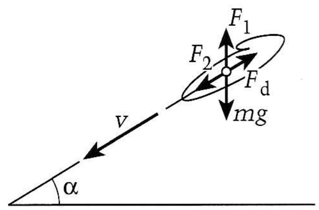


Figure 4

Therefore, $\mathbf{F}_1 = -m\mathbf{g}$ and $\mathbf{F}_2 = -\mathbf{F}_d$, where the forces $\mathbf{F}_1$ and $\mathbf{F}_2$ depend on the plane's velocity. For the sake of our estimate, we'll assume that the speed of the plane at takeoff is practically equal to v—that is, the forces $\mathbf{F}_d$, $\mathbf{F}_1$, and $\mathbf{F}_2$ are rotated through the angle α together with the velocity vector (figure 5). In this case the condition for the plane to have uniform motion without pressure on the runway (during takeoff the runway exerts no force) can be written as

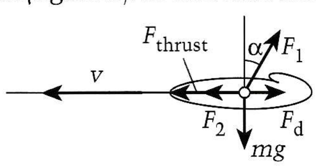


Figure 5

$$F_{thrust} = F_1 \sin\alpha,$$
$$mg = F_1 \cos\alpha.$$

Thus,

$$F_{thrust} = mg \tan\alpha \cong 1{,}700 \text{ N}.$$

There's no point in refining this estimate—the nature of the airflow around the plane near the ground is rather different from the case of a gliding plane, but we can't take that into account here.

P122

The work W performed on the system was converted to changes in the internal energy of the gas ΔU_g and in the potential energy of the piston ΔU_p:

$$W = \Delta U_g + \Delta U_p.$$

For one mole of monatomic ideal gas, the change in the internal energy is given by the expression

$$\Delta U_g = \frac{3}{2} R(T_x - T_0).$$

The change in the potential energy is equal to the work performed in moving the piston quasi-statically from its initial position to its final one. As this takes place, the external force that performs the necessary work must be equal at every moment to the force of gravity mg acting on the piston. Since the piston is in equilibrium in both the initial and final positions, this force of gravity is equal to the pressure of the gas in the vessel PS (the pressure of the external air is neglected). Denoting by Δh the change in the height of the piston, we get

$$\Delta U_p = mg\Delta h = PS\Delta h = P\Delta V,$$

where ΔV is the change in the volume of the gas. Applying the ideal gas law for 1 mole of gas yields

$$\Delta U_p = P\Delta V = R(T_x - T_0).$$

Thus,

$$W = \frac{3}{2} R(T_x - T_0) + R(T_x - T_0)$$

$$= \frac{5}{2} R(T_x - T_0),$$

and so

$$T_x = T_0 + \frac{2}{5}\frac{W}{R}.$$

P123

It is immediately obvious that if $\mu \ge \tan\alpha$, the puck won't move at all. So we'll examine the case where $\mu < \tan\alpha$.

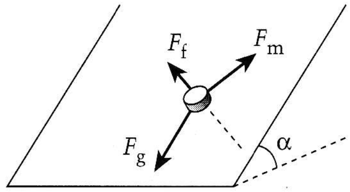


Figure 6

Let's consider the forces acting on the puck and lying in the plane (fig. 6). These are (1) the component of gravity directed down the plane and equal to $F_g = Mg \sin \alpha$, (2) the force of friction $\mathbf{F}_f$, directed counter to the puck's velocity $\mathbf{v}$ and equal to $F_f = \mu Mg \cos \alpha$, and (3) the magnetic force $\mathbf{F}_m$, which is perpendicular to the velocity $\mathbf{v}$ and equal to $F_m = QvB$. For steady-state motion the vector sum of all three forces is zero—

$$\mathbf{F}_g + \mathbf{F}_f + \mathbf{F}_m = 0;$$

or, taking into account that $\mathbf{F}_t$ is perpendicular to $\mathbf{F}_m$,

$$F_g^2 = F_f^2 + F_m^2.$$

From this we find the steady-state velocity

$$v = \frac{Mg}{QB}\sqrt{\sin^2 \alpha - \mu^2 \cos^2 \alpha}$$

and the angle between the vectors $\mathbf{v}$ and $\mathbf{F}_g$

$$\beta = \arcsin \frac{\mu}{\tan \alpha}.$$

P124

The main challenge in solving this problem is just keeping track of all the possible ways of connecting the resistors. The resistors can be connected in series, in which case the maximum power is 2 W (the maximum current in this scheme is determined by the 3-ohm resistor). When all the resistors are connected in parallel, the power will not exceed (1 + 1/2 + 1/3) W = 11/6 W, because the maximum voltage is determined by the 1-ohm resistor. There are three other diagrams corresponding to figure 7 and three more corresponding to figure 8. The "champion" is the scheme shown in figure 7, where the 2- and 3-ohm

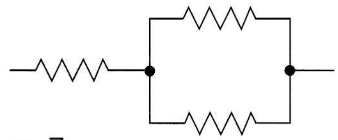


Figure 7

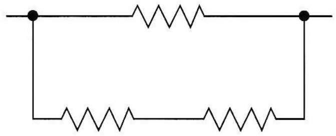


Figure 8

resistors are connected in parallel. The maximum power is

$$P_{max} = \left(1 + \frac{6}{5}\right) \text{W} = 2.2 \text{ W},$$

and the voltage is 11/5 V.

P125

The "billiard method," well known from geometry, is a handy way to construct the path of a beam undergoing multiple reflections from a flat surface. (In our case the reflecting surface is conical, but it doesn't matter for the rays we are interested in.) After colliding with a wall, a ball "jumps" onto the mirror-image table A^* and continues to move along a straight line (fig. 9). (We can replace "wall" with "reflecting surface" and "ball" with "light beam.") In our problem it's sufficient to consider the passage of the "extreme" ray at the edge of the large base. After the first reflection it enters the neighboring "added" cone, and so on (fig. 10). The small

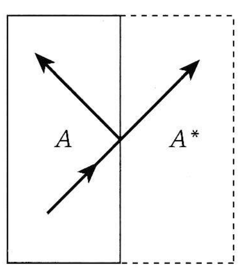


Figure 9

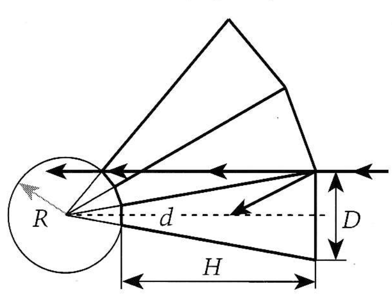


Figure 10

diameters d turn out to be the sides of an equilateral polygon inscribed in a circle of radius R. Under the conditions of the problem ($d << D << H$), we can assume that an arbitrary ray will pass through the plane of the small base if

$$\frac{D}{2} < R.$$

A little more geometry tells us that

$$R \cong \frac{d}{D-d} H.$$

From this we get

$$\frac{D}{2} < \frac{Hd}{D-d},$$

or

$$H > \frac{D(D-d)}{2d} \cong \frac{D^2}{2d}.$$

Thus, if $H > D^2/2d$, all the rays of the incident beam will pass through the plane of the small base of the conical light conductor.

Brainteasers

B121

Let K be the midpoint of AB (fig. 11). Then DK is parallel to NB, so $\angle BPM = \angle KDM$. By the symmetry of the rectangle, $\angle KDM = \angle NAM$.

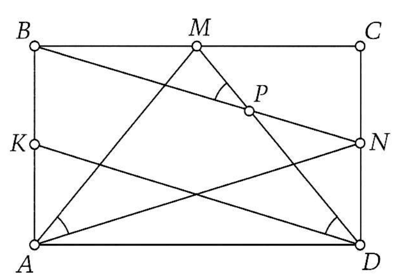


Figure 11

B122

The answer is $650 \cdot 9 = 5{,}850$. The equation can be rewritten in the form $9(100 \cdot \text{O} + \text{NE}) = 100 \cdot \text{NI} + \text{NE}$, or $900 \cdot \text{O} = 100 \cdot \text{NI} - 8\text{NE}$. It follows that NE is divisible by 25—that is, we must consider three cases: NE = 25, NE = 50, and NE = 75. Substitution leads to the equations

(1) NE = 25, $18 + \mathrm{I} = 9 \cdot \mathrm{O}$,
(2) NE = 50, $46 + \mathrm{I} = 9 \cdot \mathrm{O}$,
(3) NE = 75, $64 + \mathrm{I} = 9 \cdot \mathrm{O}$.

Equation (1) has no solution, because $\mathrm{I} < 9$, $\mathrm{O} \neq 2$; equation (2) has a unique solution I = 8, O = 6, and yields the answer given above; and equation (3) gives I = O = 8, which is impossible. (V. Dubrovsky)

B123

Let the dimensions of the brick be a, b, c, its density ρ, and the pressure in the aforementioned three positions p_1, p_2, p_3. Then

$$p_1 = \frac{\rho gabc}{ab},$$

from which we get $c = p_1/\rho g$. Similarly, $b = p_2/\rho g$ and $a = p_3/\rho g$. Therefore, the brick's mass is

$$m = \rho V = \rho abc \frac{p_1 p_2 p_3}{\rho^3 g^3}.$$

As the pressure exerted by a wall of height h is $p = \rho gh$, then substituting this value ρ into the equation for the brick's mass yields

$$m = \frac{p_1 p_2 p_3 h^2}{p^2 g} = \frac{1{,}368 \cdot 2{,}581 \cdot 5{,}404 \cdot 16}{88{,}200^2 \cdot 9.8} \cong 4 \text{ kg}.$$

B124

The word *scretibo* is followed by *scorbite*; *biscrote* is preceded by *bisector*; *iscetbor* is followed by *trobiecs*; and the last word in the dictionary is *scetibor*.

It's convenient to replace the letters of the "key word" *robitecs* with the digits 1, 2, 3, ..., 8, respectively. Then every "word" becomes an eight-digit number composed of these digits in a certain order, and the problem is reduced to arranging these numbers in increasing order. The general rule can be formulated as follows: read a given number (permutation) $a_1a_2...a_8$ from the right, find the first digit a_n such that $a_{n-1} < a_n$, and find the *smallest* of the digits a_n, a_{n+1}, ..., a_8 that is greater than a_{n-1}—say, a_k. Then write the digits $a_1...a_{n-2}$ in the initial order; after them write a_k and then all the remaining numbers in increasing order. For instance, in 75348621, $a_n = 8$ ($n = 5$), $a_k = 6$, and the next number in the list is 75361248. (V. Dubrovsky)

B125

In the position shown in figure 12 none of the kings will lose if it goes every time to the opposite end of the longer diagonal it occupies. So it's a draw. However, it's not very obvious whether this position can emerge at the end of a real game in which the players don't miss any opportunities to give away their last king.

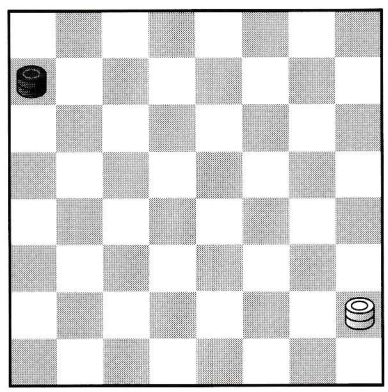


Figure 12

Polish Checkers, also known as CONTINENTAL CHECKERS, board game, variety of Checkers (Draughts) most played in continental Europe. The game is played on a board of 100 squares with 20 men on a side. The men move and capture as in English Draughts and American Checkers, except that in capturing they may move backward as well as forward. A piece is promoted to queen (or king) when it ends its move on the opponent's back line. If on reaching that line more captures are possible, they must be made; the piece involved is not promoted until it again reaches the back line and remains there at the end of a move. A queen may move any number of squares along a diagonal, capturing any unguarded pieces in its path and changing diagonals to take any other unguarded pieces on the board. In capturing, the player must choose the direction by which he can take the greatest number of pieces, whether men or queens, under penalty of huffing (forfeiting the offending piece).

The German game Damenspiel is Polish Checkers played on the 64-square chessboard with 12 men on a side, as in Draughts or Checkers. (*Encyclopaedia Britannica*, 15th ed.)

Toy Store

A thief and a clever wife. See figure 13.

A greedy moneychanger. See figure 14

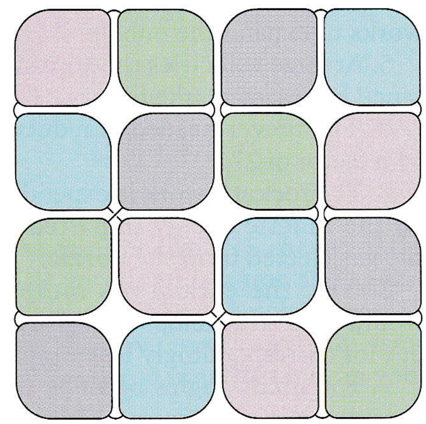


Figure 13

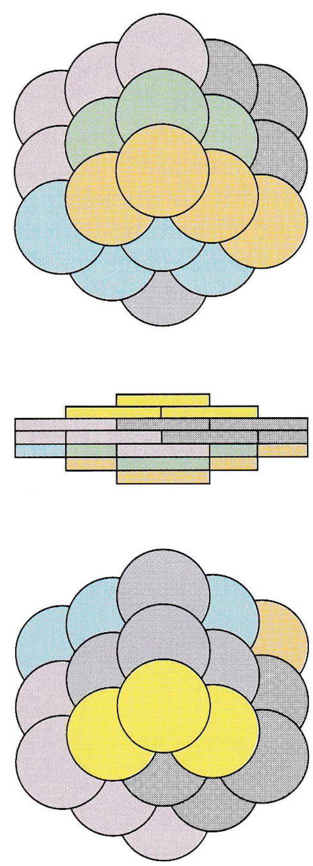


Figure 14

Kaleidoscope

1. No mechanical work is performed.

2. The total work performed equals zero both in the reference frame associated with the Earth and in that of the moving train.

3. Yes—for example, when it acts upon a weight lying on the floor of a moving railway car.

4. Yes: as the bubble rises, the hydraulic pressure drops and the gas works to expand the bubble.

5. At large velocities the air resistance increases markedly.

6. The power needed is reduced by a factor of 27.

7. The rocket's work is expended on the kinetic energy imparted to the gas ejected from the nozzle.

8. Yes, the escalator's motors would need to develop more power, but for correspondingly less time.

9. For small angles of slope the conveyer is more efficient, since the rolling friction is less than the sliding friction.

10. Yes, but its efficiency would be very low, because most of the work will be expended on compressing the gas itself.

11. It's doubtful, because there will be an increase in energy losses associated with thickening of the crankcase oil, and also with the need to warm up the engine and the air in the passenger compartment.

12. The heat source is the combustion chamber and the heat sink is the surroundings.

13. The efficiency will be greater when the loads are connected in series.

14. No, because the efficiency approaches one as the load's resistance approaches infinity. However, the power of the battery and the power of the resistor both approach zero.

Microexperiment. As the temperature increases, the energy losses due to radiation and heat convection also increase.

Algebraic identities

1. The equality is true if points O, A, B are collinear.

2. Show that if O is the origin, then the complex number h represented by the orthocenter H is equal to $a + b + c$ (where a, b, c are the numbers represented by A, B, C). Note that $h - a = \alpha(b + c)$, $h - b = \beta(c + a)$, $h - c = \gamma(a + b)$, where α, β, γ are real numbers, and derive therefrom that $\alpha = \beta\gamma = 1$.

3. Use the identity $(a - b)(d - c) + (b - c)(d - a) + (c - a)(d - b) = 0$. To prove Ptolemy's theorem, show that the complex numbers $(a - b)(d - c)$ and $(b - c)(d - a)$ have equal arguments (this follows from the Inscribed Angle Theorem).

4. Check that replacing the lengths AB, BC, CA, AM, BM, CM with the complex numbers $b - a$, $c - b$, $a - c$, $m - a$, and so on, yields an identity for complex numbers.

Index of Advertisers

"BULLETIN BOARD" CONTINUED FROM PAGE 55

The deadline for entries is February 1, 1995. If you don't receive an entry kit by October 15, contact Toshiba/NSTA ExploraVision Awards Program, 1840 Wilson Blvd., Arlington VA 22201-3000 or call toll free 800 EXPLOR-9.

Learn and Serve America grants

In June the Clinton Administration named 65 colleges and universities winners of new Learn and Serve America higher-education grants to establish service-learning programs that benefit local communities. Totaling \$6.75 million, the new Learn and Serve America grants are provided by the Corporation for National Service and enable colleges to make community service an integral part of students' educational experience.

The Corporation's flagship program, AmeriCorps, will be fully launched in September. By year's end it will engage up to 20,000 members in critical service to communities nationwide. In exchange for one or two years of service, AmeriCorps members will receive education awards to finance their higher education or repay college loans. AmeriCorps members work within the AmeriCorps national service priorities of education, public safety, human needs, and the environment.

For more information, contact the Corporation for National Service, 1100 Vermont Ave. NW, Washington DC 20525 (phone: 202 606-5000).

Gelfand named a MacArthur Fellow

I. M. Gelfand, a Distinguished Visiting Professor at Rutgers University, has received a \$375,000 fellowship from the John D. and Catherine T. MacArthur Foundation. He is the founder of the Gelfand Outreach Program and the Gelfand Mathematics Correspondence School (see the November/December 1993 and January/February 1991 issues of *Quantum*). For information about these programs, contact Harriet Schweitzer at 908 932-0669 (e-mail: harriets@gandalf.rutgers.edu).

Diamonds from a jug

Two tales with a brainteasing twist

by Sergey Grabarchuk

EDITOR'S NOTE: IN THIS INstallment of our Toy Store we present two stories from *A Jug of Diamonds*—a collection of tales by Sergey Grabarchuk. Each of the stories is built around a puzzle (generally a "dissection," "assembly," or "fit-in" puzzle) created by this talented puzzle designer from Uzhgorod (Ukraine). *Quantum* readers should already be familiar with his "triad" puzzle (see the November/December 1993 issue).

A thief and a clever wife

One night a thief sneaked into the house of a merchant and stole the jewelry of the merchant's wife. As he was leaving, he failed to notice the master of the house returning at the same time. When the merchant saw a stranger going out the gate, he sensed that something was amiss and secretly followed him. Some time later he found himself under the windows of the thief's home, where he could hear very well what was going on inside. So he heard the thief entering his home and saying to his girlfriend: "Quick, put on all this jewelry, and if anybody asks, tell them you inherited it from your mother."

Early the next morning the merchant went to the judge and recounted the whole story. The judge ordered the guards to bring in the thief, his girlfriend, and the merchant's wife, and he addressed the thief:

"Is it true that you stole the valuables of this worthy man's wife?"

"No, Your Honor, that's a vicious slander," the thief replied, since he was not only deft but very sly, too. "This is my girlfriend's jewelry—she inherited it from her mother."

And the thief's girlfriend swore that that's exactly the way it was, and

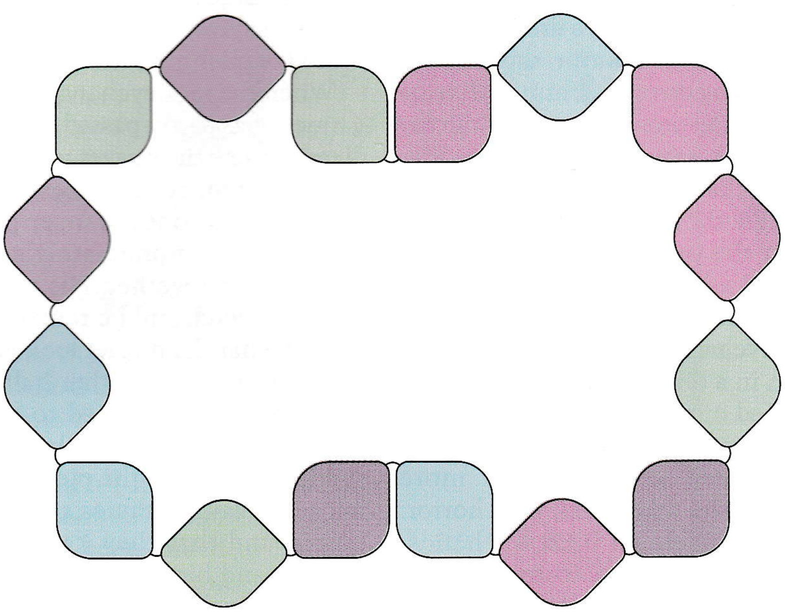


Figure 1

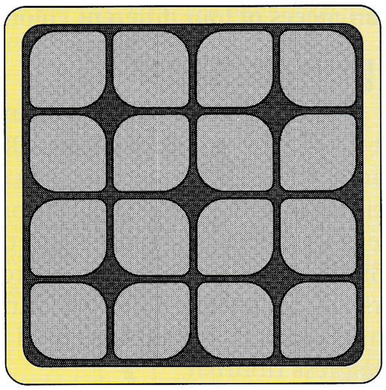


Figure 2

that the merchant lied. But then the merchant's wife said:

"I have proof that this bracelet—which this brazen hussy has defiled by putting on her arm—belongs to me."

She produced a small box and gave it to the judge, saying:

"Let Your Honor order this good-for-nothing to hand you the bracelet. Then I'll be able to prove that the bracelet and the box were made for each other and match each other perfectly."

The judge opened the box and saw hollows carved in its bottom—there were sixteen of them. He ordered that the bracelet be removed from the thief's girlfriend's arm and gave it to the merchant's wife. With a few deft movements, the merchant's wife laid the bracelet in the box so that each piece of the bracelet fit exactly into its nest, and the entire bracelet went into the box without leaving any hollows empty.

"But that's not all," the clever woman said. "Let Your Honor see that any straight row consisting of four pieces contains pieces of four different colors!"

The judge then realized that the merchant's wife was telling the truth. The thief and his girlfriend understood that they had been unmasked, and they confessed to their crime.

And now, dear reader, can you figure out how the merchant's wife laid the bracelet (fig. 1 on the preceding page) in the box (fig. 2)? An indispensable condition is that all straight rows of four pieces—horizontal, vertical, and diagonal—must contain pieces of four different colors.

A greedy moneychanger

There once was a moneychanger at a bazaar in Damascus. This moneychanger was very greedy and stingy, but he had the reputation of being a good-hearted simpleton. Thanks to this reputation, customers visited his shop more willingly than others, and this helped his business flourish.

One day a blind old man entered his shop and asked to exchange a small bag of gold coins for dinars. When the moneychanger first saw the coins that the old man gave him, he was dumfounded: the coins were unknown to him and very strange. But, biting one, he convinced himself that they were pure gold. He was just about to count out the required number of dinars when the inevitable occurred: the moneychanger was seized by greed and, unable to control himself, with trembling hands counted out only half of the dinars he owed the old man. He poured them into the bag and handed it to the blind visitor. The old man thanked him and went away. The moneychanger closed the shop and secluded himself in a distant room, where he took out the old man's coins to examine them thoroughly. In the semidarkness of the room they glimmered magically, and that soothed the swindler's conscience. He sorted the coins and then began to arrange them in neat piles, flat rows, or just small heaps. He played with the coins for a long time, until evening fell and he ceased this activity. The moneychanger prayed, thanked Allah for a successful day, and went to bed. Contrary to his habit, he didn't hide the old man's coins in a chest, but left them on a small table.

He slept very well and had no dreams. But towards morning he saw the blind old man in a dream. To the moneychanger's horror, the old man was staring directly at him with tears rolling from one of his eyes, while the other one squinted strangely. The moneychanger came up to the table and saw the coins where he had left them the previous evening. They formed a beautiful pile and looked like seashells. This startled the moneychanger, because he had left them in a different arrangement. He reached out for the coins, but felt a certain awkwardness in his hands. He looked at his hands more closely—he was struck with horror: all the fingers on both his hands, except his thumbs, were stuck together as if glued. He broke out in a cold sweat, felt weak in the knees, and sank onto a mat near the table. He shot another look at the coins and tried to sweep them off in his hand. What happened next only intensified his horror. The coins, like his fingers, had become glued together, too, and became a kind of nugget, or one solid piece. The moneychanger kept turning this nugget in his hands, unable to shake off his horror and astonishment.

Suddenly something rustled in the darkest corner of the room. The moneychanger turned his head, saw the same blind old man, and remembered his last dream. He wanted to get up and rush toward the old man, but his legs wouldn't obey him. He tried to say something, but he couldn't even move his tongue. The old man, as if stopping him, raised his hand and spoke: "You see, wicked man, how, cheating me, you outwitted yourself. What good is my money to you if you cannot use it? You won't dare to appear in public with such hands. Your business will be ruined, and you'll become a beggar like me. But I feel that you haven't fallen low enough yet, and I want to give you a chance to expiate your guilt. Throw the thing you're holding onto the floor, and when it splits into several parts try to assemble it again from the pieces. You'll see what will happen then. But hurry, because every hour it will become harder and harder to do." And so saying, the old man disappeared, vanishing into thin air.

When the moneychanger's astonishment has partly passed, he immediately threw the nugget to the floor. Sure enough, it split into several pieces. The moneychanger picked them up and immediately started fitting them together. His memory was very good, and he remembered clearly what the nugget looked like. So he easily put together half of it, and more, but he failed to get beyond that point. Not only that, he suddenly felt that the pieces of the nugget became heavier and a little bigger, and that they kept getting heavier and bigger by the minute. When he finally reached the point of inserting the last piece, the moneychanger could hardly lift it.

He managed to put one of its edges in place and was trying to adjust the whole piece, when all of a sudden it slipped out of his hands and sank into place, pinching the money-changer's thumb. He was pierced by a sharp pain and fainted, collapsing near the nugget . . .

The moneychanger woke up, as always, before sunrise. He didn't remember his dream. He said his morning prayers and walked up to the table. The coins were lying there as he had left them the previous evening. When he reached out for them, he suddenly recalled everything he saw in his dream. He broke out in a cold sweat. With a sinking heart he tried to sweep them off into his hand and . . . managed to do it. The coins fell one after another with a gentle tinkling. And when the last coin fell, the moneychanger suddenly felt weak in the knees and sank down onto the mat near the table. It's hard to say how long he sat there. When he regained consciousness, he walked into his shop, stuffed his pockets full of money, and went out into the bazaar. He generously gave alms to all the beggars along the way.

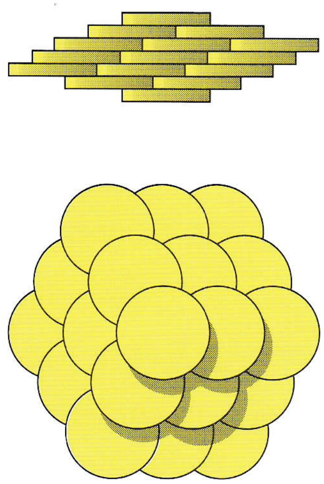


Figure 3

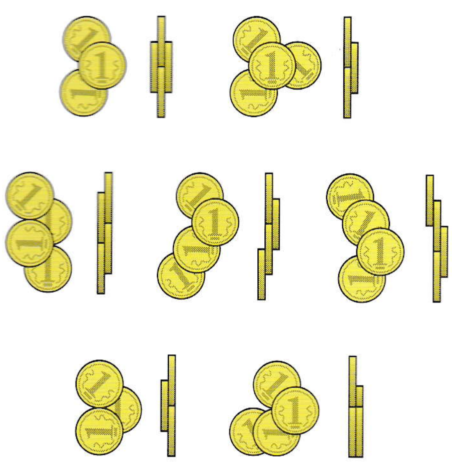


Figure 4

From that day on he did the same thing every morning, but he never came across that blind old man. And all his customers soon noticed that an inexplicable generosity had been added to his good heart and simple mind.

And now, dear reader, can you construct the nugget in figure 3 from the seven pieces in figure 4?

SOLUTIONS ON PAGE 61