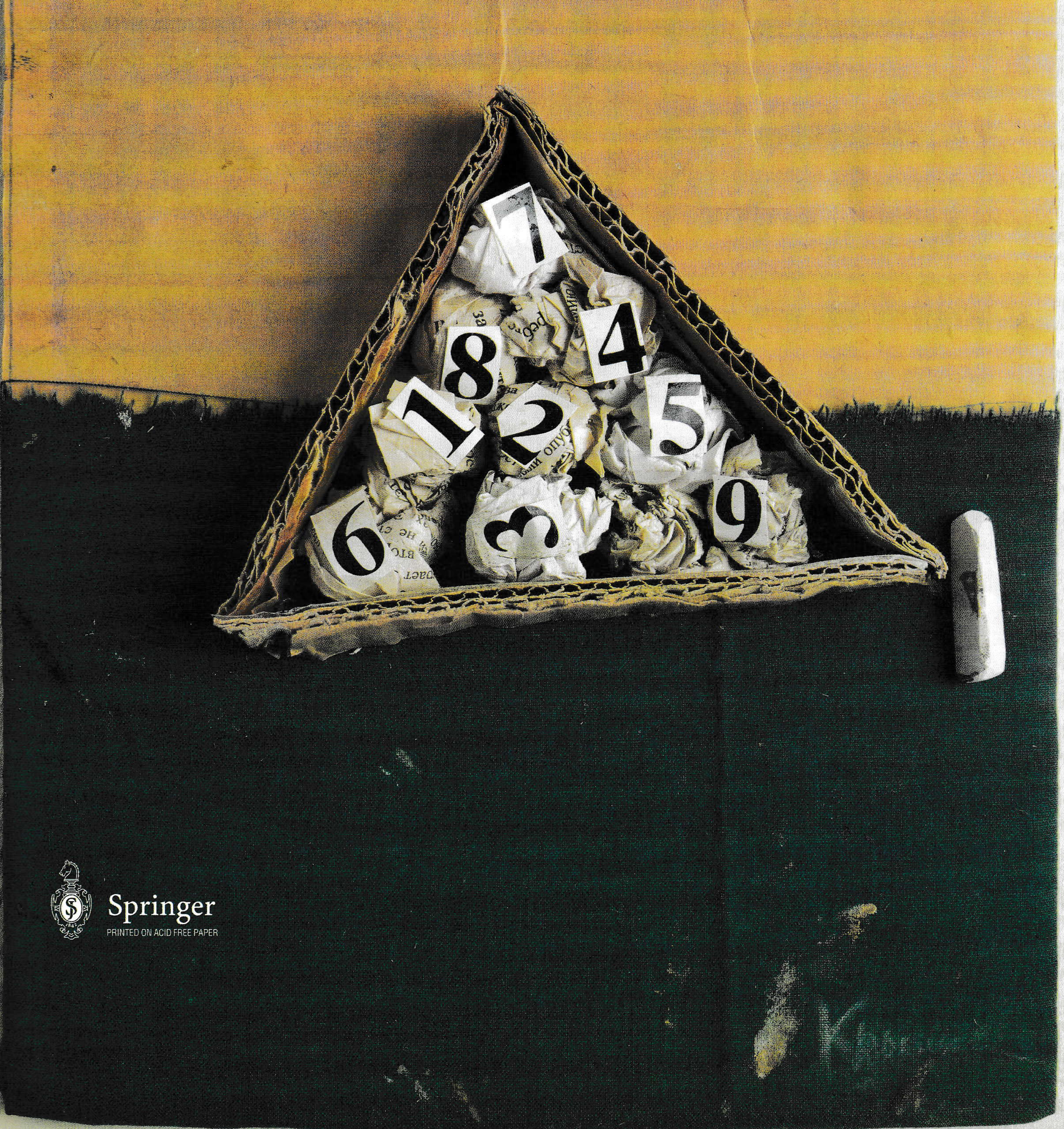
QUANTUM

NOVEMBER/DECEMBER 1996

US $7.50/CAN $10.50



Springer
PRINTED ON ACID FREE PAPER

The Fall of Phaeton (c. 1605) by Sir Peter Paul Rubens

EVERY CULTURE, IT SEEMS, HAS MADE A GOD OF the Sun at some point in its history, or has created a story that places it squarely at the center of our lives—at the intersection of our hopes and fears. The power and stability of that fiery orb are supremely important for life on this planet. One can easily understand the panic our ancestors felt when the Sun inexplicably disappeared in the middle of the day. Even today we shiver during a solar eclipse, and not just because of the temperature suddenly drops several degrees.

The story of Phaethon (sometimes spelled Phaeton) deals to some extent with an ancient fear that the Sun will do the harm it seems so capable of. But like most Greek myths, it resonates on many levels. Phaethon was the son of the sun god Helios and the sea goddess Clymene. Taunted with illegitimacy, he appealed to his father. Helios promised to prove his paternity by giving Phaethon whatever he wanted. Phaethon asked that he be allowed to drive the chariot of the sun through the heavens for just one day. Full of misgivings but bound by his oath, Helios agreed. Phaethon set off, but was unable to control the horses of the sun chariot. He swerved too close to the earth and began to scorch it. The mighty Zeus was displeased and struck Phaethon down with a lightning bolt to prevent further damage.

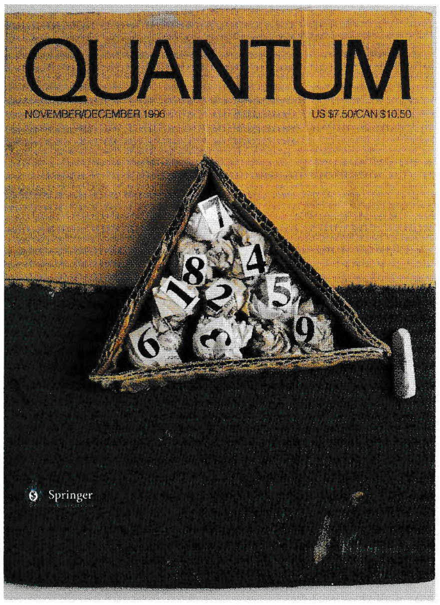
Readers may be familiar with another Greek story involving the Sun. The legendary architect and sculptor Daedalus fashioned two marvelous sets of wings that would allow him and his son Icarus escape the island of Crete, where Daedalus had fallen out of favor with King Minos. Icarus, however, flew too close to the Sun. The wax holding the wings together melted, and Icarus fell into the sea and drowned. "This is what comes of our vain aspirations," some would conclude. "Better to stay put."

But, of course, we don't stay put. We fly off into space, or send our machines where we can't go ourselves. Turn to page 16 to learn what it really takes to "fly to the Sun."

QUANTUM

NOVEMBER/DECEMBER 1996

VOLUME 7, NUMBER 2



Cover art by Dmitry Krymov

The game of billiards has so often been associated with the study of physics, it has become a cliché to speak of them in the same breath. This relaxing pursuit has provided a metaphorical anchor for discussing the collision of particles, the reflection of light rays, and undoubtedly other topics as well.

But the game also appeals to others with a more purely mathematical turn of mind. The Kaleidoscope in this issue exposes some of the math behind the fancy shots. It also experiments with some new and, at times, bizarre shapes for the pool table itself. (In rural Wisconsin, somewhere between Milwaukee and Madison, there is a pool hall that invites passers-by to play on its L-shaped and Z-shaped tables. Could the proprietor be a mathematician manqué?)

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE,* and *General Science Source.* Available in microform, electronic, or paper format from University Microfilms International.

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Lunar miscalculation

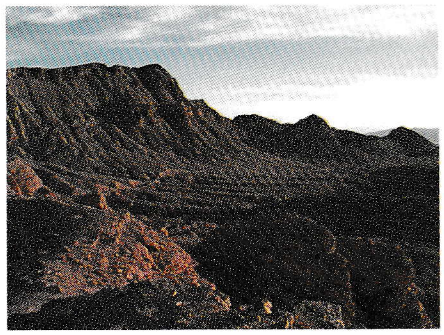
How I got stranded on a remote, rocky bluff in pitch darkness

ON THURSDAY, SEPTEMBER 26, 1996, there was a lunar eclipse. I live on the southeast edge of Las Vegas, in an area called Green Valley. The eclipse was to begin at 6:15 P.M. PDT, which was shortly before sunset.

At about 4:30 P.M. I loaded my car, packing a digital camera that uses no film, but rather stores images in memory on a PCMCIA card. You can then download the shots to a computer and use them however you wish. I also loaded my 8mm video camera and tripod. Then I drove off toward Lake Mead, trying to get away from the city lights—of which there are plenty in Las Vegas—and find a suitable observation spot.

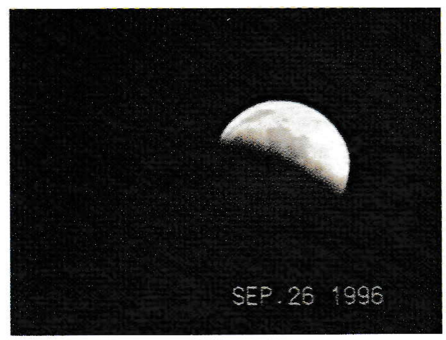
I drove to the Lake Mead National Recreational Area and along the North Shore Road some 20 miles or so until I found a stop-off that had a trail leading to the top of a very rocky and rugged hill. It would be a great vantage point for observing the eclipse.

Remembering the horror stories of people in the desert dehydrating and dying for lack of water, I always carry water in my car. So I carried a bottle with me as I made my way up the very difficult trail, stopping frequently to catch my breath, since I don't get enough exercise. The following pictures show one view from the top of that hill. The first was taken with my digital camera; the second is a frame from the video I took and shows the surrounding terrain more vividly.



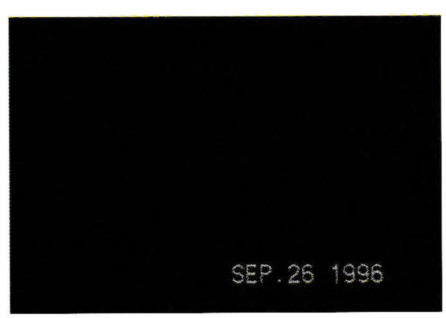
I set up my equipment and aimed my video camera toward where I expected the moon to rise in the east. Unfortunately there were some clouds on the horizon in that direction, so I would be unable to record the earliest part of the eclipse.

At about 6:30, I still hadn't observed anything, and the Sun had not yet set. A backpacker walked by, said hello, and headed for the interior of the park. A few minutes later a woman by the name of Marcia Palmer hiked up to my location—she was also there to see the eclipse. Later she would be joined by a friend, Darlene Guadalupe, and Darlene's 14-year-old son Kevin. We all sat on the rocks, waiting for the eclipse to be visible. At about 7:15, we began to see the eclipse, and I began to photograph it. A little later, the video captured the image printed below. As you can see, the eclipse had started, and we had a nice view of it.



All was going well until the eclipse became full. At that time, I had expected it would last 10 or 15 minutes and be over with. Then I would pack up and go home. But I hadn't done my homework. I was confusing lunar and solar eclipses. A solar eclipse doesn't last more than about 10 minutes at any one location, whereas a lunar eclipse lasts for more than 3 hours, and for an hour and forty minutes of that time, the eclipse is *total*!

The following frame from the video shows what my camera was able to record during totality:

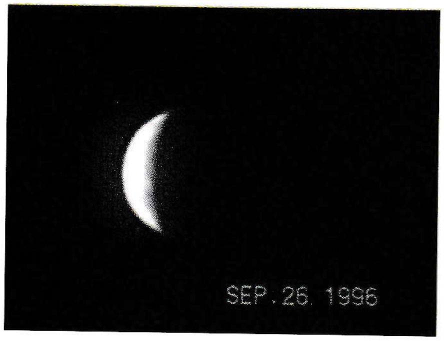


Note that the camera didn't pick up the peculiar illumination of the moon most of us observed with our eyes. It was pitch black out there on that mountain!

We waited, and we waited. But the total eclipse seemed never to change! It was then that I realized that this thing would last awhile. So I thought I'd pack up and go home. But then I looked around for the trail. I could see *nothing*! None of us had a flashlight. We didn't even have matches. There was nothing we could do. We simply had to wait until there was enough moonlight that we could see the trail, so that we could make our way down the mountain.

At 9:30 P.M., the crescent of the moon began to appear, and it produced enough light that we could make out the trail down the mountain. We all moved slowly and carefully, trying to be sure we didn't disturb some rattlesnake basking in the moonlight on the trail. I got home at 10 P.M., so I had been gone some six hours—not quite the one-hour jaunt I thought I would be taking!

Here is the picture I captured right before packing up and heading down the mountain. This is the source of our light for the return trip.



Next time I go to observe a lunar eclipse, I'll try to remember a few of the "Eclipse Rules of Thumb" found in *The Moon Book* by Kim Long (Johnson Books, 1880 South 57th Court, Boulder CO 80301). These rules involve information like the length of an eclipse and the fact that they always occur in pairs: "A solar eclipse is always followed or preceded by a lunar eclipse within an

CONTINUED ON PAGE 14

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 7 (6 issues) will be published in 1996–1997. *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.

Periodicals 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: $17.70; Personal rate (non-student): $23.70; Institutional rate: $45.00; Single Issue Price: $7.50. 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: Personal rate is US$31.70 including surface delivery charges; institutional rate is US$45.00; airmail delivery is US$13.00 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

The advent of radio

How and why the wireless telegraph was invented 100 years ago

by Pavel Bliokh

BY THE MIDDLE OF THE 19th century the telegraph had become a common method of communication. Wires ran between towns large and small, and in 1866 a special cable was laid on the floor of the Atlantic Ocean—the telegraph now connected two continents. Messages could be quickly and reliably transmitted over huge distances (to and from anywhere on Earth), but only if one indispensable condition was met: before you could send a signal from one place to another, someone had to have traveled the entire distance, stretching an electric wire behind. This was no trifling problem, and it was actually insoluble when it came to moving objects (such as ships).

It's easy to imagine the great attention devoted to the problem of wireless transmission over large distances by scientists and engineers at the end of the 19th century. You might think this problem could have been solved quite easily on the basis of well-know physical principles that had been discovered more than 100 years ago. Some simple methods of transmitting signals were indeed available, and we begin our story with them.

Ghost invention No. 1

The "electrostatic" wireless telegraph could have been invented as early as the end of the 18th century. It's known that any electrically charged body generates an electric field with intensity

$$E = \frac{k_e Q}{r^2}. \qquad (1)$$

This formula is true when the body is "pointlike"—that is, when its size is very small in comparison with the distance r from it. Here Q is an electric charge and k_e is a proportionality factor, which depends on the system of units. It's very important that the electric field arises around a charged body without any wires, even in a vacuum. One can detect this field by the electric force

$$F = qE = k_e \frac{Qq}{r^2} \qquad (2)$$

acting on a test point charge q, which is either attracted to the source of the field (when the signs of Q and q are opposite) or repulsed from it (if the signs are the same).

The simplest diagram for a wireless telegraph that uses the force F is shown in figure 1. The "transmitter" is a body whose charge $Q(t)$ varies according to the transmitted signal $V(t)$. The "receiver" is made of two opposite charges $\pm q$ fixed on a small rod that is suspended on an

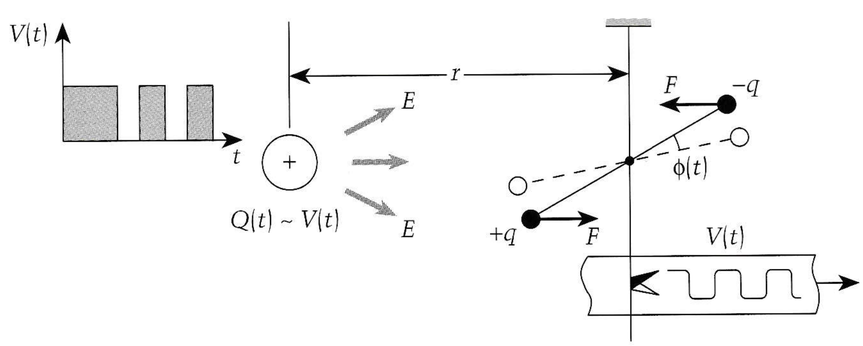
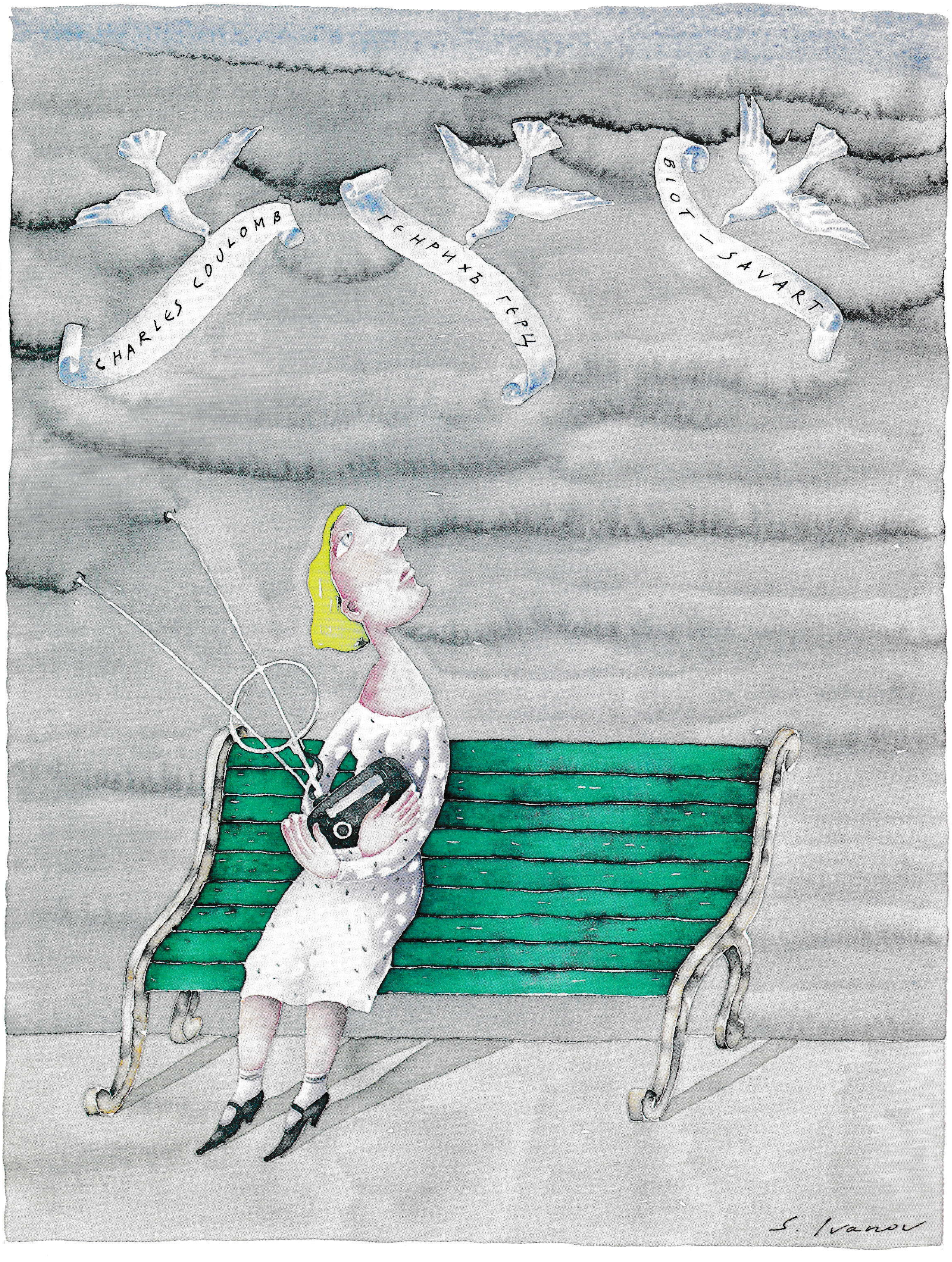


Figure 1
Simplified diagram for an "electrostatic" wireless telegraph.

Art by Sergey Ivanov


CHARLES COULOMB
ГЕНРИХ ГЕРЦ
BIOT — SAVART
S. Ivanov

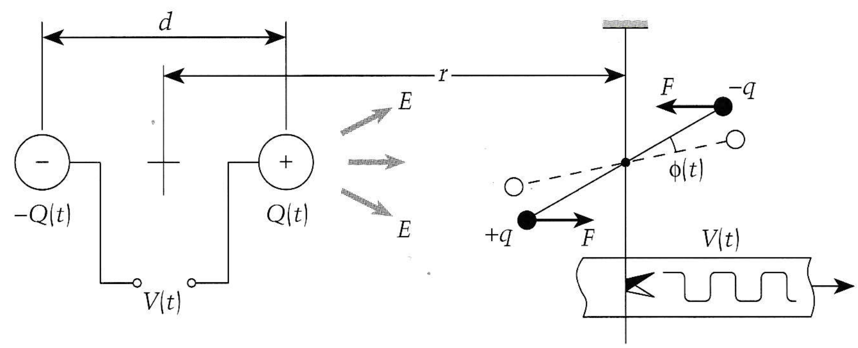


Figure 2
Actual diagram for an "electrostatic" wireless telegraph.

elastic string. In principle, the receiver needs only one charge, but then it would be less sensitive. The electric force F produces a torque that turns the rod by some angle $\phi(t)$ proportional to $Q(t)$. By registering that angle on a chart recorder, we can reproduce the source signal $V(t)$—so the device works!

The receiver described (known as the Cavendish torsion balance) was actually used more than 200 years ago by the French physicist Charles Coulomb. Using this balance, he measured the force of interaction F between two charges Q and q, and discovered the fundamental law of electrostatics (formula (2)), which was named after him.

Since the design we've looked at seems quite workable, the question arises: why wasn't it suitable for transmitting signals over large distances? First of all, remember that the field strength E fades with distance as $1/r^2$. Skipping ahead a few steps, I should point out that the field strength of a radio wave in free space decreases much more slowly—as $\sim 1/r$. This is a rather significant difference, but in reality it's even more drastic, as we see below.

The problem is, figure 1 contains a fundamental mistake. In reality the electric charge cannot increase or decrease by itself, because charge is conserved. For an object's charge Q to change, it must be connected to a voltage source, which always has two poles and simultaneously produces charges of opposite sign. So an actual implementation of an electrostatic wireless telegraph would look like figure 2.

We see that the work of a transmitter composed of two pulsating charges $\pm Q(t)$ will be less effective, because the two generated electric fields $\mathbf{E}_+$ and $\mathbf{E}_-$ partially cancel each other (the resulting field is the vector sum $\mathbf{E}_+ + \mathbf{E}_-$). We'll simplify the corresponding calculation and consider the case when the receiver is located along the line connecting the two charges $\pm Q$. Then $\mathbf{E}_+ = k_e Q/r_+^2$ and $\mathbf{E}_- = -k_e Q/r_-^2$, where $r_+ = r - d/2$, $r_- = r + d/2$, and d is the distance between the charges. In the case we're considering, the vector sum becomes algebraic, so

$$E = k_e Q\left(\frac{1}{r_+^2} - \frac{1}{r_-^2}\right) = \frac{2k_e Qrd}{\left(r^2 - d^2/4\right)^2}.$$

It's reasonable to consider the size of the receiver d to be very small compared to the distance r ($d \ll r$). So the term $d^2/4$ can be ignored in the denominator:

$$E \cong 2k_e \frac{p_e}{r^3}, \qquad (3)$$

where $p_e = Qd$ is referred to as the electric moment. Thus the real electric fields generated by a dipole transmitter decrease very quickly with distance ($\sim 1/r^3$), which makes them unsuitable for transmitting a signal over *great* distances.

We have found the electric field along the dipole axis. In any other direction E depends also on the angular coordinates, but the $1/r^3$ dependence on distance remains the same.

Now, one might imagine a more complicated transmitter consisting of three or more charges arranged in such a way that the decrease in field strength obeys a different law. Indeed, one can do this (the corresponding fields are called *multipole*) and obtain a different function for the decrease in field strength: $1/r^n$. However, in each case $n \geq 3$, so the multipolar fields are even less suitable for wireless telegraphy than the dipole fields.

Ghost invention No. 2

The "magnetostatic" wireless telegraph could have appeared in the early 19th century. The attempt to use an electric field came to naught, but maybe a magnetic field will do the trick? The basis for such an experiment is the Biot–Savart law, discovered by the French physicists Jean Biot and Francois Savart in 1820. According to this law, an electric current I flowing in a straight conductor of length l produces a magnetic field around the wire given by

$$B = \frac{k_m Il}{r^2}, \qquad (4)$$

where r is the distance from the conductor and k_m is a proportionality constant. The observation point is assumed to lie far away ($r \gg l$) in the plane perpendicular to the conductor and through its center. For all other directions formula (4) becomes a little complicated, but the general dependence $B \sim 1/r^2$ remains true.

In the plane considered above, the magnetic field lines are circles (fig. 3). The existence of the magnetic field can be demonstrated with a magnetized needle, which comes to rest along a magnetic field line. The overall setup looks very similar to our first electrostatic project (see figure 1).

To transmit a signal $V(t)$, the current $I(t)$ must be changed in a certain way. Accordingly, the magnetic field will vary and so will the torque act-

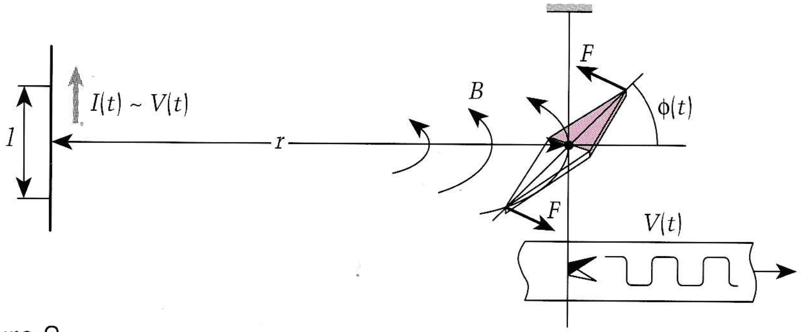


Figure 3

Simplified diagram for a "magnetostatic" wireless telegraph.

ing on a needle suspended from an elastic string. The needle's deviations are traced by a chart recorder, thereby reproducing the signal.

Everything seems to be in order, and the decrease in the magnetic field ($B \sim 1/r^2$) is much slower than that of the electric dipole field ($E \sim 1/r^3$). But alas—here we made the same mistake that forced us to substitute figure 2 for figure 1. The problem is, electric current exists in the conducting wire only if the circuit is *closed*—that is, when it connects two poles of a voltage source. So the correct diagram looks like the one in figure 4.

However, in a closed circuit any portion of it can be compared with another portion whose current flows in the opposite direction. In figure 4 these portions are labeled I and II. The fact that we've drawn our circuit in the shape of a rectangle makes no real difference. The essence of our reasoning will be the same for a circuit of any shape.

The magnetic fields produced by the opposite currents partially cancel each other, so the resulting field will be determined by a formula that is very similar to formula (3):

$$B \cong 2k_m \frac{p_m}{r^3}, \qquad (5)$$

where $p_m = IS$ is the magnetic moment of a closed loop of current encompassing an area S. In our case $S = ld$ and $p_m = Ild$. Formula (5) is valid for an observation point located on the frame's axis and sufficiently far from the frame ($r \gg l$, $r \gg d$). In any other direction B depends on the angular coordinates, but the law $B \sim 1/r^3$ remains true.

Once again we've bumped up against the disagreeable dependence $\sim 1/r^3$, which is indeed enough to bring our project to a standstill. Of course, we can "trick" formula (5) by placing the second wire at so large a distance d that its magnetic field would hardly affect the field produced by the first wire. However, to do this we would need a "transmitter" (that is, a loop of wire) of approximately the same length as the distance to the "receiver." So much for a "wireless" telegraph!

One and indivisible

In the middle of the 19th century it had been established that the alternating fields $E(t)$ and $B(t)$ form a single electromagnetic field. It is this field that made it possible to solve the problem of wireless telegraphy.

When we considered the two modes of communication above, we used formulas for electrostatics (Coulomb's law) and magnetostatics (Biot–Savart's law). It was tacitly assumed that the same relationships would be valid even during the changes in the charge or current that are required to transmit a signal. However, something was wrong in our reasoning, and this can be clearly seen if we think about how quickly a signal is sent from the "transmitter" to the "receiver" in the systems considered above.

Since the fields $E(t)$ and $B(t)$ mimic the changes in $Q(t)$ and $I(t)$ at any distance (only the amplitudes of E and B vary with distance, not the times involved), we come to the conclusion that the signals are transmitted *with infinite velocity*. However, this inference runs counter to a very important postulate of the theory of relativity, which says that no signal can travel faster than the speed of light. So something important fell out of our reasoning—namely, the fact that the electro- and magnetostatic formulas cannot be applied to varying charges and currents. (It's not for nothing that we put the words "electrostatic" and "magnetostatic" in quotes when naming our devices.)

When we're dealing with variable **E** and **B** fields, we can't consider them in isolation from each other. Their interrelations are so tight that they form a single indivisible electromagnetic field. The basic propositions of the modern theory of the *electromagnetic field* were formulated in 1860–65 by the English physicist James Clerk Maxwell, whose work grew out of the ideas and experiments of his precursor Michael Faraday. Faraday's name is linked with the phenomenon of

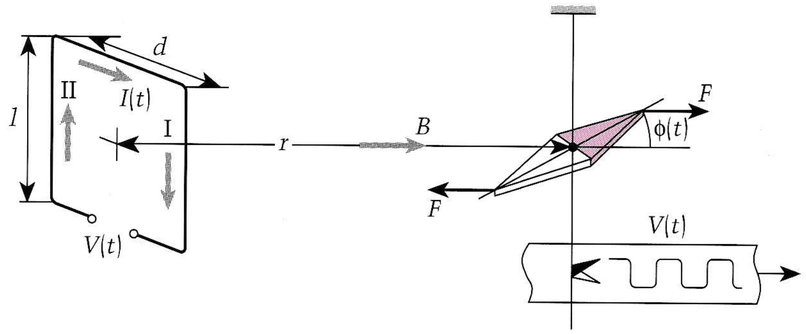


Figure 4

Actual diagram for a "magnetostatic" wireless telegraph.

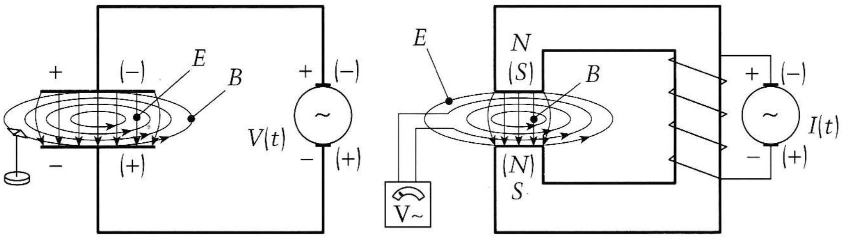


Figure 5

Experiments demonstrating Maxwell's first and second equations.

electromagnetic induction, and Maxwell's famous equations play a role in electrodynamics as fundamental as Newton's laws in classical physics.

I'll illustrate the essence of Maxwell's equations with two imaginary experiments (they can easily be reproduced in a lab). The left-hand side of figure 5 shows a capacitor connected to a source of variable voltage $V(t)$. The capacitor's charge varies constantly, as does the electric field between the plates. According to *Maxwell's first equation*, a varying electric field $\mathbf{E}(t)$ generates a magnetic field $\mathbf{B}(\mathbf{r},t)$. The two variables $\mathbf{r}$ and t show that the magnetic field varies with time and with position. The magnetic field lines are circles encompassing the electric field lines.

Notice that in the "empty" space between the capacitor plates there are no wires carrying electric current. Nevertheless, a magnetic field exists there! This means that a varying electric field is a source of magnetic field just as a flow of current is. To distinguish these two sources, they are referred to as *conduction current* (the motion of electric charges in a conductor) and *displacement current* (a varying electric field in the "empty" space of a dielectric). It was Maxwell who introduced the concept of displacement current in physics.

Now look at the right-hand side of figure 5. It shows a solenoid with an "empty" core. An alternating current flows through the winding and generates a varying magnetic field between the poles. Its field lines circle around the poles in just the same way as the electric lines do around the capacitor on the left-hand side. According to the *Maxwell's second equation*, a varying magnetic field $\mathbf{B}(t)$ induces an electric field $\mathbf{E}(\mathbf{r}, t)$, which can be detected with a coil and a voltmeter. You can see the symmetry of the properties of the **E** and **B** fields in the figure.

Our story of imaginary experiments is purely qualitative. We didn't use any calculations (they're rather complicated), but Maxwell's equations can be numerically solved to yield detailed quantitative descriptions of these experiments.

Electromagnetic waves and radio

Figure 6 shows a radio transmitter with an antenna. The transmitter has a high-frequency generator that operates under the control of the signal to be transmitted. The antenna consists of two wires connected to the transmitter's output.

Let's look at how the antenna works. The arrows in figure 6 show the currents, but there's something puzzling here—how can current exist in an open circuit? Everything becomes clear, however, when we realize that the two arms of the antenna are the "plates" of a kind of capacitor. The gradual transfiguration of a common capacitor into an antenna is shown on the left-hand side of figure 7. The alternating current supplied by the transmitter is the charge-and-discharge current of the antenna's capacitor, shown by the broken lines in figure 6. The diagram has been simplified, in that there is no single capacitor connected to particular points along the wires, but rather many "capacitors" distributed along the entire length of the antenna. In other words, the antenna has a distributed capacitance.

The charge-discharge cycle of the antenna's "capacitor" caused by the electric current $I(t)$ produces two opposite charges $\pm Q(t)$ in the arms of the antenna. In the space around antenna the current generates a varying magnetic field $\mathbf{B}(t)$, and the charges induce a varying electric field $\mathbf{E}(t)$.

But as I mentioned above, the varying fields $\mathbf{E}(t)$ and $\mathbf{B}(t)$ are intertwined and can't be considered individually. For example, look at point 1 in figure 7. A change in $\mathbf{E}(t)$ will induce the formation of a field $\mathbf{B}(t)$ not only at point 1 but also at point 2 nearby (compare this with the left-hand side of figure 5). The varying magnetic field **B** at point 2 produces an electric field at point 3 (compare this with the right-hand side of figure 5), and so on. Thus the electric and magnetic fields do not originate simultaneously at every point around the antenna, but propagate from one point to another with a finite velocity. This set of interlinked **E** and **B** fields moving off from the antenna is known as an *electromagnetic wave*, or *radio wave*.

The speed of radio waves can be calculated from Maxwell's equations. The calculations show that this speed is the same as the speed of light c = 300,000 km/s. It turns out that light waves and radio waves are basically the same phenomenon—electromagnetic waves of different frequency. According to the modern terminology, the frequencies of radio waves range from several hertz to several gigahertz (1 gHz = 10^9 Hz). However, our systems of mass communication (radio and television) only use high-fre-

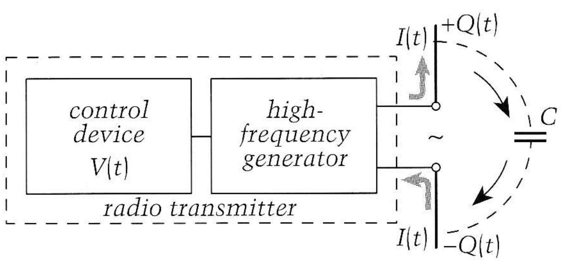


Figure 6

Radio transmitter with antenna.

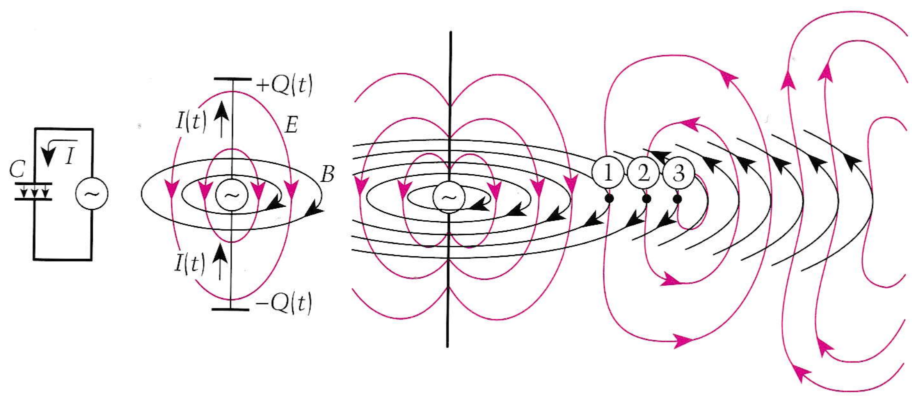


Figure 7

Transformation of a capacitor into an antenna and the generation of radio waves.

quency oscillations in the range of 10^5–10^9 Hz.

The frequency, in fact, is what sets the "electrostatic" transmitter apart from the radio transmitter. It makes no particular difference that the charges in figure 2 are concentrated at the ends of the wires, while in figure 6 they're distributed along the entire length of the wire. What's important is that in the first case the charges *vary in the same way* as the transmitted signal $V(t)$, but in the second case the charges and currents oscillate at a very high frequency—higher by far than the frequency of the transmitted signal. The transmitted signal merely regulates (modulates) the oscillations generated by the transmitter.

If we take the simple case of transmitting a telegraph signal (in Morse code), the transmitter is just turned on and off by the operator's key. When audio or video signals are transmitted, control of the transmitter is much more complicated, but the antenna radiates only high-frequency radio waves, not the low-frequency oscillations of the transmitted signal $V(t)$.

The theoretical prediction of electromagnetic waves called for an experimental proof. It was provided in 1888 by the German physicist Heinrich Hertz. Now the path to wireless telegraphy was wide open, but it still took several years to turn the idea into an apparatus. When, in the spring 1896, Alexander Popov demonstrated his radio telegraph, the first words transmitted over a distance of 250 m were "Генрих Герцъ" ("Heinrich Hertz" in Russian).

After Hertz's experiments, the invention of radio was "in the air," so to speak. Independently of Popov and at virtually the same time, a similar radio device was constructed by the Italian physicist Guglielmo Marconi, who was working in England. The range of wireless telegraphy steadily increased, and in 1901 Marconi successfully transmitted radio signals across the Atlantic Ocean. Radio communication now connected the continents.

As our story draws to a close, we need to explain why the transmission of information by electromagnetic waves proved to be more effective than the "electrostatic" and "magnetostatic" approach. Let's put a transmitter and an antenna at point O (fig. 8). At $t = 0$ the transmitter is turned on, and after a time Δt

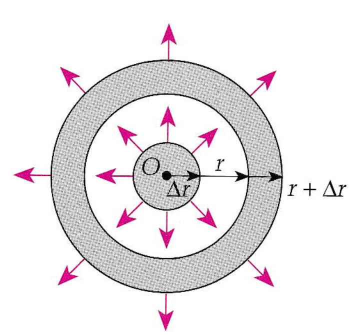


Figure 8

Volume of space containing electromagnetic energy at the times Δt (sphere) and $t + \Delta t$ (spherical layer).

it's turned off. While the transmitter is active, a radio wave spreads from the antenna to a distance $\Delta r = c\Delta t$. Assuming that the intensity of the radiation doesn't depend on its direction, we conclude that a sphere of this radius is filled homogeneously with electromagnetic energy $\Delta W = P\Delta t$, where P is the radiating power. After the transmitter is turned off, the radio wave spreads farther, and after time t the same energy fills a spherical layer limited by the radii $r = ct$ and $r + \Delta r = c(t + \Delta t)$. The volume of this layer at the moment $t \gg \Delta t$ is $\Delta V \cong 4\pi r^2 \Delta r$, so a unit volume located at a distance r from the antenna contains the energy

$$w = \frac{\Delta W}{\Delta V} = \frac{\Delta W}{4\pi r^2 c\Delta t} = \frac{P}{4\pi r^2 c}.$$

So the volumetric density decreases with distance as $1/r^2$.

Now we recall that the volumetric density of electric field energy is $w_E = \varepsilon_0 E^2/2 = E^2/8\pi k_e$ (its average value per unit time is one half as large), and that the energy of the magnetic and electric fields in the electromagnetic wave are equal. Thus

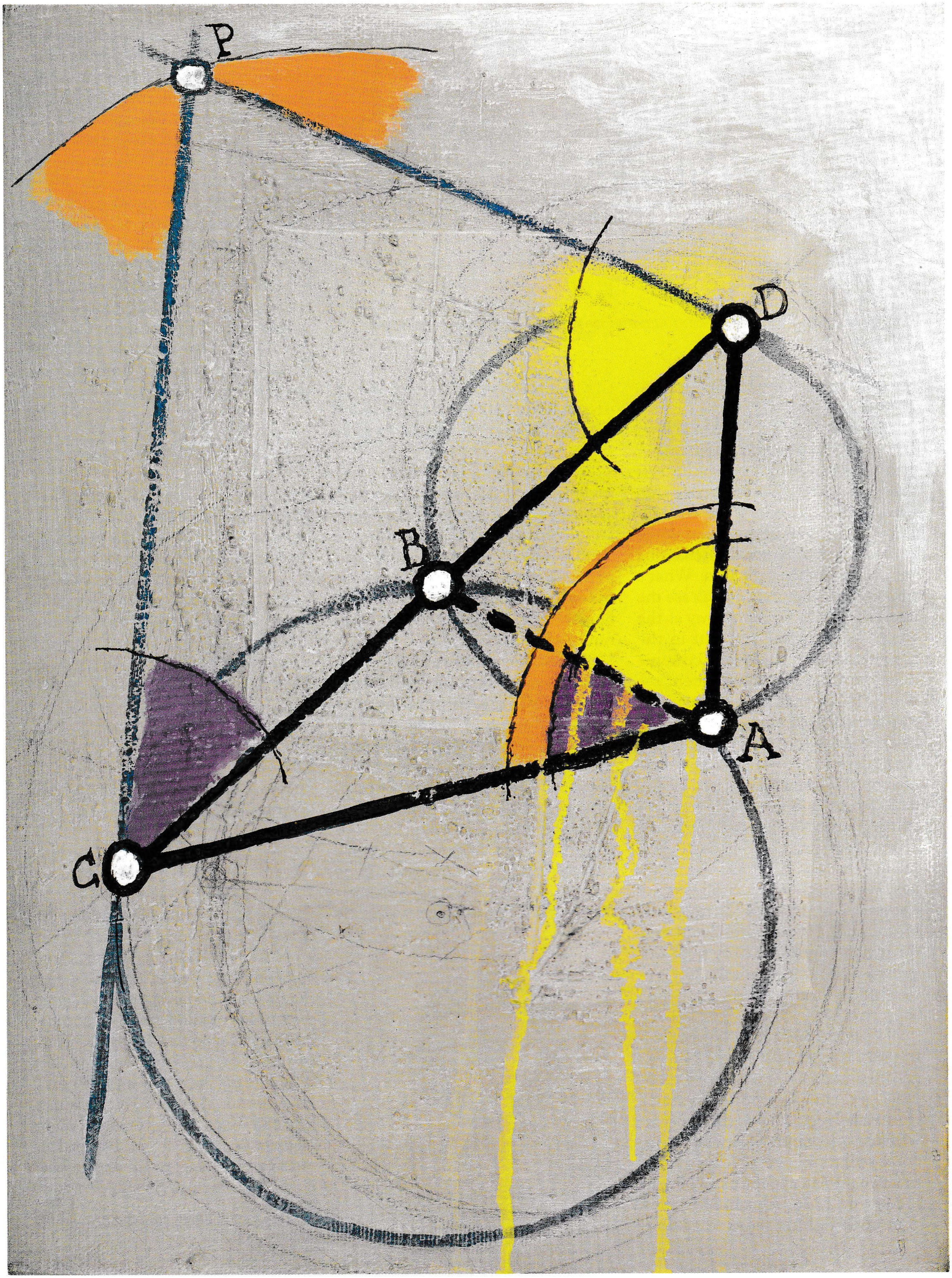
$$\frac{E^2}{8\pi k_e} = \frac{P}{4\pi r^2 c},$$

and the amplitude of the electric oscillations is

$$E = \frac{1}{r}\sqrt{\frac{2k_e P}{c}}.$$

This formula shows that the decrease in the electric field strength is proportional to $1/r$. The same is true of the magnetic field. So here's why the transmission of signals over *large distances* by radio waves is incomparably more effective than the "electrostatic" or "magnetostatic" modes (actually, it's the only way we know). Assume, for example, that at a distance of 10 m the intensities of an electrostatic field and a high-frequency electric field are equal: $E_0 = E_{\sim} = E$. Then, at a distance of, say,

CONTINUED ON PAGE 14


P
D
B
A
C

Inscribe, subtend, circumscribe

"Or draw a triangle inside a semicircle
That would have no right angle."
—*Dante,* The Divine Comedy (Heaven)

by Vladimir Uroyev and Mikhail Shabunin

THE TRIANGLE THAT DANTE writes about (fig. 1) is impossible. The author of the immortal poem gives this fact together with a number of other scientific truths that apparently were considered to be universally known to his educated contemporaries. It would be appropriate here to recall when the great Dante Alighieri lived (1265–1321) and to take note of the fact that he wasn't a mathematician. He made his mark in history as a poet and philosopher, the creator of the Italian literary language.

In this article we'll deal with triangles inscribed in semicircles as well as other configurations that are based on this well-known theorem:

THEOREM 1 (Inscribed Angle Theorem). *An inscribed angle is equal to half the central angle subtended by the same arc.*

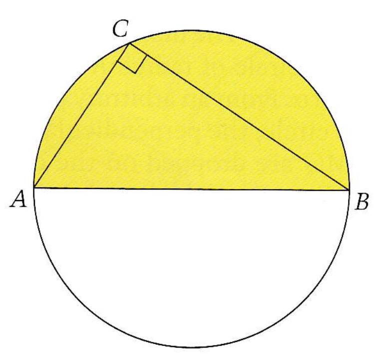


Figure 1

Art by Yury Vashchenko

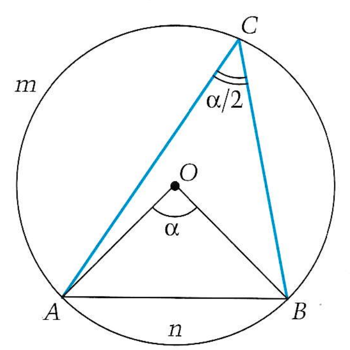


Figure 2

This theorem is often encountered in the following equivalent wording:

Let α be the central angle subtended by the arc AnB of a given circle (fig. 2). Then from any of the points on the arc AmB, the chord AB subtends the angle $\alpha/2$.

According to this theorem, an angle subtended by a diameter is always a right angle, so "a triangle inside a semicircle" is necessarily a right triangle. This fact turns out to be useful, for instance, in the following problem.

Problem 1. In a circle C_1 constructed on a segment *AB* as diameter, a chord *DB* is drawn. A circle C_2 touches *AB* at *A* and *DB* at *K*. Prove that *AK* bisects angle *DAB* (fig. 3).

Solution. Let point *O*, which lies on *AB*, be the center of C_2. The radius drawn from O to the point of tangency *K* is perpendicular to *DB* and so is parallel to *AD*, because the angle *ADB* is subtended by a diameter. It follows that angle *DAK* equals angle *AKO*. Finally, we notice that angle *AKO* equals *KAO*, because $OA = OK$, and $\angle DAK = \angle OAK$.

However, the following simple consequence of Theorem 1 is used even more often:

Inscribed angles subtended by the same or congruent arcs are congruent to each other.

As a rule, when we want to solve

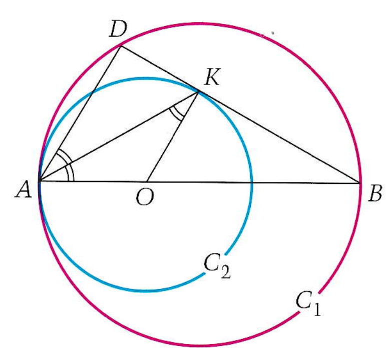


Figure 3

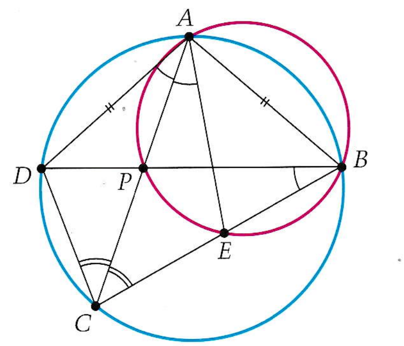


Figure 4

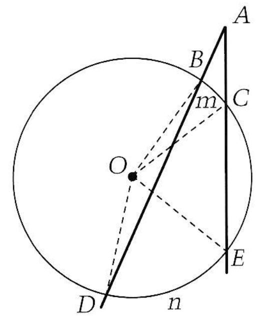


Figure 5

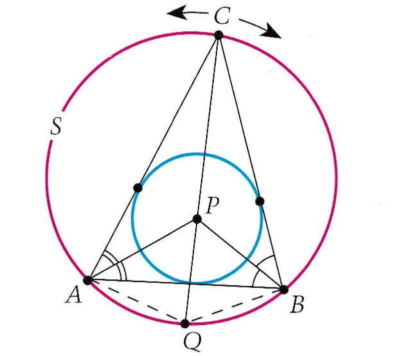


Figure 7

a problem with circles and polygons inscribed in them, we search for equal inscribed angles or try to create them by drawing various chords. The more inscribed angles we find, the clearer the path to a solution.

Problem 2. The diagonals of a quadrilateral *ABCD* inscribed in a circle intersect at *P*. Another circle, drawn through *A*, *B*, and *P*, meets *BC* at *E*. Prove that if $AB = AD$, then $CD = CE$ (fig. 4).

Solution. Draw the chord *AE* and find equal inscribed angles. In one of the circles there are the angles *DAC* and *DBC*; in the other, *PAE* and *PBE*. Therefore, $\angle DAC = \angle CAE$. In addition, the angles *DCA* and *ACB* subtended by equal arcs are also congruent. From here it follows that the triangles *DCA* and *ECA*, and so their corresponding sides *DC* and *EC*, are congruent.

Exercise 1. Prove the Intersecting Chords Theorem: *If the chords AB and CD of a given circle intersect at M, then* $AM \cdot MB = CM \cdot MD$. In other words, the product of the segments into which the point *M* divides an arbitrary chord through it does not depend on the chord. (This product is called the power of point *M* with respect to the given circle).

The Inscribed Angle Theorem can be generalized to the case of an angle whose vertex lies inside or outside the circle.

Exercise 2. (a) Suppose that an angle whose vertex *A* lies outside a given circle intercepts the arcs *BmC* and *DnE* on it (see figure 5). Prove that the measure of this angles is half the difference between the measures of the central angles subtended by these arcs—that is, $\angle DAE = {}^1/_2(\angle DOE - \angle BOC)$. (b) Formulate and prove a similar statement for the case where the vertex *A* is inside the circle.

Often it's convenient to use the Inscribed Angle Theorem in the following stronger form:

THEOREM 2. *The locus of the points from which a given segment AB subtends a certain fixed angle consists of two arcs with endpoints A and B (the points themselves excluded) symmetric about AB (fig.6).*

Proof. The fact that the segment *AB* subtends the same angle from any point of the two arcs immediately follows from theorem1. So we have to prove the converse statement: Any point *P* such that $\angle APB = \alpha$ lies on one of the arcs. But this statement follows from exercise 2: if point *P* lies outside the figure bounded by the arcs, then $\angle APB < \alpha$; if *P* lies inside this figure, we have $\angle APB > \alpha$. This completes the proof.

The most obvious class of problems where theorem 2 can be applied are construction problems.

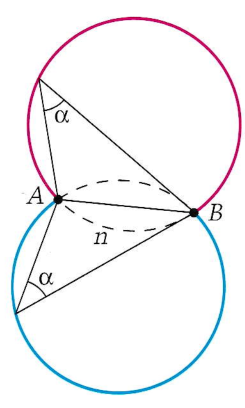


Figure 6

Exercise 3. Construct a triangle given an angle, its opposite side, and (a) the altitude, (b) the median to this side.

Problem 3. A point *C* runs along an arc *AB* of a circle *S*. Find the path of the incenter *P* of the triangle *ABC*.

Solution. The incenter lies at the intersection of the bisectors of the triangle *ABC* (fig. 7). Therefore, we have the following equations:

$$\begin{aligned}\angle APB &= 180° - ({}^1/_2\angle CAB + {}^1/_2\angle CBA)\\ &= 180° - {}^1/_2(180° - \angle ACB)\\ &= 90° + {}^1/_2\angle ACB.\end{aligned}$$

By theorem 1, the measure of $\angle ACB$ is constant, so the measure of $\angle APB$ is constant as well. Then by theorem 2, point *P* sweeps out the arc *AB* of a certain circle. We'll see later that the center *Q* of this circle is the midpoint of the arc *AB* of circle *S*.

Theorem 2 can be used to determine whether a certain set of points belongs to a circle. In particular, it implies the following important corollary:

The vertex of the right angle of a right triangle lies on the circle constructed on the hypotenuse as diameter.

Exercise 4. The diameters *AB* and *CD* of a circle of radius *R* make an angle of α. From an arbitrary point *M* of the circle, the perpendiculars *MP* and *MQ* are dropped on the diameters. Prove that the length *PQ* does not depend on *M* and express it in terms of *R* and α.

Exercise 5. The vertices *A* and *B* of the acute angles of a fixed right triangle slide along two perpendicular lines. Find the trajectory of the triangle's third vertex C.

When you do the next two exercises, take into account the fact that the bases D and E of the altitudes DB and CE of a triangle ABC lie on a circle with diameter BC.

Exercise 6. From the vertices B and C of a triangle ABC the altitudes DB and CE are dropped and the perpendiculars DE and CG to the line DE are drawn. Prove that $EF = DG$.

Exercise 7. Prove that the altitudes of a triangle ABC are the angle bisectors of the triangle DEK formed by the feet of the altitudes of $\triangle ABC$.

Problem 4. Prove that the four circles constructed on the sides of a quadrilateral as diameters cover the entire quadrilateral.

Solution. If the quadrilateral contains a point outside all four circles, then each of its sides subtends an acute angle at this point. But this is impossible, because the sum of these angles is 360° (for a convex quadrilateral) or more (if the quadrilateral is not convex).

Theorems 1 and 2 immediately lead to the following criterion for inscribed quadrilaterals:

A quadrilateral can be inscribed in a circle if and only if the sum of its opposite angles equals 180°.

Problem 5. Three circles pass through a point P and intersect in pairs at points A, B, C. The lines drawn from an arbitrary point M of the circle PBC through B and C meet the other two circles at S and T (fig. 8). Prove that the line ST passes through A.

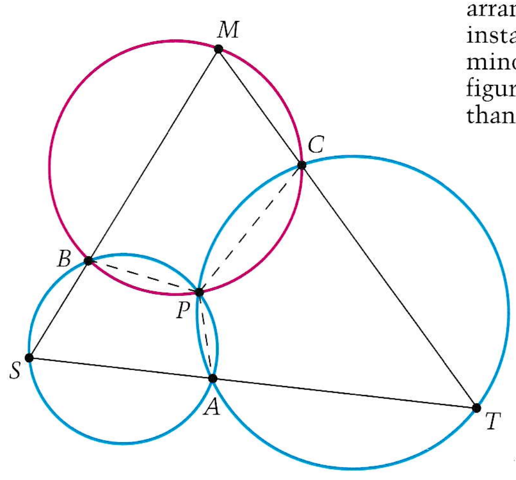


Figure 8

Solution. From the inscribed quadrilaterals $MBPC$, $BPAS$, $CPAT$, we derive the equations

$$\angle BPC = 180° - \angle M,$$
$$\angle BPA = 180° - \angle S,$$
$$\angle CPA = 180° - \angle T.$$

Adding them up, we get

$$360° = \angle BPC + \angle BPA + \angle CPA = 180° \cdot 3 - (\angle M + \angle S + \angle T),$$

or

$$\angle M + \angle S + \angle T = 180°.$$

This means that the quadrilateral $MSAT$ is in fact a triangle—that is, $\angle SAT = 180°$.[1]

In the next problem the criterion for inscribed quadrilaterals must be supplemented by the following extension of the Inscribed Angle Theorem:

THEOREM 3. *The angle between a chord and a tangent at its endpoint equals half the angular measure of the arc subtended by the chord.*

Exercise 8. Prove this theorem.

Problem 6. Two circles intersect at points A and B. A line through B meets the circles at C and D (fig. 9). Prove that the points A, C, D, and the intersection point P of the tangents to the circles at C and D are concyclic (lie on the same circle).

Solution. Draw the chord AB and notice that, by theorems 1 and 3, $\angle BAD = \angle BDP$, $\angle CAB = \angle BCP$. It follows that $\angle CAD = \angle BAD + \angle CAB = \angle BDP + \angle BCP = 180° - \angle CPD$. So

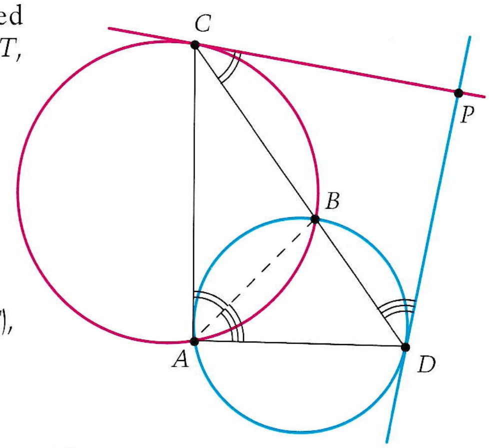


Figure 9

the quadrilateral $ACPD$ can be inscribed in a circle by the criterion given above.

Exercise 9. Lines l_1 and l_2 touch a circle at points A and B. The distances from a point M on the circle to the lines equal a and b. Prove that the distance from M to AB is $\sqrt{ab}$.

Exercise 10. Two circles touch each other externally at D. A line touches one of them at A and intersects the other one at B and C. Prove that A is equidistant from BD and CD.

Exercise 11. From a point A outside a circle a tangent AP and a secant meeting the circle at B and C are drawn. Prove that $AP^2 = AB \cdot AC$. (Compare with exercise 1.)

Exercise 12. Prove that the center of the circle around which the point P in problem 3 moves is the midpoint Q of the arc AB (fig. 7).

Problem 7. Let the center O of a circle S_1 lie on a circle S_2 and let A and B be the intersection points of the two circles. Draw a chord AC in S_1 meeting S_2 at D (fig. 10). Prove that OD is perpendicular to BC.

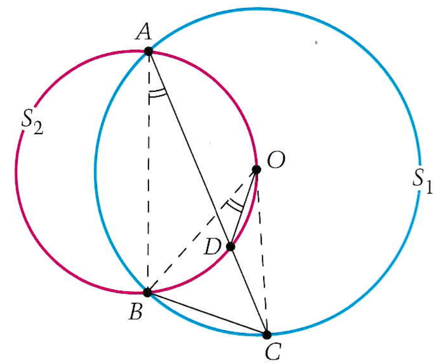


Figure 10

[1] Actually, this solution relies rather heavily on the specific arrangement shown in figure 8. For instance, if M were taken on the minor arc BP of the circle BPC in that figure, we'd get $\angle BPA = \angle BSA$ rather than $180° - \angle BSA$, and so on. However, after we make all the necessary corrections, things will fall into place and we'll have either $\angle SAT = 180°$ or $\angle SAT = 0°$.

The complete solution requires an exhaustive search through all possible cases, but this tedious work can be avoided by using the notion of oriented angles, which reduces the number of cases to one. The careful reader may have noticed a similar situation in problem 2—*Ed.*

Solution. The angles BAD and BOD are inscribed in S_2 and subtended by the same arc, so they are equal; angle BAC is inscribed in S_1, so $\angle BAC = {}^1/_2 \angle BOC$. Consequently, $\angle BOD = \angle BAD = {}^1/_2 \angle BOC$—that is, OD is the bisector of the angle BOC. It remains to notice that the bisector OD of the isosceles triangle BOC coincides with its altitude.

Problem 8. In a trapezoid $ABCD$ ($AD \parallel BC$), the angle ADB is half the angle ACB; $BC = AC = 5$; $AD = 6$. Find the area of the trapezoid. (Hint: the problem can be solved in a standard way, using, say, the Law of Cosines. But a much shorter solution emerges if you notice that by theorem 2 point D lies on the circle through A and B centered at C. The answer is 22.)

"THE ADVENT OF RADIO" CONTINUED FROM PAGE 9

100 km, the strength of the electrostatic field decreases to $10^{-12}E$, while that of the electromagnetic field decreases only to $10^{-4}E$. In other words, the electromagnetic field is stronger than the electrostatic field by a factor of 10^8 at 100 km! At greater distances (100 km is a trifle for radio communication), the predominance of $E_{\sim}$ over E_0 is even greater.

There is another advantage of radio waves that we didn't discuss here. These waves can propagate in any medium as long as its electric conductivity isn't too high. True, in this case $E_{\sim}$ is subjected to some additional attenuation, but a constant field E_0 can't pass through such a medium at all.

"PUBLISHER'S PAGE" CONTINUED FROM PAGE 3

interval of 14 days."

Perhaps our readers can provide us with a calculation showing why a lunar eclipse can last 3 hours 40 minutes, with totality lasting up to 1 hour 40 minutes.

Get out your "cowculators"!

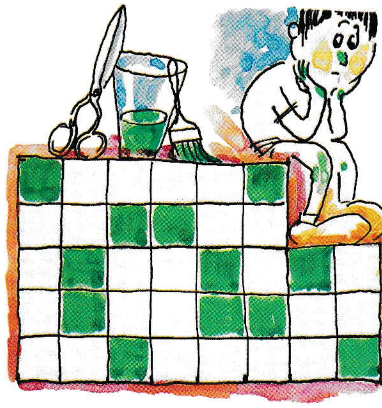
No, I'll not revisit the subject of my previous Publisher's Page, "Solar Calculator." I just wanted to alert our readers to a new feature that makes its debut in this issue of *Quantum*. It's a column called "Cowculations," and its author is the esteemed Dr. Mu of the University of Wisconsin–Parkside. This addition to our magazine comes as a response to requests from readers for a column devoted to computer science and programming. We hope you enjoy the algorithmic challenges Dr. Mu offers, beginning with the one on page 37.

Thanks for your company

This will be the last Publisher's Page I'll write, although I may return to this spot occasionally with a guest editorial. Dr. Gerald F. Wheeler, the executive director of NSTA and an outstanding teaching physicist, will become the publisher of *Quantum*. He will write the next column, and I look forward to reading what he has to say.

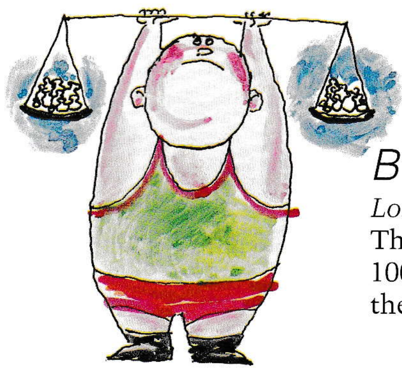
—Bill G. Aldridge

Just for the fun of it!



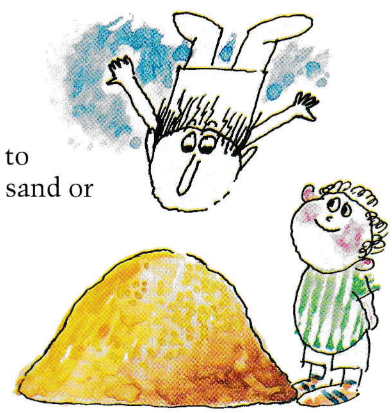
B186

Cutting a thick El. Cut the L-shape in the figure into two pieces and make a square out of them so that the colored squares form a pattern that is symmetric in the same lines as the large square itself.



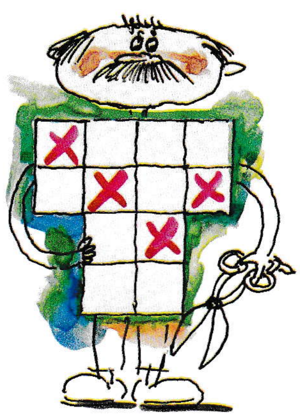
B187

Lost weight. A set of weights have masses 1 g, 2 g, 3 g, ..., 101 g. The 19-g weight was lost. Is it possible to divide the remaining 100 weights into two groups with the same number of pieces and the same total weight? (V. Proizvolov)



B188

Physics of safety. It is safer to jump from a precipice onto sand or onto firm ground. Why?

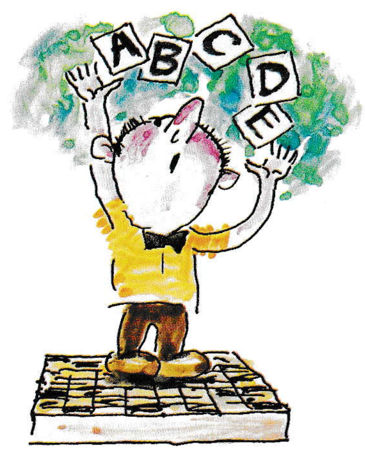


B189

Cutting a thick Tee. Cut the T-shape in the figure into four equal pieces along the gridlines so that each piece contains exactly one marked square.

B190

Logic of chess battles. After the end of a one-round chess tournament in which every two participants played each other once, the players *A*, *B*, *C*, *D*, and *E*, listed here in the order they finished, shared their impressions. "I would never have imagined that I'd be the only one who'd finish without a single loss," player *B* said. "And I'm the only one who didn't win a single game," *E* sighed. Based on this information, try to restore the entire table of the tournament: how each of the five participants played against all the others. (In a chess tournament, the winner of each game gets one point, and if the game is a stalemate (a tie) each player gets half a point. In this particualr tournament, no two players got the same number of points.) (S. Guba)



ANSWERS, HINTS & SOLUTIONS ON PAGE 53

Art by Pavel Chernusky

A flight to the Sun

An article from April 1986 whose speculations would soon become reality

by Alexey Byalko

THERE IS A SAYING: "THE most interesting surface is the human face." But is there a surface in Nature that rivals the human face in its pithiness and unpredictability? Yes, indeed—the surface of the Sun.

Look at the images of the Sun in figure 1. Unfortunately they can't give a complete picture of the beauty and complexity of the Sun. Its surface is very heterogeneous—no two places on it are alike—and all this variety is constantly changing. In addition, the appearance of the Sun's surface varies depending on the wavelengths of radiation being recorded. This occurs because solar radiation of various wavelengths is produced at different altitudes in the Sun's "atmosphere."

The variety of phenomena at play on the Sun's surface—solar flares, solar prominences, the spasmodic appearance, movement, and disappearance of sunspots, and so on—all reflect the complicated processes going on in the depths of our daystar. Scientists understand what's happening in the Sun on the atomic scale. For example, it's known how light affects the individual atoms and how the atoms interact with one another (atomic hydrogen and helium account for 99.9% of the Sun's mass). However, there is a huge gap between atomic distances (about 10 m^{-10}) and the size of objects discernible on the Sun with a modern telescope—that is, of the order of 10^6 m. Scientists devise hypotheses, trying to explain theoretically what goes on within this interval of 16 orders of magnitude. There are hypotheses dealing with the hydrodynamics of the Sun's interior, the mechanism of energy transfer from the interior to the surface, and the structure of its magnetic fields. But we still have a long way to go in reaching a real understanding of solar physics. At present we still lack a complete explanation of solar flares, sunspots, and many other phenomena on the surface of the Sun.

> IF IT WERE GRANTED unto mortals
> To climb an unimagined height,
> There, penetrating heavenly portals,
> They would perceive a dazzling sight:
> A burning and eternal ocean
> Subject to endless inner motion.
>
> There fiery ramparts are extended
> Further than human eyes explore,
> There whirlwinds flame, have ne'er suspended
> A love of fated, mutual war.
> There stones, grown liquid, boil and seethe,
> And burning rains far space enwreathe.
>
> —M. V. Lomonosov

A question naturally arises: what keeps us from getting a closer look at the Sun? And not only a look, but a chance to study it with all the means at our disposal?

Maybe the high temperature prevents such a project? Electronic devices are designed to work at ordinary terrestrial temperatures (around 300 K), but an apparatus passing near the Sun would endure far more intense heat. However, this technical problem can be solved—one can construct a refrigerated compartment in a space vehicle flying to the Sun and keep it at room temperature.

Of course, it might be harmful for instruments in a space probe to pass directly through the Sun's corona and the upper strata of the solar atmosphere not only because of overheating but for other reasons as well. For example, powerful streams of charged particles can cause radiation damage in electronic devices. Such streams of electrons and ions are emitted during solar flares. Paradoxically, a flight near the Sun at a minimum distance of 4–5 solar radii is relatively safe, because a space probe passing near the Sun at a distance less than 10 solar radii will not remain there for longer than a few hours. The time is so small because the spacecraft's speed is so large. At a distance of four solar radii from the Sun's center, the probe's speed will reach 300 km/s. This value shows how powerfully the Sun's gravitational field accelerates a body.

The primary reason we have never sent a probe to the Sun is completely unexpected at first glance. It turns out to be very difficult to place a spacecraft into a trajectory that passes near the Sun. This seems rather strange: the Sun is the major source of attraction in the solar system, so one might think it would draw any massive object to itself all on its own. However, the planets revolve around the Sun and do not fall into it. Such a fall is prevented by the velocity of each planet, which is directed perpendicular to the line connecting the planet and the Sun. And that's the point: to come near the Sun, one must compensate for the initial velocity, which for a rocket launched from the Earth is equal to the velocity of the Earth's revolution around the Sun.

This velocity is known to be about 30 km/s. If we could stop the rocket in its tracks by giving it a velocity of 30 km/s opposite the Earth's velocity, then the fall into the Sun would be inevitable. However, 30 km/s is a very large velocity. (The orbital speed for satellites near the Earth's surface is about 7.9 km/s.) Up to now no rocket has been accelerated to such a speed. Certainly it's possible in principle: one can construct a multistage rocket, but then the payload—that is, the mass of the probe—would be too small. Nevertheless, let's calculate the time necessary for such a flight from the Earth to the Sun (we'll need it below).

The trajectory of a fall to the Sun (a straight line segment) is the limiting case of an elongated elliptical orbit whose semimajor axis is equal to half the radius of the Earth's orbit. According to Kepler's third law, the squares of the periods of revolution of bodies are related as the cubes of the semimajor axes of their orbits. So the time t of a fall to the Sun (that is, half the period of revolution along an elongated orbit with semimajor axis $R_E/2$) can be found from the equation

$$\frac{(2t)^2}{T_E^2} = \frac{(R_E/2)^3}{R_E^3},$$

where T_E and R_E are the period of revolution and the radius of the Earth's orbit, respectively. Since $T_E = 1$ year, $t = 1/4\sqrt{2}$ year$\cong 0.177$ year.

The period of a flight straight to the Sun turns out to be not very long, but the necessary initial veloc-

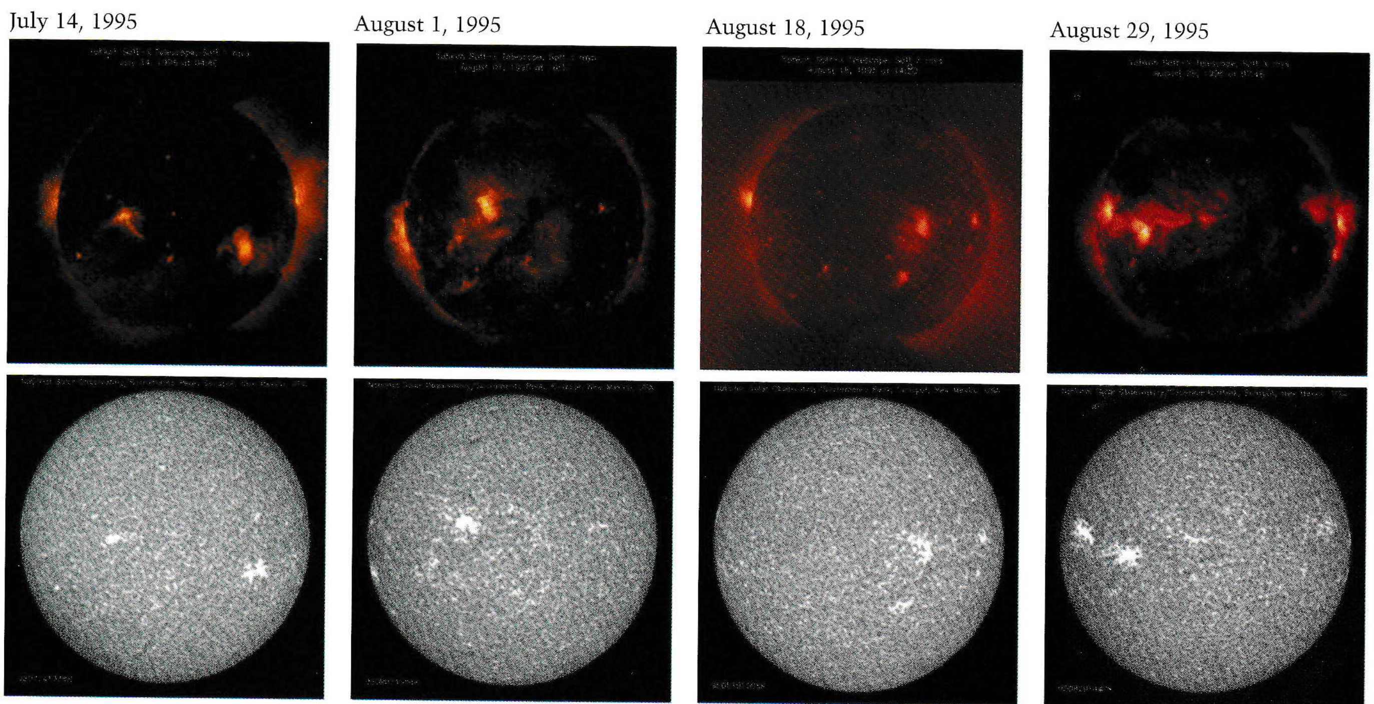


Figure 1

The changing face of the Sun. These images were taken from two sources, the joint ISAS/NASA spacecraft Yokoh (top) and the National Solar Observatory at Sacramento Peak in Sunspot, New Mexico (bottom). The images fromYokoh are all in soft x-rays, and the ones from NSO are the K emission line from Ca II. The images in each pair were taken on the same day, though not necessarily at the same time. (From the National Space Science Data Center's Photo Gallery on the World Wide Web: nssdc.gsfc.nasa.gov/photo_gallery/photogallery-solar.html) ity for such a flight is still unattainable. How, then, can we set up a flight to the Sun with minimal energy expenditure—that is, with a minimal launch velocity? The most economical flight uses Jupiter's gravitational field to decelerate the spacecraft in its orbit around the Sun. To carry out such a maneuver, the probe must approach Jupiter with a specific velocity. And to ensure this, the probe must overcome the Earth's attraction and be placed in orbit with an initial velocity relative to the Sun of v_0 = 40.5 km/s. Consequently, relative to the Earth, the probe's velocity must be equal to 10.5 km/s. Therefore, the flight begins with an acceleration of the rocket in the direction of the Earth's orbital velocity. What velocity must be imparted to the rocket?

To overcome the Earth's attraction, a body must be given kinetic energy of no less than gR_E. This kinetic energy corresponds to an escape velocity $v_e = \sqrt{2gR_E} \cong 11.2$ km/s. Pay attention, now! Don't add this number to the previously calculated value of 10.5 km/s to obtain the initial velocity! It's the law of conservation of energy that gives the initial starting velocity for a rocket launched from the Earth:

$$v_{st} = \sqrt{(v_0 - v_E)^2 + 2gR_E} \cong 15.5 \text{ km/s}.$$

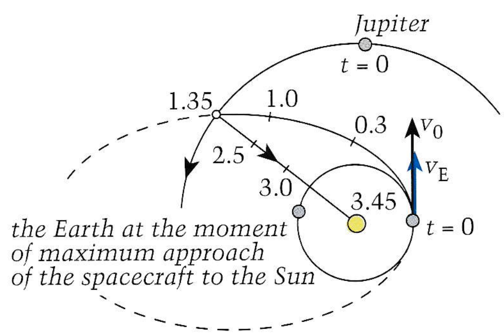


Figure 2

Trajectory of a flight to the Sun with minimal starting velocity. The numbers along the trajectory show the flight time in years. When the spacecraft crosses Jupiter's orbit after a 1.35-year flight, it approaches Jupiter and the planet reduces the probe's velocity to zero relative to the Sun. At this point the probe begins to free fall to the Sun.

By the way, do you know where the best launch location is on Earth (taking into account only the energy requirements of the launch itself)? It's Mt. Kilimanjaro in Tanzania, the highest mountain in Africa (alt. 5,900 m), which is located almost at the equator. Due to the Earth's flattened shape, the top of Mt. Kilimanjaro is the point on Earth that is most remote from the Earth's center (farther than Everest!). Here the acceleration due to gravity is the smallest on Earth. The mountaintop is always covered with snow, however, and as yet no one is planning to build a launch facility there.

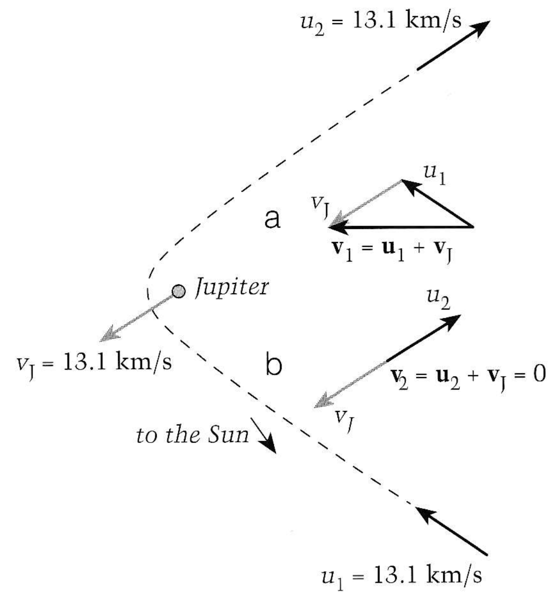


Figure 3

Trajectory of the spacecraft near Jupiter in the planet's reference system. A diagram of the velocities (a) before approach and (b) after approach is given on the right.

Let's continue our flight. The spacecraft passes along part of an elliptical orbit, crossing the orbit of Mars and approaching that of Jupiter (fig. 2), the largest planet in the solar system and the probe's intermediate goal. Jupiter will decelerate the spacecraft relative to the Sun, and then the probe begins its fall toward the Sun. For Jupiter to be at the intersection of its orbit and that of the probe, at the time of launch Jupiter must be located at the point marked with zero time in figure 2. We also have to make sure that the probe is far enough from Mars when it crosses that planet's orbit so that its trajectory isn't affected by the gravitational field of the red planet.

After the spacecraft enters the gravitational field of Jupiter, it continues to travel along a hyperbolic trajectory around this planet (fig. 3). As a result of the law of conservation of energy, the velocities of the flight to and from Jupiter at symmetrical points along the hyperbola are identical in magnitude, but their directions might differ drastically. So the probe must be placed in a trajectory around Jupiter such that its speed (relative to the planet) u_1 is equal to Jupiter's orbital speed (relative to the Sun) v_J = 13.1 km/s. The direction of this fly-by must be such that, as it moves away from Jupiter, the spacecraft's velocity at the point symmetrical to the starting point of the fly-by is directed counter to Jupiter's orbital velocity. At this point the spacecraft's velocity relative to the Sun is approximately zero and, under the influence of the Sun's attraction, however small it is at this distance, the spacecraft will begin its slow approach to the Sun. These conditions are met when the station's speed relative to the Sun at the initial moment of the flight around Jupiter is equal to v_1 = 14.3 km/s (see figure 3). This value in turn determines the starting speed of the spacecraft relative to the Sun: 40.5 km/s.

Thus, in the reference frame of Jupiter, the probe's trajectory (near the planet) is a hyperbola. In the Sun's reference frame, this part of the trajectory looks more complicated (fig. 4 on page 20).

Art by Ula Likhtman

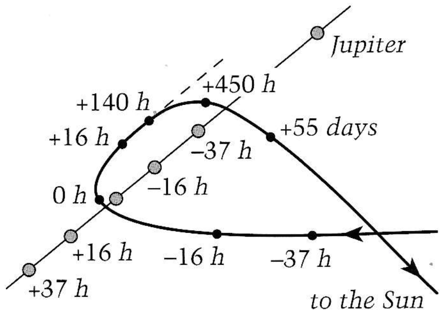


Figure 4

Trajectory of the spacecraft during its approach to Jupiter in the Sun's reference system. The numbers along the curve show the times relative to the moment of the probe's closest approach.

What is the duration of such a flight to the Sun? Earlier we estimated the duration of the fall to the Sun along a straight line. In the present case, when free fall begins from Jupiter's orbit, we must use the period of Jupiter's revolution around the Sun, which is $T_J = 11.86$ years, rather than an Earth year (although we continue to measure the time in Earth years). This yields a free-fall time of $T_J/4\sqrt{2} \cong 2.1$ years. This value is added to the time it takes to travel along the elliptical part of the trajectory (before the "meeting" with Jupiter), which is equal to 1.35 years. The total time necessary for the flight to the Sun is almost 3.5 years.

Looking at the trajectory of such a flight (fig. 4), we might think that it is Jupiter that sends the probe to the Sun. In reality, it is people who direct the spacecraft. As the probe approaches Jupiter, it is necessary to make a correction to the trajectory—even a small error in the choice of the altitude of the fly-by or the speed of approach to the planet could ruin the entire solar research effort. One other correction must be made after the fly-by to eliminate residual deviations and to precisely assign the distance at which the probe will approach the Sun.

One of the key concepts in this method—using the attraction by an intermediate planet to change the velocity—has already been used in actual space flights. One example is the flight to Venus by the Vega-1 and Vega-2 spacecraft, which, after flying past the morning star, headed off toward Halley's comet.

We need to understand the physical phenomena occurring in the Sun. We cannot predict all the consequences of the knowledge we will acquire. But by way of comparison, recall the heroic era of the great geographical discoveries, when we became familiar with our own planet. It's quite possible that our times will someday be called the era of the great space discoveries. The first flight to the Sun will surely be one of its shining achievements.

Addendum

While Alexey Byalko was writing this article, plans were underway at the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) to launch a probe to the Sun, which came to be called Ulysses. Launched by the Space Shuttle *Discovery* in October 1990, Ulysses indeed flew by Jupiter in February 1992, as the author prescribed above. The spacecraft entered a polar orbit of the Sun, passing over the south pole in 1994 and the north pole in 1995.

As described by NASA, the Ulysses mission

> explored for the first time the high latitude heliosphere away from the plane of the ecliptic. The primary results of the mission have been to discover at these high latitudes the properties of the solar corona, the solar wind, the heliospheric magnetic field, solar energetic particles, galactic cosmic rays, solar radio bursts and plasma waves. Other investigations include study of cosmic dust, gamma ray bursts, and studies of the Jovian magnetosphere obtained during the Jupiter fly-by.

Ulysses has completed the first phase of its mission and has now embarked on a second orbit of the Sun. A wealth of information about the project is available on the World Wide Web. Good places to start are the NASA Ulysses home page (http://ulysses.jpl.nasa.gov) and the ESA Ulysses home page (http://helio.estec.esa.nl/ulysses). ◘

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M186

From corner to corner. Nine checkers are placed in the lower left corner of an ordinary 8×8 chessboard to form a 3×3 square. A checker a can jump over any other checker b onto a square symmetric to a about b if this square is free. Is it possible to jump the entire 3×3 square of checkers into the (a) upper left, (b) upper right corner of the chessboard following this rule? (Y. Briskin)

M187

Circumscribed cutoffs. Prove that of the n quadrilaterals cut from a convex n-gon by its diagonals (subtending triples of consecutive sides) no more than $n/2$ can have inscribed circles. Give an example of an octagon that has four such quadrilaterals. (N. Sedrakyan)

M188

Battleship strategy. The game of Battleships is played on a 7×7 "ocean." What is the smallest number of shots that must be made to hit (at least once) a four-square ship if (a) it is a 1×4 rectangle of unit squares, (b) it is a "tetramino" piece of unknown shape (that is, it consists of four squares connected across their sides)? (A. Kholodov)

M189

Fibonacci function. Prove that the Fibonacci series 1, 1, 2, 3, 5, 8, 13, ..., each of whose numbers is the sum of the two preceding numbers, contains no less than four and no more than five m-digit numbers for each $m \geq 2$.

M190

Distribution of arc lengths. Twenty-one points are marked on a circle. Prove that at least 100 arcs with these points as endpoints have degree-measures not exceeding 120°. (A. Sidorenko)

Physics

P186

String on a ball. One end of a string of length l is attached to the top of a sphere of radius R (fig. 1). At a certain moment the string is released. Find its acceleration at this moment. (Ignore the effect of friction.) (A. Bytsko)

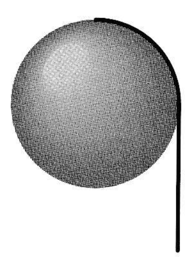


Figure 1

P187

Falling charge. A point mass M carrying a charge Q is placed at a distance L from an infinite conducting plane and then released. How long will it take the particle to reach the plane? The force of gravity is absent. (Hint: use the method of mirror images.) (A. Bytsko)

P188

Cooking under pressure. A small amount of water is poured into a pressure cooker, which is then closed tightly and set on a hot plate. By the time all the water has evaporated, the temperature of the pot is 115°C and the pressure inside is 3 atm. What portion of the volume of the pressure cooker was occupied by water before it was heated? The initial temperature was 20°C. (A. Sheronov)

P189

Sunlit plate. One side of a thin metal plate is illuminated by the Sun. When the air temperature is T_0, the temperature of the illuminated side is T_1, while that of the opposite side is T_2. What would be the temperatures of the both sides if this plate is replaced by another that is twice as thick? (E. Ponomarev)

P190

Fresnel prism. When light strikes a Fresnel prism at a right angle (fig. 2), it's split into two beams that are refracted by each half of the prism and then interfere with each other. At what maximum distance from the prism will the interference pattern will still be observed? The distance between the apexes of the prism S = 4 cm, the refractive index of the glass n = 1.4, and the prism angle α = 0.001 rad. (V. Deryabkin)

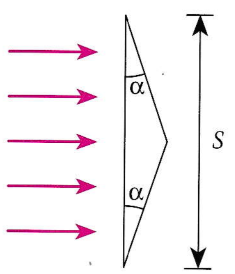


Figure 2

ANSWERS, HINTS & SOLUTIONS ON PAGE 50


3t + 1 + y + 2t = 333;
y = 332 − 5t.
y
x
0

Ax by sea

Actually, ax + by = c—a problem genre that's 1,600 years old

by Boris Kordemsky

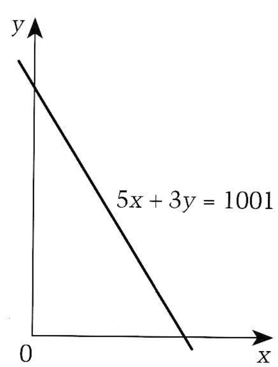
THE PLOT AND THE NOTATION of the problem kept changing, but its essence remained the same: find integers (usually positive integers) x and y that satisfy the equation

$$ax + by = c \qquad (1)$$

with given integer coefficients a, b, and c.

Perhaps c is the Thousand and One Tales of the Arabian Nights, and we want to know how many nights Scheherazade will be able to spin them if she tells five stories a night for x nights and three stories a night for the other y nights. Clearly, her stock of stories will suffice for $x + y$ nights, where x and y are the positive integer roots of the equation $5x + 3y = 1{,}001$.

Or maybe c is 10 rubles and 1 kopeck (1/100 of a ruble) spent by someone for x bus trips (at 5 kopecks per trip) and y streetcar trips (at 3 kopecks per a trip).[1] Then the answer to the question of how many



trips were made is given by the same equation $5x + 3y = 1{,}001$.

The solutions to this equation also tell us something about the points on the line $5x + 3y = 1{,}001$ (see the figure above) whose roots are both positive integers (or just integers, perhaps, in some other problem).

Equations in integers are often called *Diophantine equations* after Diophantus of Alexandria, the famous Greek mathematician of the 2nd and 3rd centuries A.D.

The conditions for the existence of integer solutions to equations of the form $ax + by = c$ can be found in the article "Divisive Devices" by V. N. Vaguten in the September/October 1991 issue of *Quantum* (see also "Go 'Mod' with Your Equations" in the January/February 1992 issue). We won't dwell on the theory here. Rather, we'll simply take up the particular equation $5x + 3y = 1{,}001$ to demonstrate various techniques for solving equations in integers.[2]

The method of the "ingenious student"

Divide both sides of the equation $5x + 3y = 1{,}001$ by the smaller coefficient:

$$\frac{5}{3}x + y = \frac{1001}{3};$$

extract the integer parts on the right and on the left:

$$x + \frac{2}{3}x + y = 333 + \frac{2}{3},$$

or

$$x + y + \frac{2(x-1)}{3} = 333. \qquad (2)$$

Since x and y are integers, the number $(x - 1)/3$ must be an integer as well. Denote it by t. Then $x = 3t + 1$. Plugging this into equation (2), we

[1]This certainly doesn't matter for the mathematics, but you may be interested to know that due to inflation and the increase in prices, a bus trip in Moscow is currently 30,000 times as expensive, and for the street car you have to pay 50,000 times more than when this article was written.—*Ed.*

[2]Recall only that the existence of a solution to equation (1) is ensured by the condition GCD$(a, b) = 1$ (where GCD is the greatest common divisor), so the equation $5x + 3y = 1{,}001$ does have an integer solution.

Art by Vasily Vlasov

get $3t + 1 + y + 2t = 333$, and so $y = 332 - 5t$.

In "Divisive Devices," it is proved that these expressions for x and y form the *general solution* to our equation—that is, they run through all its integer solutions as t ranges over the set of all integers.

Assigning the values 0, 1, 2, ..., 66 to the parameter t, we'll find 67 pairs of possible positive integer roots of the equation.

Now suppose additionally that Scheherazade wants to spread her thousand and one tales over as many nights as possible.[3] That is, we have to find $\max(x + y)$—the maximum sum of the pairs of roots of our equation.

Since $x + y = 333 - 2t$, $\max(x + y)$ is attained for $t = 0$. So Scheherazade can spread the tales over at most 333 nights (by telling three stories a night for 332 nights and five stories only once). She can shorten her "work" to 201 nights (which is hardly in her best interests) by relating five stories a night for 199 nights and two stories a night only twice. This solution results if we take the largest possible t—that is, $t = 66$.

An additional question to mull over

Suppose that, solving a certain equation in integers by the "ingenious student" method, you arrive at the equation

$$x + y + \frac{4y - 1}{3} = 77.$$

What should your line of reasoning be, and what should you do to properly express x and then y in terms of an integer parameter t?

The method of the "ingenious mathematician"

For the general equation $ax + by = c$ (with, say, $a > b$), the procedure[4] is as follows: find the remainder m of a when divided by b and the remainder n of c when divided by b. If $n = 0$, then we immediately get

$$x = bt,$$
$$y = \frac{c}{b} - at,$$

where $t = 0, \pm 1, \pm 2, \ldots$. If $n \neq 0$, we successively multiply m by 1, 2, ..., $b - 1$ and write out the sequence of the remainders of these products when divided by b. This sequence will contain the number n (otherwise, the equation has no integer solutions). Then one of possible values of x is simply the number of the position of n in the sequence.

For example, let's apply this method to our equation $5x + 3y = 1{,}001$. We have $m = 2$, $n = 2$. Multiply $m = 2$ by each of the numbers $1, 2 = b - 1$ to obtain 2, 4. Divide these by 3 and write out the remainders: 2, 1. The number n is in the first position here, so we can take $x = 1$. This determines the corresponding value of y: $y = 332$.

And there we have it: an easy and completely general method of solving linear indeterminate equations in two integer variables!

But the "ingenious mathematician" method, unlike the previous method, gave us only one pair of roots: (1, 332). Is this a defect of the method? Not at all! Consider the general solution of the problem obtained by the "ingenuous pupil" method: $x = 1 + 3t$, $y = 332 - 5t$.

The *partial solutions* $x_0 = 1$, $y_0 = 332$ are the same in both cases, while the coefficients at t (3 and -5) are determined by the coefficients of the equation in question: $3 = b$, $-5 = -a$. And this isn't just a matter of chance. On the contrary, it's a general rule: if (x_0, y_0) is a solution to the equation $ax + by = c$, where a and b are relatively prime numbers, then all of its integer solutions are given by the formulas $x = x_0 + bt$, $y = y_0 - at$, where $t = 0, \pm 1, \pm 2, \ldots$.

The method of congruence

This is a pleasant way, and often the quickest way, to find integer solutions to the equation $ax + by = c$.

It will suffice to recall that the notation $a \equiv b \pmod{m}$ (read as "a is congruent to b modulo m") means that $a - b$ is divisible by m; or, equivalently, that $a = b + km$ (m is a positive integer, and a and b are arbitrary integers).

Further, if $a \equiv b \pmod{m}$, then $a \equiv b + km \pmod{m}$ for any integer k. For instance, let $3x \equiv 2 \pmod{5}$. Then we have $3x \equiv 2 + 2 \cdot 5 \pmod{5}$, $3x \equiv 12 \pmod{5}$, and, finally, $x \equiv 4 \pmod{5}$. (It should be noted that the division of both parts of a congruence modulo m by their common factor q is legitimate only if q and m are relatively prime numbers.)[5]

A preliminary example. Suppose we want to solve the congruence $11x \equiv 2 \pmod{23}$. It would be impractical to keep adding 23 to the right-hand side until we get a multiple of 11. Let's look for a more elegant procedure. For instance, we note that $22x \equiv 4 \pmod{23}$ and subtract 23 from the left side. This yields $-x \equiv 4 \pmod{23}$, and, finally, $x \equiv 19 \pmod{23}$.

In the case of our initial equation $5x + 3y = 1{,}001$, the congruences work as follows: $3y \equiv 1{,}001 - 5x$; $3y \equiv 1{,}001 \pmod{5}$. Since $1{,}001 = 200 \cdot 5 + 1$, we have $3y \equiv 1 \pmod{5}$, or $3y \equiv 6 \pmod{5}$. So $y \equiv 2 \pmod{5}$, which means that $y = 2 + 5k$ ($k = 0, \pm 1, \pm 2, \ldots$). This solution is clearly equivalent to the one obtained above: $y = 332 - 5t$ ($t = 0, \pm 1, \pm 2, \ldots$).

The method of continued fractions

This method involves transforming an ordinary fraction a/b, composed of the coefficients of the equation $ax + by = c$, into the *continued fraction*

$$\frac{a}{b} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots \cfrac{1}{a_n}}},$$

which is written briefly as

$$\frac{a}{b} = [a_0; a_1, a_2, \ldots a_n].$$

[3]Of course, under the condition that she relates either three or five tales each night.

[4]The inventor of this method, L. F. Taylor (*Numbers*, London, 1970), didn't give any special name to it.

[5]Many interesting facts about congruence can be found in the two articles mentioned in the first section of this article.

The numbers

$$\frac{P_k}{Q_k} = [a_0; a_1, a_2, \ldots, a_k], \ k = 0, 1, \ldots, n$$

—called the *convergents* of the continued fraction—can be calculated by an inductive process using the recurrent formulas

$$P_{k+1} = P_k \cdot a_{k+1} + P_{k-1};$$
$$Q_{k+1} = Q_k \cdot a_{k+1} + Q_{k-1},$$

where $P_0 = a_0$, $Q_0 = 1$, $P_1 = a_0 \cdot a_1 + 1$, and $Q_1 = a_1$, $k = 0, 1, 2, \ldots$. Having computed the numerator and denominator of the next-to-last convergent P_{n-1}/Q_{n-1} directly or using the recurrences above, we can complete the solution of the equation by applying the ready-made formulas[6] representing its general solution in terms of these numbers:

$$\begin{cases} x = (-1)^{n-1} \cdot c \cdot Q_{n-1} + b \cdot t, \\ y = (-1)^n \cdot c \cdot P_{n-1} - a \cdot t, \\ t = 0, \pm 1, \pm 2, \ldots. \end{cases} \qquad (3)$$

Let's return to the equation $5x + 3y = 1{,}001$ for one last time. Perform the transformation of the number 5/3 into a continued fraction in detail:

$$\begin{array}{l} 3\overline{)5}\quad 1 \rightarrow a_0 = 1 \\ \quad 3 \\ \quad 2\overline{)3}\quad 1 \rightarrow a_1 = 1 \\ \qquad 2 \\ \qquad 1\overline{)2}\quad 2 \rightarrow a_2 = 2, \end{array}$$

so $5/3 = [1; 1, 2]$. The convergents are

$$\frac{P_0}{Q_0} = \frac{a_0}{1} = 1;$$
$$\frac{P_1}{Q_1} = a_0 + \frac{1}{a_1} = \frac{2}{1};$$
$$\frac{P_2}{Q_2} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2}} = \frac{5}{3}.$$

(Actually, we don't need the last convergent. Besides, it always equals the given fraction. We give this formula only to recall once again how the convergents are calculated directly from the definition.)

Since we have here $n = 2$, the numerator P_{n-1} and the denominator Q_{n-1} of the next-to-last convergent are

$$P_{n-1} = P_1 = 2;$$
$$Q_{n-1} = Q_1 = 1.$$

Now we're in a position where formulas (3) can be applied:

$$x = -1 \cdot 1{,}001 \cdot 1 + 3t,$$
$$y = 1 \cdot 1{,}001 \cdot 2 - 5t,$$

and, finally,

$$x = -1{,}001 + 3t,$$
$$y = 2{,}002 - 5t,$$

where $t = 0, \pm 1, \pm 2, \ldots$.

The answer looks different again, but you can easily verify that it's equivalent to its preceding forms. Here the particular solution $x = 1$, $y = 332$ emerges for $t = 334$.

Now let's solve one final problem.

After a shipwreck

Five sailors disembarked on an island and gathered a pile of coconuts before it got dark. They postponed dividing them up until the next morning. One of the sailors woke up at night, counted what they had collected, gave one coconut to the monkey they had with them, and took exactly 1/5 of the rest for himself. Then he went back to bed and was asleep in a minute. Some time later a second sailor woke up and repeated all these actions. Then, in turn, so did the other three. None of them had the slightest idea of what the others had done. In the morning they divided the remaining coconuts in equal parts, but this time the monkey got nothing. How many coconuts did the sailors gather?

Solution. Denote by x the unknown number of coconuts. We can write down the transformations that the number of the coconuts in the pile underwent as the following chain of equations: $x = 5a + 1$; $4a = 5b + 1$; $4b = 5c + 1$; $4c = 5d + 1$; $4d = 25y + 1$. (I leave it to the reader to tease out the sense of these equations.)

This system of simultaneous equations reduces to one indeterminate equation:

$$256x = 2{,}101 + 15{,}625y.$$

A quick solution of this cumbersome equation will be a nice reward for your patient work on the four methods presented above. You can choose the method that's most efficient for this particular problem. The smallest (positive, of course) answer is $x = 3{,}121$.

In *Mathematical Puzzles and Diversions*, Martin Gardner describes this problem as one of the most frequently attempted yet least mastered Diophantine puzzles. After this problem appeared in *The Saturday Evening Post* in 1926, letters continued to arrive for some 20 years, either requesting or proposing a solution.

Exercises

1. Find integer solutions of the equation $10x + 21y = 23$ using each of the four methods described in this article.

2. Find a two-digit number such that the digit in the ones place times eight is 13 less than the digit in the tens place times three.

3. A number of tourists were brought by bus to a railroad station in five equal groups. (Each bus holds no more than 54 persons.) Seven other persons joined them there, and they all were evenly distributed in 14 railroad cars. What was the total number of tourists?

4. Are there any integer points on the line $13x - 5y + 96 = 0$ whose coordinates do not exceed 10 in absolute value?

5. Let n be a positive integer. Find all integer solutions to the equation

$$nx + (n + 1)y = 2n + 1.$$

6. Fifteen liters of a liquid must be poured into bottles of volumes 0.5 l and 0.8 l so as to fill all the bottles completely. How many bottles of each kind will be needed?

7. Prove that for any odd x the congruence $x^2 \equiv 1 \pmod 8$ is true—that is, the square of any odd integer yields a remainder of one when divided by eight.

[6]These formulas are derived in most books on the elementary number theory. See, for instance, *Mathematical Excursions* by H. Merril (Dover Paperback, 1957).

In Foucault's footsteps

A simple experiment demonstrating the Coriolis force

by M. Emelyanov, A. Zharkov, V. Zagainov, and V. Matochkin

THE FIRST EXPERIMENT proving that the Earth rotates on its axis was done by Jean Bernard Léon Foucault in 1851. This article describes a simple experiment that also demonstrates the Earth's rotation quite clearly.

Imagine a large cylindrical vessel with a funnel at the bottom (fig. 1). The vessel is filled with water and suspended from the ceiling by a long cord. Initially the funnel is closed and the vessel is at rest relative to the Earth. What will occur when the funnel is opened at the bottom? For simplicity let's imagine we're conducting our experiment at the North Pole.

At the point where the vessel joins the funnel, the water flowing out has not only a vertical but also a horizontal velocity. Let's denote the horizontal projection of the water's velocity relative to the vessel by $\mathbf{v}_0$. This value $\mathbf{v}_0$ depends on both the height of the water column above this level and the distance from the vessel's axis.

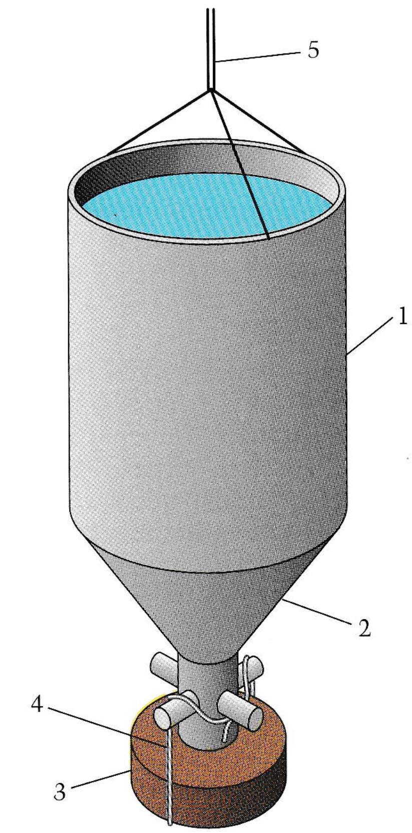


Figure 1

Experimental setup: (1) cylindrical vessel; (2) funnel; (3) soft rubber with a stiff base; (4) cord; (5) string from a badminton racket.

The presence of a nonzero horizontal velocity v_0 gives rise to a Coriolis force. Figure 2 shows the counterclockwise direction of the Earth's rotation, the velocity v_0, and the Coriolis force F_C acting on a certain layer of water. This force affects every bit of the water and produces a torque relative to the cylinder's symmetry axis. This causes the water to rotate. Since there is friction between the water and the wall of the vessel, the vessel will also rotate, and the direction of this rotation coincides with that of the Earth.

As the water pours out of the funnel, the velocity v_0 decreases, as does the Coriolis force (because $F_C \sim v_0$). How will the water move under the influence of this variable force F_C? The angular acceleration produced by this force decreases, but the angular velocity of the water increases, though more and more slowly.

The rotation of the vessel causes a winding of the cord that holds it. This produces a restoring torque, which increases with the angle of winding. Since the Coriolis force (which eventually causes the water in the vessel to rotate) decreases with time, there will be a moment when the torque due to this force is canceled by the increasing torque of the elastic force due to the cord. Then the torque due to the cord becomes greater than that produced by the Coriolis force. As a result, the angular velocity of the vessel is slowed, although the cylinder still rotates in the same direction as the Earth.

At a certain point the vessel stops and begins to rotate in the opposite

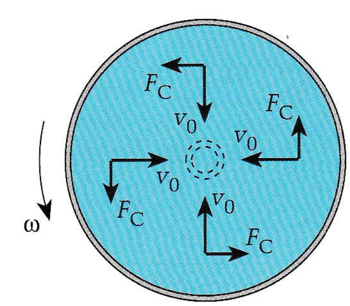


Figure 2

Top view of the whirlpool in the vessel.

direction due to the torque of the cord acting on it. The cylinder's angular velocity increases, and the restoring torque decreases as the cord unwinds. The torque due to the Coriolis force (which decreases as before) acts counter to the rotation of the vessel and slows its acceleration. At a certain moment the decreasing torque due to the cord becomes less than the torque due to the Coriolis force, and then the angular velocity begins to decrease. Then the cycle repeats itself.

When we did the experiment, we used a cylindrical vessel 25 cm in diameter and 30 cm high (fig. 1). At the bottom of the vessel we attached a funnel with an opening at the bottom that was 8 mm in diameter. We then attached a sleeve on the spout with four symmetrically placed screws. Using a string wound around each pair of opposing screws, we rigged a piece of soft rubber that had a stiff backing so that it closed the spout. We knotted the string on one of the screws.

To hang the device from the ceiling, we used a string from a badminton racket (about 2.5 m long). (You could use two strings, wound in opposite directions, to eliminate sidewards motion during the rotation.) We filled the vessel with water almost up to the rim, and after the system settled down we burned the knot on the thread. In every case we observed the same phenomenon: as the water flowed through the funnel, the suspended vessel rotated first counterclockwise, then clockwise.

You can also do this experiment with a glass funnel suspended from a pair of fishing lines. It turns out that the angle of the vessel's rotation depends very much on the diameter of the funnel's spout. On the one hand, the smaller the hole, the more pronounced the effect from the Coriolis force. On the other hand, the role of viscosity increases as the diameter of the opening decreases. We managed to observe a significant amount of rotation using a funnel containing 1 l of water that had a 5-mm opening at the bottom. ◘

THE POPULAR GAME OF BILliards requires not only a sharp eye and a sure hand, but also accurate calculation. The legendary hero of the Civil War in Russia, Marshal Semyon Mikhailovich Budyonny, used to say, "When I play billiards, I take lessons in physics and mathematics."

The art of playing billiards includes many fine tricks involving off-center strokes. These make the ball spin and, due to the friction against the cloth covering the table, curve its path. Effects of this sort were described by the well-known French engineer and physicist Gaspard Coriolis in his book *The Mathematical Theory of the Phenomena of Billiards*, published in 1835.

Here we'll consider only the simplest rectilinear motion of a billiard ball and the trajectories that result from collisions with the bumpers of variously shaped tables. After each collision the ball is reflected according to the familiar law of optics: *The angle of incidence equals the angle of reflection* (fig. 1). So the path of the ball coincides with the path of a ray of light. Note, by the way, that a photon can be regarded (in problems involving reflection) as a small billiard ball.

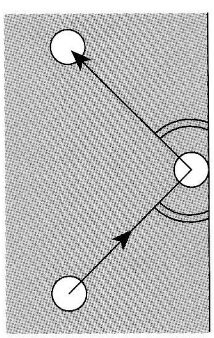


Figure 1

Let's begin with this problem: Given the location of the ball on a billiard table, determine the direction in which it must be struck so as to hit a bumper and reflect into the pocket at a given corner.

To solve this problem, imagine that the bumper is replaced by a mirror (fig. 2). Then the motion of the ball in the mirror after it hits the bumper (mirror) will continue the straight motion it had before the collision. Draw the image of the billiard ball as

Art by Dmitry Krymov

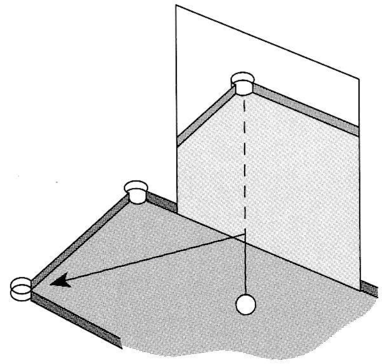


Figure 2

reflected in the given bumper and join the initial position of the ball with the reflection of the chosen corner (fig. 3). Now the required path of the ball is obtained by reflecting the segments of these lines about the corresponding bumpers.

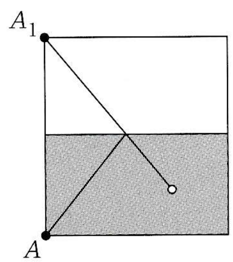


Figure 3

The next problem is a bit more difficult. Two balls are on the table, one red, the other white. Strike the red ball so that, after reflecting from the bumpers AB and BC, it hits the white ball.

Here again mirror reflections are helpful. First we reflect the table about AB and let C_1 be the image of C. Next we reflect the table's reflection about BC_1. Then we join the reflection of BC about AB. Now we join the second reflection of the white ball with the red ball and "fold back" the line thus obtained. It will give the required path (fig. 4).

And how will the ball move on a circular billiard table? It's clear that all the

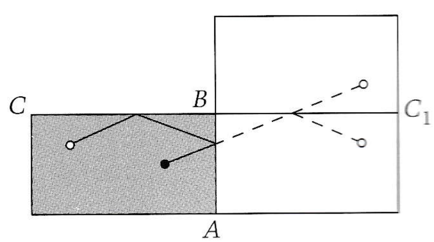


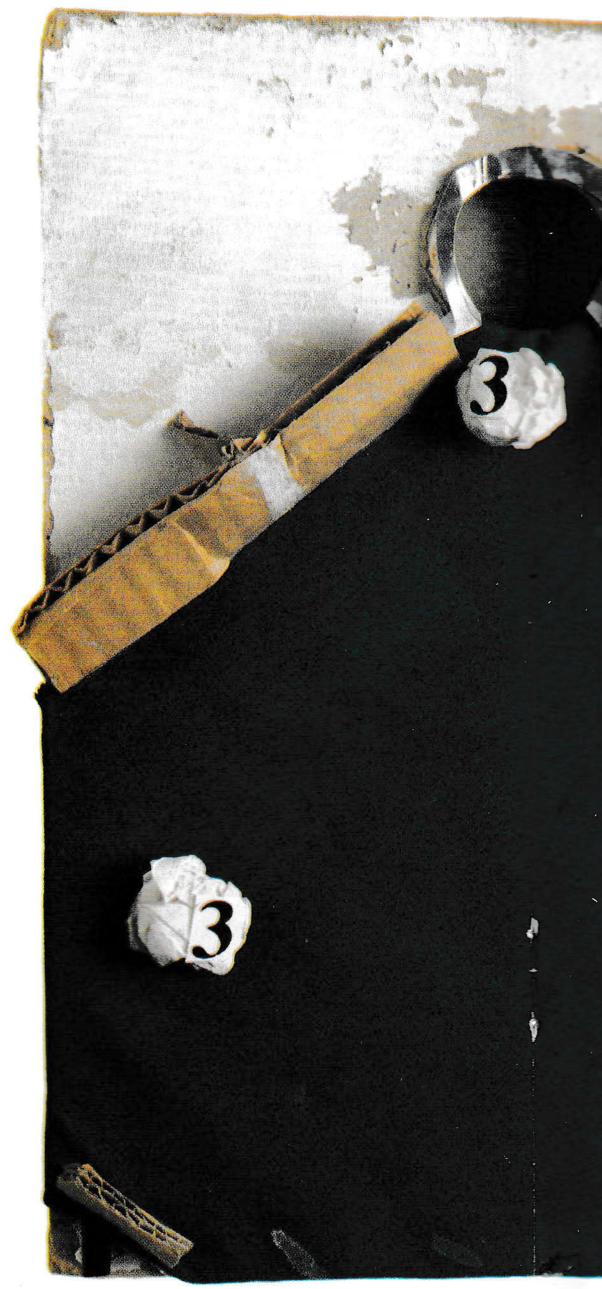
Figure 4

KALEIDOSC

Billiard r

Or, What I learned ir

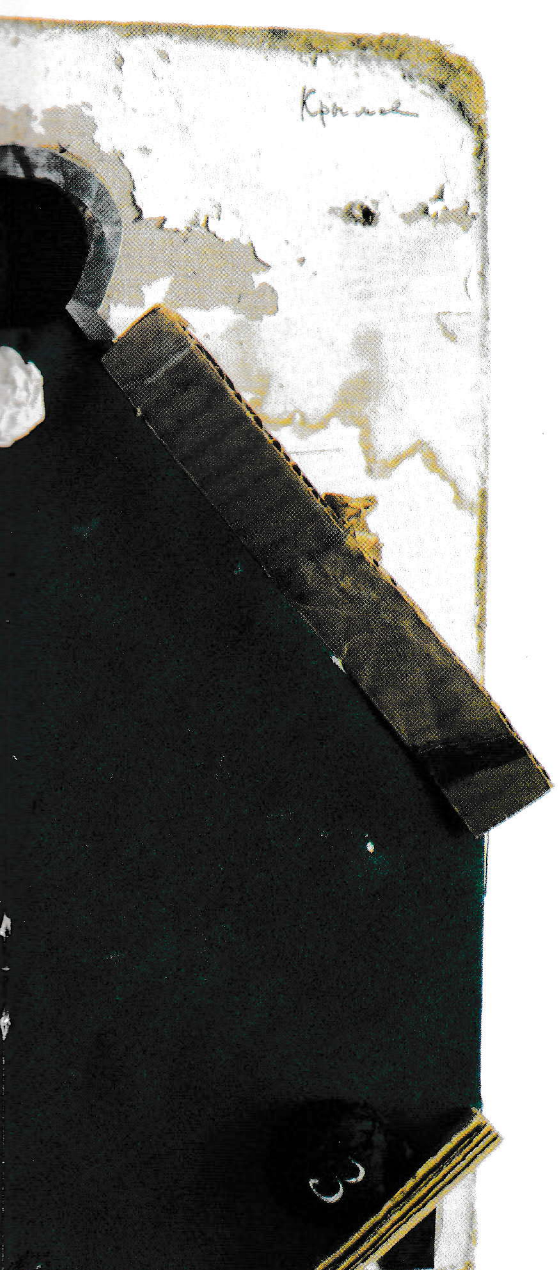
by Anatoly Sa



d math

ed in the pool hall

oly Savin



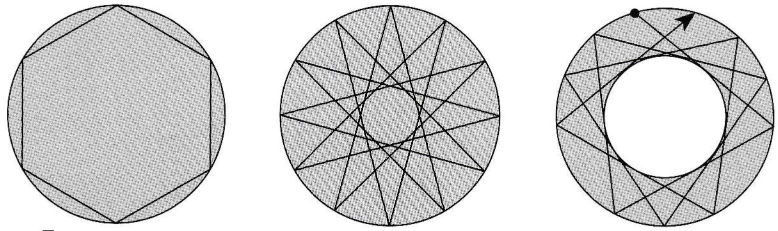


Figure 5

chords traced by the ball between two collisions with the bumper are the same length. Therefore, the ball's path will either be a regular convex polygon or a regular star polygon, or it will never close and the ball will sweep out a certain ring (fig. 5).

It's very interesting to watch the motion of the ball on an elliptical billiard table. The boundary of such a table can be defined as the locus of points M the sum of whose distances from two fixed points F_1 and F_2 (called the foci of the ellipse) is constant: $F_1M + F_2M = 2d$, for some number d (fig. 6).

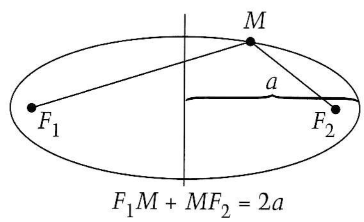


Figure 6

The ellipse has a remarkable property: a ball started from one of the foci will pass through the other after reflection on the table. This is called the optical property of the ellipse. One of its consequences is the fact that a ball that starts inside the ellipse, on a point on line F_1F_2 but *outside* the segment F_1F_2, will never cross the segment (fig. 7).

Mathematicians have long sought a polygon that would have two points M_1 and M_2 inside it such that a ball that starts at M_1 could never reach M_2 (fig. 8). At present most mathematicians believe that such a polygon is impossible, although this has not been proved.

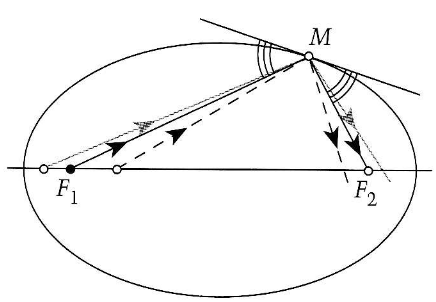


Figure 7

At the same time, it's not difficult to devise a curved "billiard table"

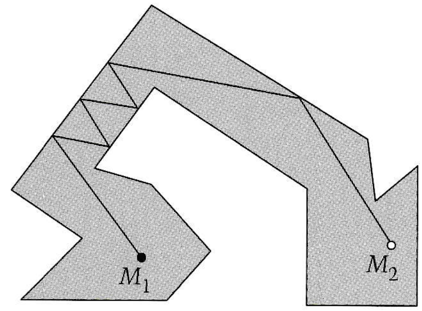


Figure 8

with this property. One example of such a table is shown in figure 9. The arc AD here is half an ellipse with foci at points B and C; the arcs AB, BC, and CD are semicircles. As was mentioned above, a ball that starts from a point M_1 in one of the smaller semicircles will never cross the segment BC, so it will never hit any point M_2 in the bigger semicircle.

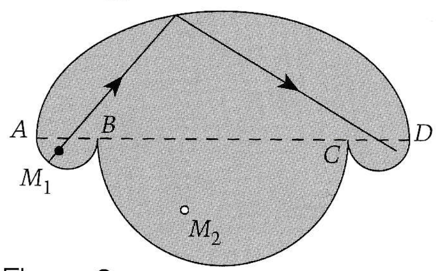


Figure 9

Mathematicians study the trajectories of a ball in even more intricately curved billiard tables. Why? Solutions to these types of problems help us understand the laws of motion of gas molecules or beams of particles in closed volumes, and these laws are useful in many areas of physics—in particular, in quantum electronics. The molecules are reflected off the walls exactly like a ball from the bumper of a billiard table. ◘

PHYSICS CONTEST

The nature of light

"The invisible influences of gravitation and electromagnetic fields remain magic; describable, but nevertheless implacable, nonhuman, alien, magic."—B. K. Ridley

by Arthur Eisenkraft and Larry D. Kirkpatrick

LIGHT PLAYS SUCH A CRUCIAL role in our lives that it's very hard to imagine a universe without light. Almost all of the information that we receive from outside our solar system comes to us in the form of light. Observations of the heavenly bodies and attempts to find regularities in their motions led to tremendous advancements in science and the modern-day scientific method. Studies of light and color revolutionized painting and the fine arts. The invention of the electric light allowed us to work and study at night. More recently the invention of lasers has had profound effects on our abilities to understand the world around us and to make great advances in technology used in such areas as surgery, cutting, welding, surveying, communication, the arts, advertising, and manufacturing.

But what is light? How do we describe its behavior? We have two basic models that we can use to describe light—particle behavior and light behavior. The debate over the best way to describe light has been waged for centuries. Newton felt that light was composed of tiny particles that traveled very fast. He used this idea to predict that light would travel faster in transparent materials like water and glass than in air. By the time the speed of light in water was measured by Jean Foucault in 1862 as slower than in air, the particle model for light was already out of favor. By 1801 Thomas Young had demonstrated the interference of light, a uniquely wavelike effect, and the wave model for light dominated for the next century.

As we discussed in the May/June 1995 issue of *Quantum*, Albert Einstein reintroduced the particle aspect of light in 1905 to explain the observations of the photoelectric effect. Einstein said the light can behave like a particle (known as a photon) that has an energy $E = hf$, where $h = 6.63 \cdot 10^{-34}$ J-s is Planck's constant and f is the frequency of the light.

Additional verification of these revolutionary ideas was provided by Arthur Holly Compton, the son of a Presbyterian minister. Compton was "turned on" to the study of X rays by his older brother and friend Karl. It was known that X rays are another form of electromagnetic radiation similar to visible light and radio, infrared, and ultraviolet waves. Therefore, the puzzle of light was the puzzle of X rays.

Compton's investigations began with the study of the angular distribution of X rays from crystals, for which he was awarded a Ph.D. from Princeton University in 1916. In his research Compton learned about Bragg scattering and was able to measure the wavelengths of X rays rather accurately. He discovered that the wavelengths of some of the X rays scattered by matter were lengthened.

After rejecting classical explanations for these observations, Compton combined Einstein's ideas about photons and relativity and emerged with a simple explanation for his observations. He assumed that X rays consisted of photons with energy $E = hf$ and momentum $p = E/c$. When a photon undergoes a collision with an electron, some of the energy and momentum of the photon is transferred to the electron, reducing the energy of the photon and consequently increasing its wavelength. Using the relationship $\lambda f = c$ for electromagnetic waves, where λ is the wavelength and c is the speed of light, Compton was able to calculate that

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta), \qquad (1)$$

where λ' and θ are the wavelength and scattering angle of the scattered photon and m is the mass of the electron. Compton verified this effect

Art by Tomas Bunk


BAR
PRIVATE
RADIOLOGY
DR. ROENTGEN
Photon
ELECTRON
ELECTRON
TOMAS
BUNK

experimentally and it is now known as the Compton effect. It is interesting that Compton was the one who suggested the name "photon" for light when it acts like a particle. Compton shared the 1927 Nobel prize in physics with Charles Wilson, the inventor of the cloud chamber.

Later Compton studied cosmic rays and helped to establish that cosmic rays are charged particles rather than high-energy electromagnetic waves. After working on the Manhattan Project during World War II, Compton became the chancellor of Washington University in St. Louis. His brothers became presidents of the Massachusetts Institute of Technology and Washington State University, both alma maters of one of your authors (LDK).

By the way, how does the debate about the nature of light come out? We now believe that light exhibits both particle and wave aspects depending on the types of measurement that we make. This is known as wave–particle duality and is a property of all particles at the subatomic level, including electrons and protons.

One of the problems on the semifinal exam used to select the 1996 US Physics Team was a one-dimensional, nonrelativistic derivation of the Compton effect and is the basis for this month's contest problem.

A. Consider the one-dimensional collision of a photon with a free electron initially at rest. Assume that the energy of the photon is much less than the rest energy mc^2 of the electron and the photon recoils in the backward direction with frequency f'. Write expressions for the conservation of energy and linear momentum.

B. Neglecting additive terms of order v^2/c^2, show that

$$h^2 ff' = \left(\frac{1}{2}mv^2\right)\left(\frac{1}{2}mc^2\right), \qquad (2)$$

where v is the electron's speed after the collision.

C. Show that equation (2) can be written in the form

$$\lambda' - \lambda = \frac{2h}{mc},$$

in agreement with the formula for the Compton effect. Note that the change in wavelength does not depend on the original wavelength.

The quantity h/mc is known as the Compton wavelength and has the value $2.43 \cdot 10^{-12}$ m = 2.43 pm—a very small change. This change is difficult to measure unless λ is also small. Compton used X rays with a wavelength of 71.1 pm.

D. What is the energy of these X rays? Can the electrons in matter be treated as if they were free? Does the recoil energy of the electron satisfy the condition for a nonrelativistic treatment?

E. If you would like to work more with this effect, try obtaining the two-dimensional result given in equation (1). Let ϕ be the angle of the electron leaving the collision. Write down the equations for the conservation of energy and the two components of momentum. Use these three equations to eliminate ϕ and v. You will need to use the fact that the collision is nonrelativistic to neglect small additive terms. (Most textbooks on modern physics give the relativisitic derivation.)

Please send your solutions to *Quantum*, 1840 Wilson Boulevard, 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*.

Moving matter

Two readers sent in excellent solutions to the May/June contest problem. Congratulations to Noah Bray-Ali, a junior at Venice High School in Los Angeles, California, and Wolfgang Wais of Tübingen, Germany. We will follow along with Noah as he describes his approach.

A. There are three parts to this problem: the jump, the collision, and the upswing. Conservation of energy during Tarzan's (mass M) swing from height h_0 gives his speed v_0 at the bottom of the valley:

$$Mgh_0 = \frac{1}{2}Mv_0^2, \qquad v_0^2 = 2gh_0. \qquad (1)$$

During the ensuing perfectly inelastic collision with Jane (mass m), momentum is conserved, and we can solve for the couple's final velocity v':

$$Mv_0 = (M+m)v', \qquad v' = \frac{Mv_0}{M+m},$$

or, with equation (1),

$$v'^2 = \frac{M^2}{(M+m)^2}v_0^2 = 2gh_0\frac{M^2}{(M+m)^2}. \qquad (2)$$

We can now find the height h to which they rise using conservation of energy during their upswing:

$$\frac{1}{2}(M+m)v'^2 = (M+m)gh'. \qquad (3)$$

Solving for h' using equations (2) and (3) yields

$$h' = \frac{M^2}{(M+m)^2}h_0 = 4.44 \text{ m}.$$

Alas, Tarzan and Jane end up more than half a meter short of the crest of the hill they had hoped to reach.

B. When the bullet (mass m) and pendulum (mass M) collide, momentum is conserved:

$$mv_0 = (m+M)v',$$

where v_0 is the original velocity of the bullet and v' is the velocity of the system after the collision. Thus

$$v'^2 = \frac{m^2}{(m+M)^2}v_0^2. \qquad (4)$$

After the collision, the bullet–pendulum system rises to a height h, which we can find from the given parameters: the length L of the cord and the horizontal distance s traveled by the system. If we denote by θ the angle the cord makes with the vertical, we have

$$\theta = \sin^{-1}\frac{s}{L} = 3.44°,$$

$$h = L - L\cos\theta = L(1-\cos\theta) = 0.009\ \text{m},$$

From conservation of energy we can solve for the bullet's speed:

$$\frac{1}{2}(M+m)v'^2 = (M+m)gh.$$

Replacing v'^2 using equation (4), we get

$$v_0^2 = 2gh\frac{(M+m)^2}{m^2}, \quad (5)$$

or

$$v_0 = 420\ \text{m/s}.$$

C. If the spring of force constant k is initially compressed a distance d, from conservation of energy during compression we can solve for the speed v_0 at which the ball (mass m) is ejected:

$$\frac{1}{2}kd^2 = \frac{1}{2}mv_0^2,$$

$$v_0^2 = \frac{kd^2}{m}. \quad (6)$$

All of the work in part B is still true, so we can combine equations (6) and (5) and solve for h, where M is the mass of the pendulum:

$$h = \frac{kd^2}{2g}\frac{m}{(M+m)^3}.$$

To solve this for the relative masses that optimize h, we take the derivative of h with respect to m, set it equal to zero, and solve for m in terms of M. Noting that the term $kd^2/2g$ is constant and can be factored out, this reduces to

$$0 = \frac{dh}{dm} = \frac{1}{(M+m)^2} - \frac{2m}{(M+m)^2},$$

and finally

$$M = m.$$

D. This question can be broken into two parts: the collision and the upswing. From the conservation of angular momentum about the pivot point of the rod, we can calculate the final angular speed ω of the system after the collision in terms of the ball's speed v_0 after ejection, which we know from part C:

$$mv_0L = I'\omega,$$

where m is the mass of the ball, L is the length of the rod, and I' is the moment of inertia of the system after the collision. We can calculate I by the superposition principle, knowing that I for the rod is $^1/_3ML^2$:

$$I' = I + mL^2 = \frac{1}{3}ML^2 + mL^2.$$

Replacing and solving for θ gives

$$\omega = \frac{mv_0}{\left(\frac{1}{3}M + m\right)L}.$$

During the upswing, the rotational kinetic energy from this collision is converted into gravitational potential energy. The ball rises a height $L - L\cos\theta$, and the center of mass of the rod rises $L/2 - L/2\cos\theta$, where θ is the final angle we're trying to find. So conservation of energy gives

$$\frac{1}{2}I'\omega^2 = \left(m + \frac{M}{2}\right)gL(1-\cos\theta).$$

Replacing I' and ω and solving gives

$$1 - \cos\theta = \frac{3m^2v_0^2}{(M+2m)(M+3m)gL}.$$

Equation (6) is still valid, so we can replace and solve:

$$\theta = \cos^{-1}\left(1 - \frac{3kd^2m}{gL(M+2m)(M+3m)}\right),$$

which expresses the final angle in terms of the masses of rod and ball.

E. From equation (6) we know the velocity v_0 of the marble is proportional to the compression d of the spring. This horizontal velocity is constant during the flight, so the final distance traveled in the x-direction will simply depend on the time of flight, which is the same for both trials: the time it takes to fall from the table to the floor. The distance traveled in the x-direction without air resistance is thus directly proportional to v_x and, therefore, to the compression distance. If we denote the values of the first attempt by d_0 and x_0 and those of the second attempt by d and x, this proportionality gives

$$d = d_0\frac{x}{x_0} = 1.11\ \text{cm}.$$

Shady computations

Clearing up a paradox at the boundary of dark and light

by Chauncey W. Bowers

Art by Sergey Ivanov

THE SPEED OF LIGHT AS AN upper limit arises as the result of specific ramifications of the special theory of relativity. In particular, the mass/energy equivalence ($E = mc^2$) requires an infinite amount of energy to accelerate any mass to the speed of light. In addition, to maintain the notion of causality, nothing can influence events faster than the speed of light. However, the limits imposed by the laws of physics permit certain measurable attributes of our everyday world to exceed the speed of light. Specifically, the movement of a shadow's edge can greatly exceed the speed of light. In fact, shadows can exhibit behavior normally prohibited by common sense—for example, they can arrive at a destination before leaving their origin.

Consider a shadow cast by a wall AB of height Y. The angle of the incident light with the ground is θ. Figure 1 shows that this situation results in a right triangle that can be analyzed simply with basic trigonometry rules. To this end, the light is assumed to hit the wall as parallel rays, as would occur if the light source is very far away—for example, the Sun. It follows that the length of the shadow on the ground is $AC = Y \cot\theta$. If the wall is lowered a distance Δy (from point B to B'), the shadow on the ground will move Δx (from point C to C'), such that $\Delta x = \Delta y \cot \theta$. For example, if $\theta = 30°$, the cotangent $\cong 1.73$, and $\Delta x \cong 1.73\Delta y$. The average speed of anything, including the edge of a shadow, is simply the distance traveled divided by the time taken to travel that distance. For the shadow's edge, the average velocity is $\Delta x/\Delta t$, where Δt is the time it takes for the shadow to move Δx. Because $\Delta x = \Delta y \cot\theta$, all that remains is to calculate Δt for the shadow. The time for the shadow to move from point C to C' is a function of how quickly the wall is lowered. The wall cannot be lowered faster than the speed of light, so we will denote the speed of the wall as ac, where c is the speed of light and $a < 1$.

If the wall is lowered at a speed much lower than c (exactly how much slower will be shown later), the speed v_s of the shadow's edge will simply be

$$v_s = ac \cot \theta. \quad (1)$$

This is because the movement of the shadow's edge can be considered instantaneous with that of the wall—that is, the time it takes for the shadow to move can be approximated as the time it takes for the wall to move, but the distance the shadow moves is $\cot\theta$ times that of the wall.

In fact, there is a delay between the time the wall begins to move and the time the shadow begins to move—the shadow will not move until light has had time to travel the distance $BC = Y/\sin\theta$. This time delay equals

$$\Delta t_1 = \frac{Y}{c \sin\theta} \quad (2)$$

—that is, the distance traveled by the light divided by its velocity.

When does the shadow complete its movement, stopping at point C'? If the wall is lowered Δy at a speed ac, then the time it takes the *wall* to complete its journey is simply

$$\Delta t_w = \frac{\Delta y}{ac}.$$

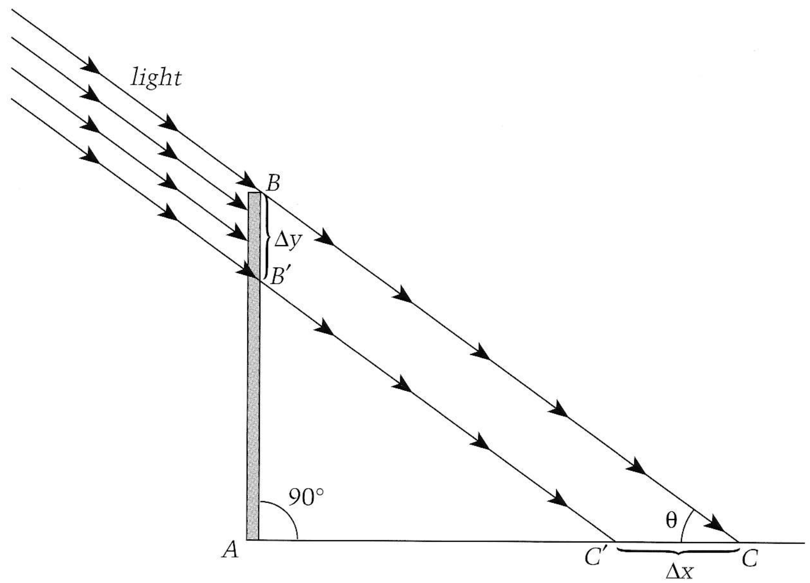


Figure 1

The shadow will not "arrive" at point C' until the wall has arrived at B' *and* light has traveled from B' to C'. This distance $B'C'$ is $(Y - \Delta y)/\sin\theta$, and the time for light to travel that distance is

$$\Delta t_2 = \frac{Y - \Delta y}{c \sin\theta}.$$

We now have everything needed to calculate the average speed of the shadow's edge. The time Δt_s for the shadow to move from C to C' is equal to the time Δt_w for the wall to move from B to B' plus the time Δt_2 for light to move from B' to C' minus the time Δt_1 it takes light to move from B to C, or

$$\Delta t_s = \frac{\Delta y}{ac} + \frac{Y - \Delta y}{c \sin\theta} - \frac{Y}{c \sin\theta} = \frac{\Delta y}{c}\left(\frac{\sin\theta - a}{a \sin\theta}\right).$$

The speed of the shadow is therefore

$$v_s = \frac{\Delta x}{\Delta t_s} = \left(\frac{ac \cos\theta}{\sin\theta - a}\right). \quad (3)$$

Equation (3) is the general equation for the speed of the edge of a shadow cast by a wall at an angle θ, where the wall is being lowered at a speed ac. As one would expect, the speed of the shadow is independent of the wall's height and the distance the wall is lowered. The shadow's velocity is only a function of the velocity of the wall and the incident angle of light. As expected, for small values of a ($a \ll \sin\theta$),

$$v_s = \left(\frac{ac \cos\theta}{\sin\theta}\right) = ac \cot\theta,$$

which is equation (1).

The behavior of the shadow described by equation (3) is somewhat surprising. For an angle θ of 30° and a wall being lowered at $0.4c$ ($a = 0.4$), the speed of the shadow is $3.5c$! In fact, the shadow can be made to travel as fast as one likes until $a = \sin\theta$, at which point there is a singularity in equation (3). The speed of the shadow is undefined mathematically at this singularity, and this correlates with the behavior of the shadow. When $a = \sin\theta$, light will appear *instantaneously* over the region C to C'. Thus the speed of the shadow's edge is physically, as well as mathematically, undefined. If the speed of the wall is increased so that $a > \sin\theta$, the sign of the shadow's velocity changes. Again, this correlates with the physical behavior of the shadow. When $a > \sin\theta$, light will appear at point C' *before* the edge of the

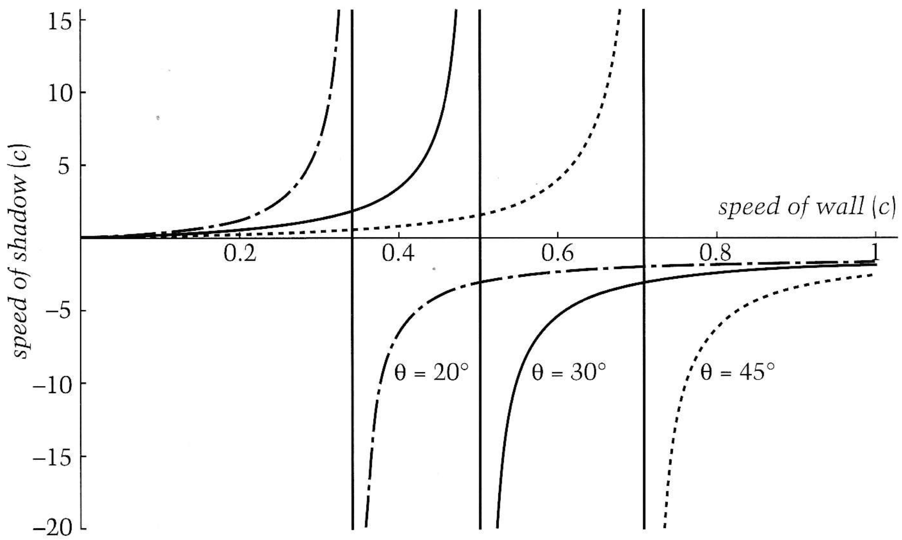


Figure 2

shadow has moved from point C. The light will then proceed from left to right in figure 1 (rather than right to left) with the velocity calculated from equation (3). The shadow will, in effect, arrive at its destination (C') before leaving its origin (C). Not only does increasing the velocity of the wall past the singularity cause the direction of the shadow's velocity to reverse, but the speed of the shadow *decreases* as the wall speed *increases* after this point.

Graphs of equation (3) for various angles θ are shown in figure 2, demonstrating more clearly the singularity and its relation to θ. Decreasing θ decreases the wall speed required to produce the singularity. For $\theta = 5°$, the wall only needs to move at $0.08c$ to produce a shadow velocity greater than $10c$.

It is left to the reader to verify that the ability to produce shadow speeds greater than c can be accomplished only by lowering the wall, not by raising it. Indeed, even if the wall could somehow be raised instantaneously, the velocity of the shadow's edge would only equal $c \cos \theta$.

This derivation of the shadow's behavior indicates why the mass–energy equivalence is not a factor for shadow velocity. Simply put, there is no mass moving in the direction of the shadow's edge. Its movement is the result of the *timing* of light traveling to the ground as the wall is lowered. Furthermore, the shadow cannot be used to influence events faster than the speed of light. In this regard, it should be remembered that the shadow does not begin to move from point C until *after* light has traveled the distance BC. The subsequent movement of the shadow at speeds exceeding the speed of light cannot then influence other events faster than this built-in delay (for $a < \sin \theta$). For $a > \sin \theta$, light will appear at point C' ahead of the delay expressed by equation (2), but after light has traveled the distance $B'C'$. Thus any *influence* is slower than the speed of light, since we must also wait for the wall to move from B to B'.

Poets have often alluded to the boundary between light and dark as an area of strange and paradoxical behavior. Scientists must, in this case, agree that the poet has seen an aspect of verifiable truth. While the edge of a shadow is easily defined and measured, its movement is not limited by the velocity of the light that creates it.

Chauncey W. Bowers *is a researcher in the Division of Neurosciences at the Beckman Research Institute of the City of Hope. His e-mail address is cbowers@coh.coh.org.*

Feeding rhythms and algorithms

The true subject of bovine ruminations

by Dr. Mu

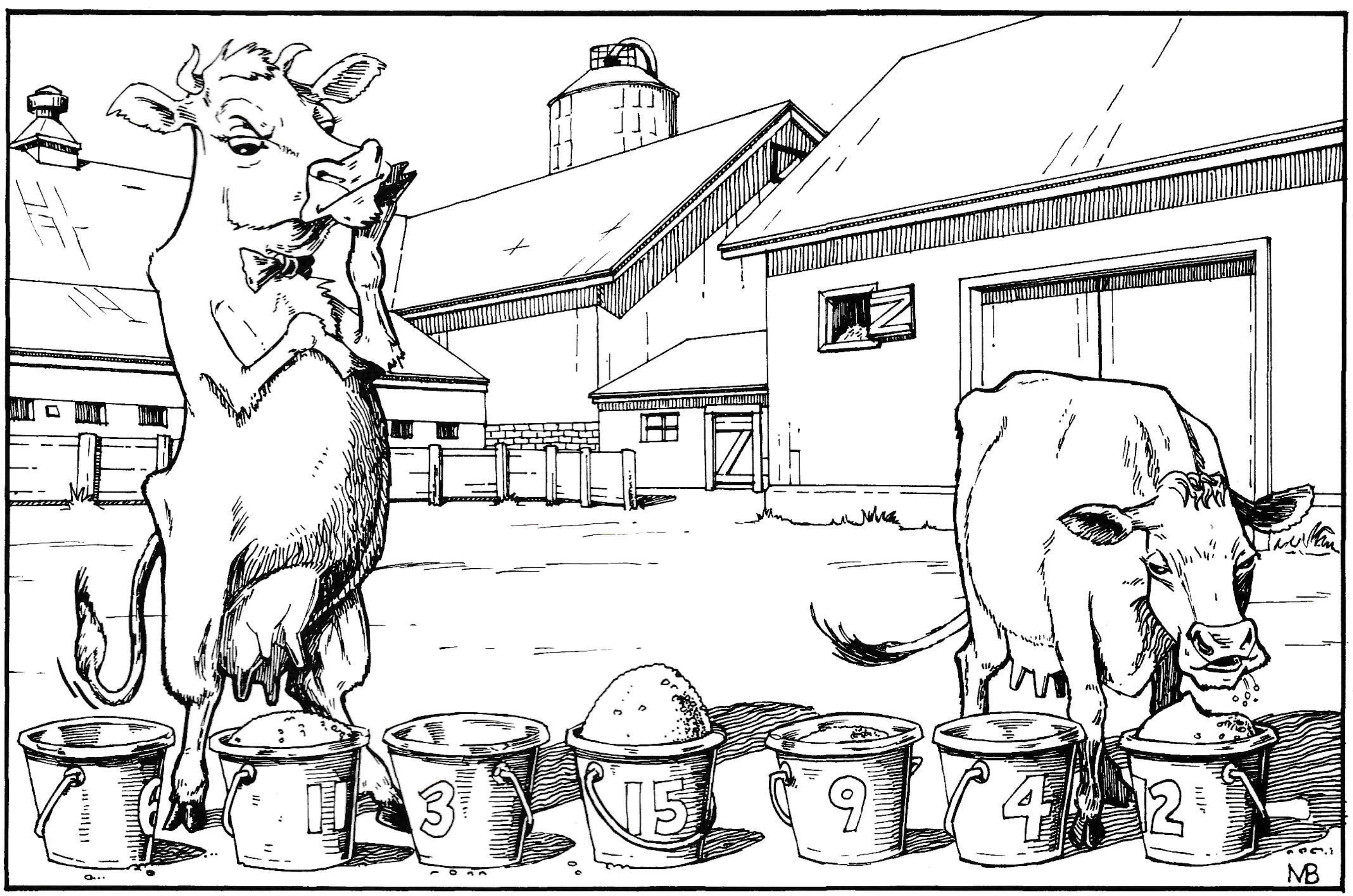
HUMANS HAVE NO IDEA WHAT WE COWS think about. You see us resting peacefully under a shady tree, contented, unconcerned with the worries of the world, chewing our cud, and assume that nothing is going on between our ears.

But you're sadly mistaken. We think a lot about problems that interest us—like our next meal. Let me say up front, I love my chow. Try producing 60 pounds of milk a day (on average) on an empty stomach or two. Being a cow who always likes to get its fair share, I've been doing some cowculations along this line.

My boss, farmer Paul, has been playing a weird game lately. When we come in for milking he's been setting out a row of eight pails, each filled with varying amounts of feed. (It looks to us as though they are randomly filled.) Two cows are assigned a row of pails and given the following instructions:

Each bucket is numbered with the number of scoops of feed inside. Going one at a time, pick a pail from either end of the row, eat the contents, and remove the pail from the row. Take turns (waiting for the other to eat and remove the pail) until all the food is gone.

Some of my casual bovine mates take a cavalier approach to this task. No matter who goes first, they flip



Art by Mark Brenneman

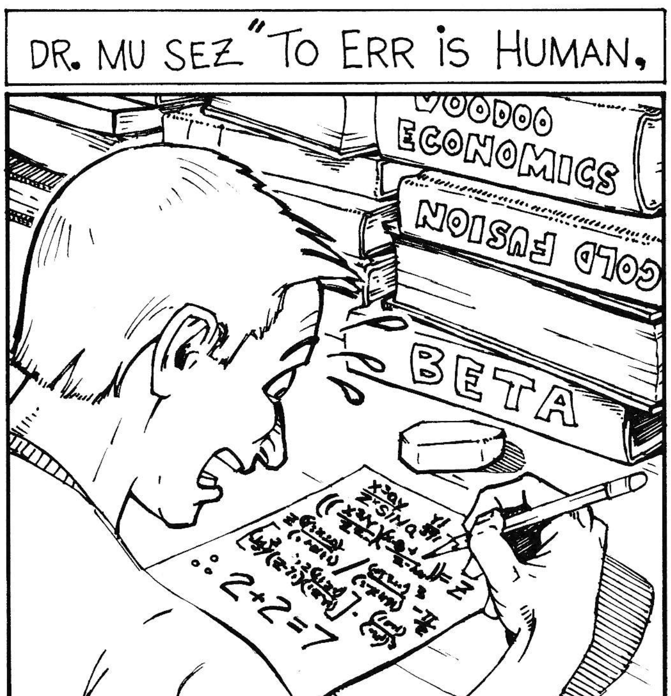
a coin to decide which side to pick. They're easy to beat—I just pick the end with the highest numbered pail. (I call this my Greedy algorithm.) This usually works, but every once in a while they get lucky and beat me.

For example, the other day farmer Paul placed the following row of feed pails {1, 2, 3, 2, 4, 2, 15, 6} for Bessie and me to select from. I went first and took 6, leaving {1, 2, 3, 2, 4, 2, 15}. Bessie picked 15, so it was clear I wasn't going to overcome her lead, and I went a bit hungry that day.

Hunger may bring out the best in us, but I want to make sure it doesn't happen again. I want to be certain that I get an equal share or better—no matter what! In other words, when farmer Paul places an even number of pails and fills each with a random amount of feed and when I pick first, I never want to end up with less than half the chow.

So here's your "Challenge Outta Wisconsin," or COW, as we say around the dairy state: With your cowculator, write a program that will fill *p* pails (*p* is even) with random amounts of feed (integers between 1 and *s*) and, by picking first, always win against any cow—or human (they think they're so smart!).

My cowculator uses the highly advanced software Mathematica™. I like Mathematica because I can easily examine a wide range of computer algorithms in a very high level language that reads like a mathematical expression. You see, Mathematica is a functional language in which programs are all mathematical expressions. It's also easy to learn for beginners who have never programmed before. But you can use Pascal, C/C++, or BASIC if you prefer. However, it's the algorithm that I intend to focus on and not the language. If you use a different language, you can implement my algorithm



using your own language's syntax.

Here is a look at my cowculations.

First I defined a function that picks out *p* random integers from 1 to *s*. This is done by defining a function **feed[p,s]** that generates a "Table" of *p* "Random Integers" between 1 and *s*:

```
feed[p_,s_]:=
Table[Random[Integer,{1,s}],{p}]
```

Try out the feed function for $p = 8$ and $s = 25$ and place the feed in a row.

```
p=8;s=25;
row=feed[p,s]
```

```
{18,19,1,11,25,12,22,14}
```

Bessie and I start with nothing, which is represented by empty lists "{}" assigned as follows:

```
Bessie={};
DrMu={};
```

Using the Greedy algorithm, if the "First" pail in the row has more (or possibly the same) amount of feed as the "Last" pail, then "Append" this to my feed list and "Drop" it from the row. Otherwise, pick the "Last" pail, "Append" it to my feed list and "Drop" it from the row.

Suppose Bessie goes first, and we "Do" this $p/2$ times, which exhausts the feed. To display the results, we "Print" it. Here is the Mathematica code that translates what I just said in a very precise and exact manner. Comments such as (*then*) and (*else*) were added for your ease of reading and are not required for the program.

```
row
Do[If[First[row]>=Last[row],
     (*then*)
     AppendTo[DrMu,First[row]];
     row=Drop[row,1],
     (*else*)

AppendTo[DrMu,Last[row]];row=Drop[row,-
1]];

Print["Dr Mu eats ",Last[DrMu], " leav-
ing ",row];

If[First[row]>=Last[row],
     (*then*)
     AppendTo[Bessie,First[row]];
     row=Drop[row,1],
     (*else*)

AppendTo[Bessie,Last[row]];row=Drop[row,-
1]];

Print["Bessie eats ",Last[Bessie], "
leaving ",row], {p/2}];
```

Here is the output when you run this program:

```
{18,19,1,11,25,12,22,14}
Dr Mu eats 18, leaving
{19,1,11,25,12,22,14}
Bessie eats 19 leaving
{1,11,25,12,22,14}
Dr Mu eats 14 leaving {1,11,25,12,22}
Bessie eats 22 leaving {1,11,25,12,}
Dr Mu eats 12 leaving {1,11,25}
Bessie eats 25 leaving {1,11}
Dr Mu eats 11 leaving {1}
Bessie eats 1  leaving {}
```

Time to add up the chow. Do this by "Apply"ing the "Plus" operation to my selections and Bessie's:

```
DrMuTotal=Apply[Plus,DrMu];
BessieTotal=Apply[Plus,Bessie];
Print[DrMu, " = Dr Mu's picks for a
total of ",DrMuTotal]
Print[Bessie, " = Bessie's selection
for a total of ",BessieTotal]
```

```
{18,14,12,11}= Dr Mu's choices for a
total of 55
{19,22,25,1} = Bessie's selection for a
total of 67
```

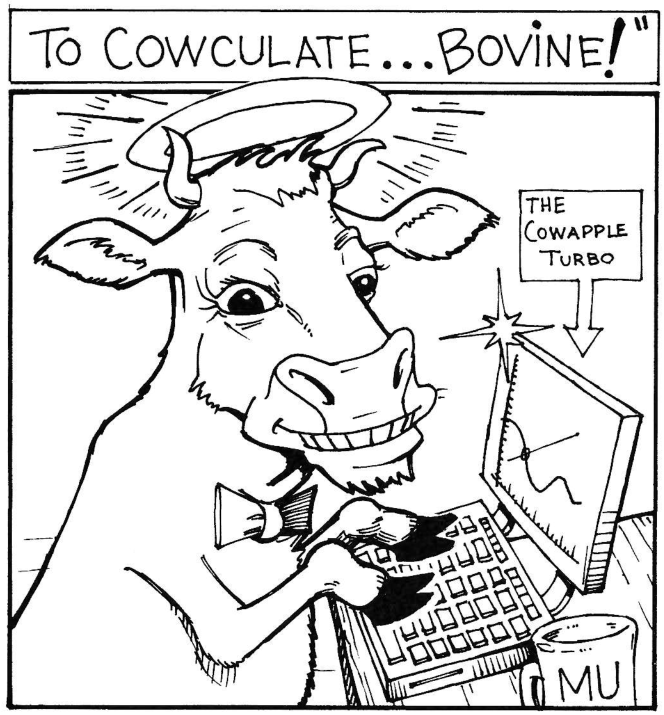
Holy cow! I came up short again! Can you help me avoid this? I don't like to go hungry. Discover an algorithm that wins every time.

Your turn to feed at the trough.

COW 1.a. Find an efficient algorithm I can use to always win at chow time over Bessie. Your solution should run within a few seconds, winning quickly even with 100 pails. Remember, I have the advantage of going first.

COW 1.b. Do some cowculations to estimate my chances of beating Bessie if we both use the Greedy algorithm. (Ties are considered a win for me.)

You can e-mail your cowculations to me, Dr. Mu, at drmu@cs.uwp.edu. I'll maintain a home page of some of the best cowculations at http://usaco.uwp.edu/cowculations. By the way, if you happen to like computer programming a lot, check out the USA Computing Olympiad at http://usaco.uwp.edu.



If you're interested in doing your cowculations in Mathematica and don't currently own a copy, consider the offer available to students from Wolfram Research at http://www.wolfram.com/mathematica/info/students.html. The student version is no different than a regular version of Mathematica—except, of course, for the price (~$109).

I need to see your cowculations before December 1, my deadline for the next column. If you have any questions, come out to the farm and I'll try to answer them after milking time. I check my e-mail daily.

Note: This problem originally appeared (unbovined of course) at the International Olympiad in Informatics in Veszprem, Hungary, July 25–August 1, 1996. The home page for IOI'96 is http://frey.inf.bme.hu/contests/ioi96.

Now it's time to do your bit and e-mail those bytes to drmu@cs.uwp.edu.

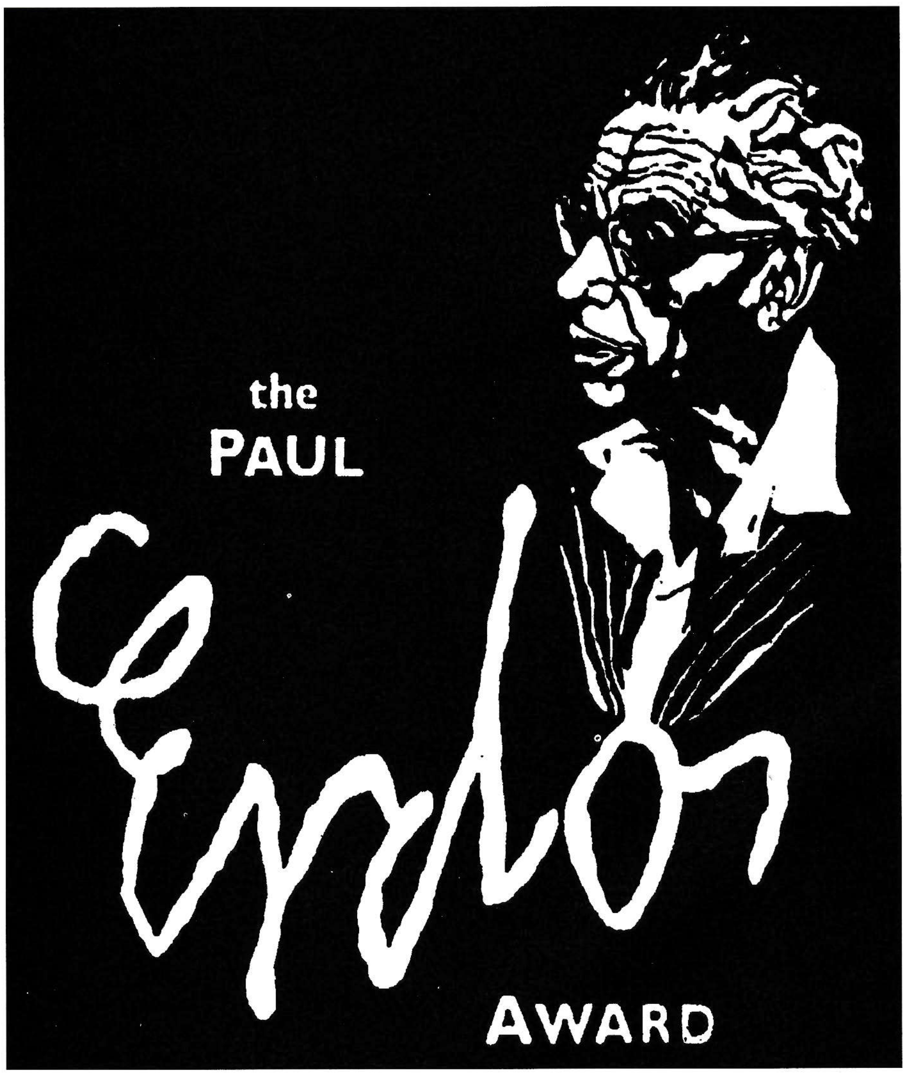
Dr. Mu *is the Bovine Professor of Cowculation Science at the University of Wisconsin–Parkside and resides in Paulsbarn.*

In memoriam: Paul Erdős (1913–1996)

May he enjoy seeing his proofs faithfully recorded in The Book

by George Berzsenyi

AT THE AWARDS CEREMOnies for the eight winners of the 1980 USA Mathematical Olympiad, I was especially happy to see Michael Finn and Eric Carlson among the honorees, since they were also two of the winners of the year-round competition I had been conducting through the Competition Corner of the now defunct journal *Mathematics Student* (*MS*) of the National Council of Teachers of Mathematics. Since my major aim with *MS* was to popularize creative mathematical problem solving among high school students, I was fully aware that the problems posed there were not challenging enough for Michael and Eric. Hence I suggested to them that they should switch their attention to more demanding problem sections, like those in *The American Mathematical Monthly* and *Mathematics Magazine*. Michael's answer was "I do them also," while Eric reminded me of the saying that "practice makes perfect." And thus I continued to receive beautiful solutions to the *MS* problems from both of them, throughout their high school years.



This image of Paul Erdős is based on a painting by Ray Paul in the possession of the University of Cincinnati's Mathematics Department Library.

Time and again when I receive various reactions and contributions to the present column I think of Michael and Eric, and I want to suggest to my readers to turn their attentions to more worthy tasks, more serious mathematical investigations. While many of the topics covered in this column may be attractive and even fascinating, they are often lacking in depth and seriousness, which are the hallmarks of real mathematical investigations. While I am capable of devising appropriate problems for competitions and assisting in the discovery and development of mathematical talents, I am not an expert in probing the frontiers of mathematics.

These were my thoughts when I recently learned about the death of Paul Erdős, who devoted his life to probing those frontiers, posing and/or solving some of the most important problems of this century. Hence I want to recommend that my readers learn more about the legacy of Erdős, follow in his footsteps, and emulate his attitude toward mathematics. To whet your appetite, I reproduce below three of his problems. They were communicated to me in 1994, in response to my tribute to him in the May/June 1994 issue of *Quantum*.

Problem 1. Let $f(n)$ be the largest integer for which there is a set of n distinct points, $S = \{x_1, x_2, \ldots, x_n\}$, in the plane with the following property: for every x_i in S there are at least $f(n)$ points in S that are equidistant from x_i. Determine $f(n)$ as accurately as possible. Is it true that $f(n) = o(n^\varepsilon)$ for every $\varepsilon > 0$?

Erdős offered $500 for a proof and "much less for a counterexample."

Problem 2. Let seven points be given in the plane. Prove that one can always choose three of the points so that the three distances determined by them are all different.

Erdős went on to say that for six points the above claim is not true, and wondered whether the only counterexample is provided by the vertices and the center of a regular pentagon. He also wanted to know how many points are needed to ensure that one can always choose four of the points so that the six distances determined by them are all different. Clearly, one can also relax the conditions on the four points and/or extend the problem to more points. Moreover, the smallest counterexamples may be of interest too.

Problem 3. An old conjecture states that $10! = 6!7!$ is the only nontrivial solution of $n! = a!b!$. (If $n = k!$, then $(k!)! = (k! - 1)!k!$ is a trivial solution—for example, $24! = 23!4!$.) Try to represent $n!$ as the product of smaller factorials (for example, $8! = 7!2!2!2!$) and prove that the density of n's for which this is possible is 0.

I last saw "Pali Bácsi" in the summer of 1994 at the Second Congress of the World Federation of National Mathematics Competitions (WFNMC) in Bulgaria. I was hoping that he would manage to attend the 8th International Congress of Mathematics Education held in Seville, Spain, this past summer, where I was honored by the WFNMC's "Erdős Award" for my contributions, but he couldn't come. And now I will never see him again! We will all miss him, not only as one of the greatest mathematicians, but also as one of the most gentle, thoughtful pillars of the society of mathematicians. One hopes there is indeed a "Great Book in the Sky" containing the most elegant proofs of every theorem in mathematics, and he is there to enjoy it. ◘

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

George Berzsenyi *is a professor of mathematics at the Rose-Hulman Institute of Technology, 5500 Wabash Avenue, Terre Haute IN 47803-3999. His e-mail address is george.berzsenyi@rose-hulman.edu.*

AT THE BLACKBOARD II

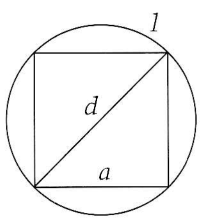
Not all is revealed

On the Uncertainty Principle and other forms of indeterminacy

by Albert Stasenko

WHAT COULD BE MORE trivial from the geometric viewpoint than a square with a diagonal, or a circle with a diameter? Any child can draw them by hand (fig. 1). And yet so much has been written about these figures, one could fill an entire encyclopedia—for example, about the irrationality of the numbers π and $\sqrt{2}$.

What would it have cost the Creator to construct Nature in such a way that the ratio of the circle's length to its diameter is exactly equal to 3, or even exactly 3.14, or 3 and a hundred (too small? let it be a million!) digits after the decimal point—just so that it were strictly equal to something! But no—scholars have proved that the number π contains an infinite number of digits after the decimal point. (This property is referred to as the *incommensurability of the circumference and the diameter*.) Students and teachers through the years have devised sayings to help them remember the first dozen digits. Others have found algorithms to calculate as many digits as we want and have translated these procedures into computer language. Even before the computer age there were devotees of science who spent their entire lives calculating a few hundred digits—such was the overwhelming desire to discover what was over the horizon. But however long one may calculate—even to the end of the time allotted to humankind—there will be no end to the sequence of digits in the number π.



$$\frac{l}{d} = \pi = 3.1415926536\ldots?$$

$$\frac{d}{a} = \sqrt{2} = 1.414\ldots?$$

Figure 1

> "THE GODS did not reveal all things to mortals. Searching on their own, people have learned little."—Xenophanes
>
> "I PLUNGE INTO the depths and stand before the mystery of the world, the secret of all that exists. And each time I am painfully aware that the existence of the world cannot be self-sufficient, cannot but have behind it, in the even greater depths, a Mystery, a secret meaning."—Nikolay Berdyayev

Why is the world constructed this way? What mystery lurks in the cross-section of a spruce log, or a Corinthian column? Isn't it outrageous that modern science in all its power cannot precisely say how many times the circle is larger than its diameter? It's a wonder mathematicians can even sleep at night!

Well, so much for mathematics. Is everything in physics nice and precise? No. Everyone now knows that physics is always an approximate model of the real world. Every measurement has an error, but as the years and centuries pass, the measurements become more and more precise. Perhaps we can hope that someday (albeit in the distant future) we will be able to say in principle where at a given moment a material point is located on the x-axis, and what its speed v is. After all, this is nothing more than the ABCs of kinematics.

Unfortunately, it is *in principle* that this is impossible—forbidden, in fact! Mathematically this is formulated in Heisenberg's famous Uncertainty Principle:

$$\Delta x \cdot \Delta p \gtrsim h, \qquad (1)$$

where Δx is the uncertainty in the measurement of the coordinate (in meters), Δp is the uncertainty of the momentum (in N · s), and $h \sim 10^{-34}$ J · s is Planck's constant.

So according to the Uncertainty Principle, we can't determine in the (p, x)- plane (that is, momentum versus position) the location of the center of mass C of the object under consideration. The more precisely we try to measure the position x of the point (that is $\Delta x \to 0$), the worse will be our simultaneous measurement of its momentum ($\Delta p \to \infty$), and vice versa. We can only say that p and x of a point C are located somewhere within a figure whose area is not less than Planck's constant h (fig. 2). Now, this area is so small, it doesn't bother us when we're analyzing the motion of airplanes, projectiles, discusses, and Ping-Pong balls. But doesn't it put you on guard, doesn't it bother you that this kind of restriction exists *in principle*? As someone once said: "I'll probably never go to Australia. But if you forbid me from going there, I'll be unhappy forthwith."

Let's modify inequality (1) and apply it to a photon. Since a photon's momentum is $p = h\nu/c$, where c is the speed of light and ν is its frequency (corresponding to the photon's "color"), then $\Delta p = h\Delta\nu/c$ (since h and c are constants, any uncertainty in the photon's momentum can only be related to that of its frequency). Dividing both members of the equation by h we get

$$\frac{\Delta x}{c}\Delta\nu \gtrsim 1,$$

or

$$\Delta t \cdot \Delta\nu \gtrsim 1. \qquad (2)$$

Here we took into account that $\Delta x = c\Delta t$.

So it turns out that the shorter the duration of the photon emission ($\Delta t \to 0$), the greater the uncertainty of its frequency ($\Delta\nu \to \infty$). For example, we see the red light of a neon sign, which is the result of radiation by excited neon atoms. The frequency of the emitted photons $\nu_0 \sim 5 \cdot 10^{14}$ Hz, while the emission

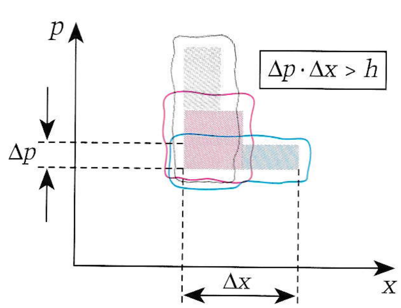


Figure 2

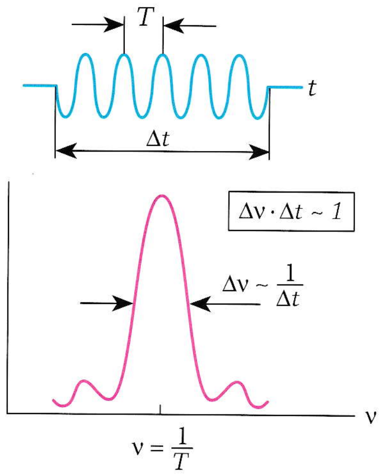


Figure 3

time for each atom $\Delta t \sim 10^{-9}$ s, so the uncertainty of the frequency is no less than $\Delta\nu \sim 1/\Delta t \sim 10^9$ Hz, which is billions of cycles per second! But even this value isn't that large compared to the photon's frequency—it's smaller by a factor of a million: $\Delta\nu/\nu_0 \sim 10^{-6}$. Still, it means that the emitted light isn't exactly "red," but rather a combination of an infinite number of other frequencies, mainly in the interval we found (fig. 3).

The same relationship (2) is also valid for musical notes (in this case c is the speed of sound in air): however hard a musician tries to produce an absolutely pure pitch—say, the note A—she can't do it. The resulting sound will in principle contain a large number of frequencies, even if the musician could force the string to keep vibrating for days, or for a whole year.

It looks like we need to go back to the beginning and look for determinacy under the canopy of mathematics. Why not start with Euclidean geometry? Isn't that a classically pure, almost marmoreal example of strictness? Not at all! Even Euclidean geometry has its own brand of uncertainty. It's not because the people who created and developed geometry formulated the fifth postulate (on the parallelism of straight lines) in a rather irritatingly awkward way, which provoked the appearance of the non-Euclidean geometries named after Lobachevsky, Riemann, and others. No, the problem is that people attempted to construct other deductive sciences similar to Euclidean geometry, and tried their best to give them the same degree of "strictness."

It seemed that all that was needed was to formulate a set of assertions (an axiomatic system), then all other assertions (which use the same terms as the axioms) could logically be either proved or disproved. Such a good system of axioms is referred to as *complete* and *consistent* (two qualities of "goodness").

To illustrate these properties of an axiomatic system, consider a simple example. Any vector in the plane can be viewed as the sum of two components—its projections on two coordinate axes:

$$\mathbf{a} = a_1\mathbf{e}_1 + a_1\mathbf{e}_2, \qquad (3)$$

where the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are of unit length and perpendicular to each other. If you try to "broaden" this basic system of vectors by adding yet another vector $\mathbf{e}_3$, you'll hear: "Excuse me, but you can't do that in two-dimensional space" (fig. 4) The new vector $\mathbf{e}_3$ can be broken down according to equation (3), so in this sense the system of vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ is *complete*. If you insist, then go ahead and add your $\mathbf{e}_3$, but this will mean that you're passing into another kind of space—three-dimensional space (or, analogously, you're constructing another axiomatic system).

So what has developed in the space of assertions (if you can imagine such a space)? In 1931 Gödel's theorem was proved, which said in

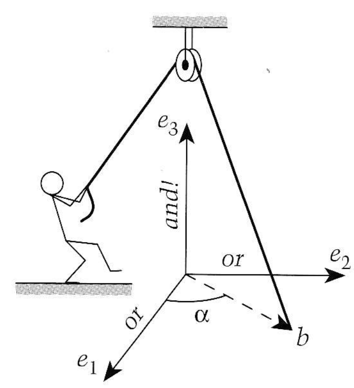


Figure 4

effect that a simultaneously complete and consistent axiomatic system does not exist. (Doesn't this sound a lot like Heisenberg's Uncertainty Principle (1)—not in the (p, x)-space, of course, but in the space of "completeness–consistency"?) If you fix the number of axioms, then sooner or later you come to an assertion that can neither be proved or disproved. If you still want to deal with this particular assertion, you must enlarge the initial axiomatic system.

Let's take a break from our heavy thoughts and recall an old story. A certain John took a certain Benjamin to court, complaining that the defendant (Benjamin) had stolen his only cow. "You should know this cow, your Honor once drank some of her milk," said John. "Of course I remember, the cow is yours," answered the judge. "But your Honor, you know that I have fourteen children, and all of them drink milk—this should be my cow," Benjamin implored. "You are also right," the judge declared, after thinking it over. "But your Honor, they both can't be right—there is only one cow!" the court clerk chimed in. This time the judge thought long and hard, and then he told the clerk: "You are also right."

This story is an illustration of a case where the judge decided to move into another space (he constructed the vector $\mathbf{e}_3$ and thereby enlarged the axiomatic system). He left behind the standard formal logic, where the law of the excluded middle is valid (that is, either one or the other must be true), and moved from dialectics to trialectics, which doesn't set white against black,[1] friend against foe. As a result there is no need for judges, or "permanent revolution," or concentration camps. It's a place where the ancient Vedic trinity (Indra, Agni, Surya), the Hindu trinity (Krishna, Shiva, Vishnu), or the Christian Trinity reigns.

But what does the value of π, the Uncertainty Principle, and Gödel's theorem have to do with all of this? Simply this: these concepts (and many others) contain, it seems, some inner meaning that frees us from the iron grip of Necessity, but not completely (total uncertainty would likely be chaos)—a kind of hint of freedom. Something like the little "window" through which Pyotr Florensky tried to peek into the "other side" of our world. Or the testimony of St. Paul: "Now we see only puzzling reflections in a mirror, but then we shall see face to face. My knowledge now is partial; then it will be whole, like God's knowledge of me." (1 Corinthians, 13:12)[2]

"There are theoretical calculations that imply a universe that consists, perhaps, of two worlds laid one on the other, very weakly linked, almost invisible to each other. . . . It is completely possible that in our neighborhood, in the very same space and time, a hidden parallel world exists, exactly like our own, or perhaps completely different." (V. A. Barashenkov, as quoted in *The View from Nowhere* by A. S. Kuzovkin, MNPP "Yanga-center," 1991, p. 31)

"The world is continually splitting into innumerable copies of itself. . . . According to a theory of Everett's, the observable universe is only one instance of the infinite variety of actually existing universes." (P. Davis, *The Accidental Universe*, Moscow: Mir, 1985, p. 149)

So where have we come to—a place where any clever student, pointing to principles of indeterminacy, has the right to study subjects only very approximately, saying: "Well, that's the Divine Plan"? Far from it! The inquisitive reader should try to learn the details with the maximum precision. That is the only way to approach any threshold in Nature, where we have the feeling of an ineffable mystery. And we wonder: why is the threshold there, and what's behind it—the possibility of taking another step, or an insurmountable prohibiting principle?

[1]Take the Sun, for example. It illuminates our days with white light, but from the point of view of thermodynamics and quantum mechanics, it's almost an ideal example of an absolutely black body—and this isn't just wordplay!

[2]As rendered in *The New English Bible* (New York: Oxford University Press, 1971). Readers may be more familiar with the King James Version: "For now we see through a glass, darkly; but then face to face: now I know in part; but then shall I know even as also I am known."

FUNDAMENTALS

A prelude to the study of physics

Some guiding thoughts for novices on the construction of models and on their role in science

by Robert J. Sciamanda

NO PHYSICIST OR ENGINEER ever solves a real problem. Instead she creates a model of the real problem and solves this model problem. The model must satisfy two requirements: it must be simple enough to be solvable, and it must be realistic enough to be useful—that is, it must be both conceptually understandable and empirically fruitful.

The theories and "laws" of physics are also models. Whether in solving a particular engineering problem or searching for the wide ranging laws of physics, the art of scientific analysis consists in the creation of useful models of reality. The model is the interface between reality and the human mind. As such, the model must be expressed in human terms; it is cast in terms of concepts that we create from the data of our experience. Our models speak as much about us, our experience, and our modes of thought as they do about the external reality being modeled.

I prefer to speak of models where others might speak of theories because the word "model" emphasizes the criterion of usefulness. We tend to think of a theory as a candidate for some absolute, objective truth; a model is used to convey useful information without the pretense of being unique, complete, or ultimate. As an example of the conception, gestation, birth, and growth of a model in physics, let's consider the history of the "ideal gas law" $PV = \alpha T$, which you undoubtedly have studied in your physics or chemistry classes.

The history of a model

Despite the voluminous abstractions of philosophers over many centuries, no useful understanding of gas behavior emerged in ancient times or the Middle Ages. The possibility of a useful model awaited the creation of the thermometer and the manometer. Each of these devices uses a thread of mercury embedded in glass in order to generate a number (the length of the mercury thread) that varies in value as the device is subjected to varying conditions. Boil, Charles, and Gay Lussac investigated the behavior of these devices when connected to a gas under controlled conditions.

To condense a very long story, their experimentation resulted in the creation of the empirical relation $PV = \alpha T$, the variables P and T representing the readings of the manometer and thermometer, respectively; α is a constant for a fixed quantity of gas. If we then *define P, V, and T to be measurements of properties of the gas*, $PV = \alpha T$ becomes a useful model of gas behavior, even though P and T, at this point, have no deeper meaning—they are merely numbers generated by the specified devices.

That there should exist any relation (let alone such a simple one) among the numbers generated by these (or any other) devices is not at all to be expected. Such serendipity can only be gratefully contemplated when it appears. It is an instance of the profound meaning in one of Einstein's most famous quotations: "The most incomprehensible thing about the world is that it is comprehensible."

The creation of the model $PV = \alpha T$ was a giant leap forward. Note that the crucial beginning step consisted in the free creation of a set of concepts in terms of which meaningful questions might be put to nature so that nature might respond in a meaningful way. These concepts are not lying in nature awaiting discovery by some passive act of looking. They must be actively created. This is how the properties of matter come to be. This is how we define into existence those measurable properties of reality that we find useful. They are human constructs in terms of which we might ask meaningful questions, read nature's answers, and organize our understanding into useful and testable models.

Each of these concepts is quantitative in nature: the number generated by a measuring device. Our empirical gas law is simply a relation (and a very useful one) among the numbers (P, V, T) generated by our measuring devices. It is an *empirical* model. The numbers generated by measuring devices have no deeper meaning except within the notion of a *conceptual* model of the system being measured and its effect upon the measuring devices.

Boyle did his experimentation in the 1600s, while the Pilgrims were colonizing America. It wasn't until the mid-1800s, while Americans were fighting over slavery, that Joule brought together the theories (models) of Newtonian mechanics and atomism (then hotly contested) to create a *conceptual* model of the ideal gas as a system of *randomly moving point particles*. In this model P is quite naturally associated with the Newtonian force concept and accounts for the behavior of the mercury manometer. However, there is no *a priori* mechanical association for the empirical quantity T—the "temperature" of the gas as generated by the thermometer.

Herein lies a wonderfully simple instance of the incredibly awesome power of an empirically based analytical science: the fruitful interaction of experimental and theoretical physics. Newton's laws drove Joule's conceptual model to a very illuminating result: the numerical value of the product PV for Joule's gas is proportional to the total kinetic energy of the randomly moving gas particles. Thus Joule's conceptual model bestows upon the empirical temperature T, in $PV = \alpha T$, a deeper meaning as a humanly invented property of the gas. It becomes a measure of the energy of random motion of the gas particles.

A model of models

Thus it is that the mathematical model $PV = \alpha T$ has foundations as both an empirical model and a conceptual model. I present it as a paradigm to illustrate the properties of the model in physics:

1. It is a human construct, the offspring of both our experience and our imagination.
2. It is quantitative and speaks of freely defined, measurable properties of matter.
3. It has both an empirical and a conceptual usefulness: it presents a testable numerical equality involving the numbers generated by specified measuring devices, and it offers a conceptual framework for associating a deeper meaning with these numbers.
4. The empirical usefulness of a model is a matter of experimental verification, and once verified this usefulness will remain, Future models of a wider scope will include it as a special case.
5. The conceptual usefulness of a model can be a cultural matter, a matter of institutional and personal taste (more of this later).

Conceptual limitations

Our conceptual models are of course produced from the data of our experience. Every now and then I close my eyes and carefully feel an object such as a piece of fruit, a table, or my own face, and try to imagine what it might be like to have never had the sense of sight. What sort of conceptual models might I fashion as I explore reality using only the sense of touch? (Try to form the concept of the shape of an object without invoking a visual image.) How could I appreciate the language of a sighted person? There is no way that a sighted person could convey to me his conscious experience of light vs. darkness, let alone red vs. green. Our conceptual models could communicate only through shaky analogies and metaphors, but our empirical models could unambiguously communicate regarding the numbers generated by measuring devices.

Conceptual models are observer dependent and observer limited. As the physicist probes into the behavior of reality, she strives to create meaningful conceptual models of that reality, using as raw materials the concepts fashioned from human experience. As she probes deeper she finds that she has to become ever more creative and imaginative, generating abstractions and cross-fertilizations of her ideas in order to conceptually model the behavior of reality in human terms.

There is no reason to expect that this process can be extended indefinitely. It seems reasonable to anticipate that beyond a certain level of analysis the behavior of reality cannot be conceptually modeled in literal human terms, even though we may continue to be clever enough to create numerical equalities involving the readings of our instruments. After all, our instruments operate on the same superficial level as our senses.

We are already on the doorstep of this conceptual barrier. The mathematical models of quantum theory defied even the imagination of Albert Einstein. He was never able to conceive a satisfactory conceptual model of the reality behind these equations. As regards creative "weirdness," modern art and music are poor seconds to modern physics, even though the arts operate completely free of any constraints, whereas physics operates under the severe constraint of empirical usefulness!

Standards, taste, and beauty

Suppose that you are shipwrecked on a desert island and, with nothing better to do, decide to create the science of physics from scratch. You decide that your first task will be to choose (or design) standards for your measurements of space and time intervals. How should you choose a standard measuring rod and a standard clock? This is a "catch-22" question: one would like to have these standards available *a priori*, so that one can perform experiments (both physical experiments and thought experiments) to ask questions of nature, read her answers, and be guided toward a theory about the behavior of matter. Yet one's choices of a standard clock and measuring rod already presuppose considerable understanding about the behavior of matter! For example, the choice of a standard clock already presupposes a theory that will be committed to the conclusion that this particular mechanism ticks at a constant rate. Logical consistency will force the theory to this conclusion. Choices among theories and choices among standards are inextricably intertwined.

The dilemma exposed in the above paragraph is not debilitating. We need only replace the word "theory" (a candidate for an absolute, objective truth) with the word "model" (a useful way of describing reality in human terms). In this view, the choice of a clock simply defines into existence a measurable parameter "t" that will be used as a linear time base for the description of the evolution of phenomena. We will be comparing the course of all other phenomena to the succession of ticks of this clock.

Clearly the choice of standards is a matter of free definition. The criterion is not one of truth; it is simply one of usefulness: which choices lead to the most "desirable" empirical and conceptual models of reality? Put another way: how "weird" does the conceptual model have to get in order to be empirically useful? The words "desirable" and "useful" must be defined by you and/or the current scientific culture; they are a matter of taste. Historically, and logically, this is an iterative process, as we see more and more details of where the model is leading.

Let me tease you with a famous example (which I hope you will study in detail later). Einstein, in his 1905 relativity theory, was the first to capitalize on this freedom of choice (of rods and clocks) in a radical way. His definitions of "desirable" and "weird" were not mainstream. To him the desirable model must preserve the invariance (sameness) of physical law (in particular, Maxwell's equations of electrodynamics) for al nonaccelerated observers. But conventional wisdom said that the velocities appearing in Maxwell's equations must be measured from an absolute frame of reference (the "aether" frame). This was "desirable" to many—they found it satisfying that the laws of physics should be simple only to an observer at absolute rest. In fact, any deviation of your experimental results from the laws of physics would then furnish you with sufficient data to measure your own absolute velocity They had been disappointed that Newton's model of mechanics did not allow us to measure our absolute velocity by *mechanical* experiments (Newton himself must have been disappointed). They were overjoyed that now Maxwell's model of electrodynamics (which includes light) would allow us to measure our absolute velocity using *optical* experiments.

Einstein conceived a completely different conceptual model for Maxwell's electrodynamics. He sought a model in which these equations could be used with equal validity by all nonaccelerated observers, each using the numerical values of all quantities (for instance, velocities) as measured from her frame. He dared to redefine the measurement of space and time intervals to make this so. It is to be expected that such a redefinition would force new and worse weirdities into the model. We surely should expect that we will have to design new clocks and measuring rods, with exotic "relativistic" properties. The remarkable result has been that the new weirdities were only cultural—that ordinary clocks and meter sticks behave relativistically, and that a vast scope of phenomena have become more simply describable, even phenomena far removed from Maxwell's equations. Widespread acceptance did not come quickly or easily, but today relativity is not only accepted as empirically and conceptually useful, it has become beautiful!

The search for beauty in our models has always been a driving force and sometimes, as with Einstein's relativity, it seems to have been the sole motivation. Today many, like Einstein, are disappointed in their search for intuitive beauty in the quantum aspects of modern physics. Unlike relativity, the beauty of quantum theory still eludes visceral human appreciation. Perhaps with time we can acquire a taste, but it must begin with an adjustment of our expectations—toward models rather than "theories." Physics does not offer any quieting and ultimate answers.

Your personal physics

Physics has not been idle. There is much for you to learn. To learn means to make your own; it is an active process that only begins with listening and reading. You must return often to listening and reading, but meaningful learning comes only from contemplation. Each person must construct his own models and his own philosophy of what physics is. These will grow and develop—construction is never complete. What I have said here is subject to criticism by scientists, philosophers, students, and even myself, as my appreciation of physics continues to develop. These words should be taken as providing only a beginning for discussion and contemplation. I have tried to express my current philosophy to you. Over the years you will build your own unique and personal version. Even more than the appreciation of a symphony or a painting, the understanding of physics is a unique and personal encounter of a consciousness with reality.

Robert J. Sciamanda *is an associate professor of physics at the Edinboro University of Pennsylvania. His e-mail address is sciamanda@edinboro.edu.*

Bulletin Board

StudyWorks! by MathSoft

Word processors have made great strides in recent years, allowing users to format equations and print them accurately and elegantly. But they won't do any computing for you, and they certainly can't recalculate every formula in your document that uses a factor you just changed. If you're looking for a single, integrated software package that will allow you to incorporate "live" equations in your homework or lab report, StudyWorks!™ for Math or StudyWorks!™ for Science may be just the thing for you.

The StudyWorks! CD-ROMs include a wealth of reference material, including common formulas that can be dragged and dropped directly into a document. StudyWorks! for Math supports algebra, geometry, precalculus, calculus, and statistics. StudyWorks! for Science supports physics, chemistry, Earth science, biology, and statistics.

Users can visit MathSoft's World Wide Web site directly from the program and open StudyWorks! files that others have uploaded. There is also a unique forum called the Collaboratory™, where users can exchange ideas, work collaboratively on a problem, or get help on homework problems. (A brief visit to the Collaboratory reassured this reviewer that the MathSoft folks provide helpful hints and guidance, not the answers outright.)

Space is lacking here to list all the features of this powerful program, developed by the company responsible for the popular Mathcad software. The Math Palette allows you to do calculus, create "live" graphs (that is, graphs that automatically change when new data are introduced), and even animate these graphs. The StudyWorks! multimedia tutorial is more than adequate to get the user up and running. The "user's guide," however, is just that—a slim 44-page booklet (identical for the math and science versions of the software), not the "manual" that some might expect. Clearly the StudyWorks! developers expect users to be comfortable with online help, which is abundant and well organized.

One aspect of the software that may cause some frustration is the way text and math are handled as separate "regions." This allows StudyWorks! to perform its computational wizardry, and to know what to work on (math) and what to leave alone (commentary). But it also prevents the program from stretching a region when you add more text to it. If your text now overlaps an equation that comes after it, you must move that equation (and everything below it, if the equation now bumps into the next region). This "cut-and-paste" approach should be fine for short assignments, but could cause headaches if you're working on a long, complex document. StudyWorks! is not a true "word processor," as we now understand the term, but its emphasis understandably lies elsewhere.

StudyWorks! for Math and StudyWorks! for Science each carry a retail price tag of US$39.95, placing it well within the reach of most students and teachers. It's relatively easy to learn, and the computation engine is powerful and fast. For more information, visit the MathSoft Web site at **http://www.mathsoft.com**.

—TMW

Weighty CyberTeaser

The November/December CyberTeaser (brainteaser B187 in this issue) proved a light task for most of those who responded to the contest at our World Wide Web site. Some used reasoning much like that of the problem's author. Others came up with ingenious approaches of their own. And some left a challenge for our CyberJudge: "Find my method (if you want to avoid a lot of tedious arithmetic verification)!"

Here are the first ten respondents who provided the correct answer (and an explanation):

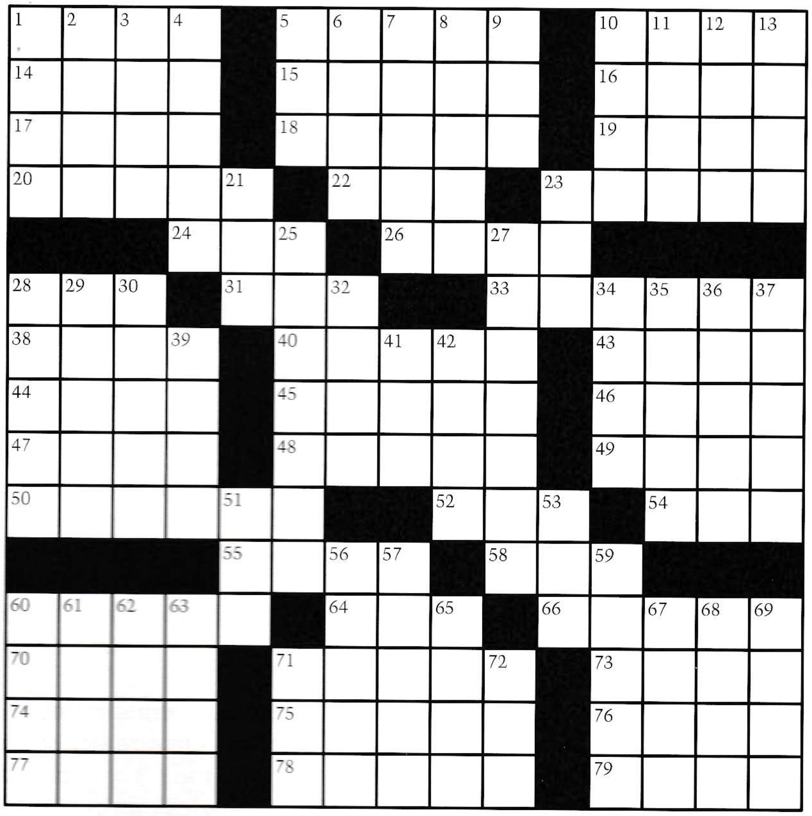
Jean-Baptiste Legros (Fontainebleau, France)
Steve Hunter (Ascot, Berkshire, UK)
Oleg Shpyrko (Cambridge, Massachusetts)
John J. Drozd (London, Ontario)
Leo Borovskiy (New York, New York)
Ted Lau (Fenton, Michigan)
Kiran Raj (Berkeley, California)
Iljong Lee (Berkeley, California)
Clarissa Lee (Perak, Malaysia)
May T. Lim (Quezon City, Philippines)

Each of them will receive a *Quantum* button and a copy of this issue. Also, everyone who submitted a correct answer is eligible to win a copy of *Quantum Quandaries*, our collection of the first 100 brainteasers from *Quantum* magazine. Congratulations to all our winners, and thanks to all who entered.

Now, who'd like to take a crack at the latest CyberTeaser? (Let's not always see the same hands . . .) Go to **http://www.nsta.org/quantum**, click on the "Contest" button, and keep on going!

Criss × cross science

by David R. Martin



Across

1 German chemist Lambert ___ (1818–99)
5 Adjust to
10 HIV retrovirus disease
14 Nucleotide sequence
15 Islamic architect (1489?–1587)
16 Type of cheese
17 Musical sound
18 Nose and mouth
19 Abampere
20 Astronomer ___ Cannon (1863–1941)
22 Computer language
23 Unit of magnetic flux
24 ___ value (avg.)
26 Mallophaga
28 Trig. function
31 4074 (in base 16)
33 Potential ___
38 German astronomer ___ Schwarzschild (1873–1916)
40 Electromagnetic radiation
43 43,694 (in base 16)
44 A reverse curve
45 Monster: comb. form
46 Ramachandra's wife
47 Anthropologist ___ Hrdlicka (1869–1934)
48 Star: comb. form
49 River in France
50 Type of fermion
52 JFK's predecessor
54 Tee's predecessor
55 List member
58 Current unit
60 ___ ganglia
64 Yellow mist (in China)
66 764,922 (in base 16)
70 English chemist Frederick Augustus ___ (1827–1902)
71 Anglican Church basin
73 A meson
74 Mongoloid or Caucasian, e.g.
75 Recent: comb. form
76 Botanist Katherine ___
77 Reverberation
78 Agriculturist Roswell ___ (1898–1977)
79 Ionizing rad. units

Down

1 ___ particle (emitted electron)
2 Nerve cell arm
3 European capital
4 Dream: comb. form
5 Donkey
6 Radar jammer
7 Unarmed: comb. form
8 ___ exclusion principle
9 Explosive: abbr.
10 German physicist Ernst ___ (1840–1905)
11 Eye part
12 Dihydric alcohol
13 Sporogonium stalk
21 Potential difference: abbr.
23 Common logarithm base
25 60°
27 Tapeworm subclass
28 Norse toast
29 American bird
30 Slow deformation of a solid
32 God of war
34 "___ does it!"
35 Increase
36 Logic circuit elements
37 Units of time
39 For fear that
41 Aesthetic stuff
42 Unit of length
51 Mixture of hydrocarbons
53 ___ agar
56 965,290 (in base 16)
57 Microwave source
59 Written research report
60 Naked
61 43,948 (in base 16)
62 Hyperbolic function
63 One lion
65 Charged particles
67 European wind
68 Bubbly liquid
69 Alimentary canal opening
71 Tuberculosis vaccine: abbr.
72 Basic logic function

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE SEPTEMBER/OCTOBER PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M186

In both cases the answer is no. To see why, color the chessboard horizontally: the first rank black, the second white, the third black again, and so on (fig. 1). We can see that

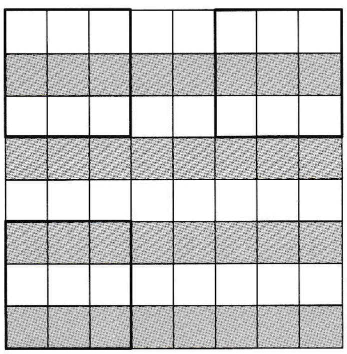


Figure 1

after a jump any checker lands on a square of the same color, so after the rearrangement the 3×3 square must cover the same number of, say, black squares as before. But this number for the bottom square is six, and for either top square it's three.

M187

If we could cut off more than $n/2$ circumscribed quadrilaterals, at least two of them would be consecutive—that is, would have two common sides. Denote them by $ABCD$ and $BCDE$ (fig. 2). If both have inscribed circles, then each of them has equal sums of opposite sides:

$$AB + CD = BC + AD,$$
$$BC + DE = CD + BE.$$

It follows that

$$AB + DE = AD + BE.$$

Since the given n-gon is convex, its

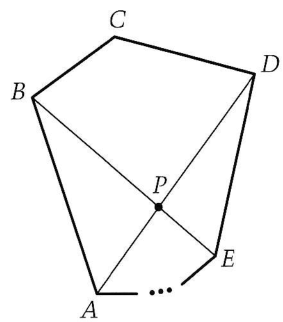


Figure 2

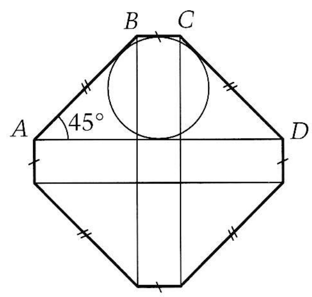


Figure 3

diagonals AD and BE intersect at a certain point P. Then, by the Triangle Inequality we have

$$AD + BE = AP + BP + PD + PE > AB + DE,$$

which contradicts the previous equation.

To construct the required octagon, consider a circumscribed isosceles trapezoid $ABCD$ with base angles of 45° and turn it into a symmetric octagon by adding congruent trapezoids, as shown in figure 3. A similar construction yields an n-gon from which $n/2$ circumscribed quadrilaterals can be cut.

M188

(a) The answer is 12. Figure 4 shows that we can fit 12 non-overlapping ships of the given shape in the 7×7 square, so fewer than 12 shots might leave one of them intact and will be insufficient. Figure 5 shows that 12 shots is always enough.

(b) The answer is 20. Figure 6 shows how 20 squares on the 7×7 board can be marked so that no four of the unmarked squares form a connected tetramino piece. On the other hand, we can place four 3×4 non-overlapping rectangles on the 7×7 board (fig.4). A direct verification, which is left to the reader, shows that the minimum number of shots needed to hit a tetramino ship hiding in a rectangle of this shape is five, so the total number of shots for the entire board that ensures detection of such a ship is no less than $4 \cdot 5 = 20$.

M189

Denote the Fibonacci numbers by f_n:

$$f_0 = 1, f_1 = 1, f_{n+1} = f_n + f_{n-1}. \quad (1)$$

We'll need the following simple estimate for the ratio $r_n = f_{n+1}/f_n$ of the neighboring Fibonacci numbers: starting with $f_3/f_2 = 3/2$, this ratio is no less than $3/2 = 1.5$ and no greater

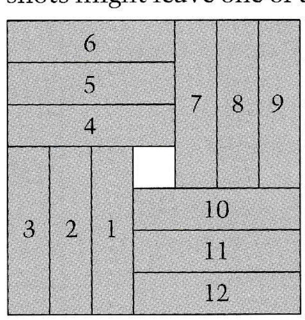


Figure 4

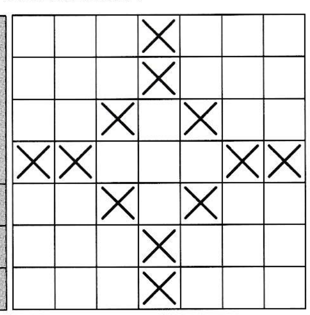


Figure 5

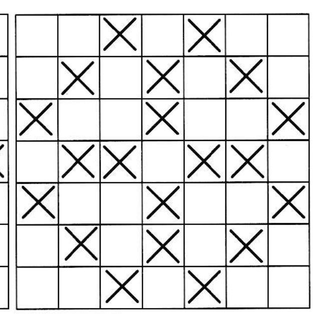


Figure 6

than $5/3 = 1.66\ldots < 1.7$. We can arrive at this estimate from the well-known fact that the ratio of consecutive Fibonacci numbers approaches the "golden ratio" $\tau = (1 + \sqrt{5})/2 \cong 1.618$.

These inequalities can easily be proved by induction. From equation (1) we have

$$r_n = \frac{f_{n+1}}{f_n} = 1 + \frac{f_{n-1}}{f_n} = 1 + \frac{1}{r_{n-1}}.$$

Therefore, if

$$\frac{3}{2} \le r_{n-1} \le \frac{5}{3} \qquad (2)$$

—that is, $2/3 \ge 1/r_{n-1} \ge 3/5$—then r_n is evaluated from above by $1 + 2/3 = 5/3$ and from below by $1 + 3/5 = 8/5 > 3/2$. Since equation (2) is true for $n = 3$, which can be checked directly, it is true for all $n \ge 3$.

Now let f_k be the smallest m-digit Fibonacci number, $m \ge 2$. Then $f_k \ge 10^{m-1}$, $f_{k+1} \ge 1.5 f_k$, and, further,

$$f_{k+2} = f_{k+1} + f_k \ge 2.5 f_k,$$
$$f_{k+3} \ge (2.5 + 1.5) f_k = 4 f_k,$$
$$f_{k+4} \ge (4 + 2.5) f_k = 6.5 f_k,$$

and so $f_{k+5} \ge 10.5 f_k > 10^m$ and has at least $m + 1$ digits. Thus, there are no more than five m-digit Fibonacci numbers.

On the other hand, $f_{k-1} < 10^{m-1}$, $f_k < 1.7 f_{k-1}$, and, following the lines of the argument above, we can show that $f_{k+3} < 7.1 f_{k-1} < 10^m$. Consequently, there are no more than four m-digit Fibonacci numbers. (N. Vasilyev)

M190

Any time we find three points on a circle, two of them can be chosen that define an arc that is at most 120°. So if we join each pair of given points that define an arc that is less than 120° with a segment, we obtain a graph in which *at least two of any three points are joined.* This property suffices to prove the statement of the problem—that is, in terms of the graph, to prove that the graph we constructed has no fewer than 100 edges.

Let A_1 be the vertex of our graph that has the smallest number of edges issuing from it. Denote these edges by

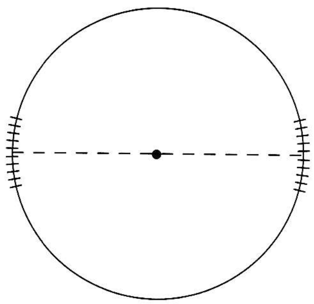


Figure 7

$A_1A_2, A_1A_3, \ldots, A_1A_k$. Each of the k points A_i, $i = 1, \ldots, k$, is an endpoint of at least $k - 1$ edges, so the total number of edges with one or both endpoints among these points is no less than $k(k-1)/2$ (we divide by 2 because each segment can be counted twice, since they each have two endpoints). Any two of the remaining $21 - k$ points must be joined by an edge. Indeed, if two of these points, B and C, are not joined, then there will be no edges at all between the points A_1, B, and C, which contradicts the property of the graph established above. This yields no fewer than $(21-k)(20-k)/2$ additional edges. So the total number of edges is at least

$$\frac{k(k-1)}{2} + \frac{(21-k)(20-k)}{2} = k^2 - 21k + 210 \ge 100$$

(for integer k). The minimum is achieved at $k = 10$ and $k = 11$. This estimate cannot be improved, as is demonstrated by the arrangement of points in figure 7, where 10 points are grouped near one point of the circle, 11 points near the diametrically opposite point.

In the general case of n points on a circle, the number 100 must be replaced with $n(n-2)/4$ for even n and with $(n-1)^2/4$ for odd n. The proof remains the same. (V. Dubrovsky, A. Sidorenko)

Physics

P186

The only force that accelerates the string is that of gravity. So we must find its component acting

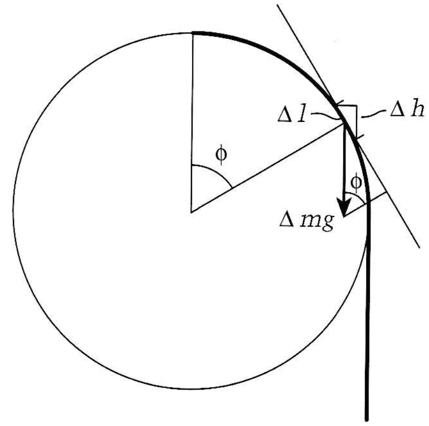


Figure 8

along the spherical surface.

Divide the string into small segments of length Δl (fig. 8). Let the angle between the radius drawn to one such segment from the sphere's center and the vertical be ϕ. Then the component of gravity acting tangentially along the sphere for this segment is

$$\Delta F = \Delta m g \sin\phi = \frac{M}{l} \Delta l g \sin\phi,$$

where M is the string's mass. Figure 8 shows that

$$\Delta l \sin\phi = \Delta h,$$

where Δh is the difference between the heights of this segment's ends. Thus

$$\Delta F = \frac{M}{l} g \Delta h.$$

The total accelerating force we seek is

$$F = \sum_{i=1}^{N} \Delta F_i = \frac{M}{l} g \sum_{i=1}^{N} \Delta h_i = \frac{M}{l} gH,$$

where $H = l - R(\pi/2 - 1)$ when $l \ge \pi R/2$. Therefore, the acceleration is

$$a = \frac{F}{M} = g\left[1 - \frac{R}{l}\left(\frac{\pi}{2} - 1\right)\right].$$

Note that $a \to g$ as $l \to \infty$ as expected. In the case where the string's length equals, say, one quarter of the circumference—that is, $l = \pi R/2$, $H = R$—the initial acceleration of the string is

$$a = \frac{F}{M} = \frac{2}{\pi} g.$$

P187

A charged particle induces a distributed surface charge on the conducting plane, which attracts the particle. The effect of the distributed charge is equivalent to the attraction of an image charge equal to $-Q$ and located at the same distance L from the plane, but on the other side (see figure 9). The force acting on a particle located a distance x from the plane is given by Coulomb's law:

$$F = \frac{kQ^2}{(2x)^2} = \frac{kQ^2}{4x^2}.$$

Rather than trying to solve this equation for the time, let's convert it to a familiar problem that we already know how to solve. Imagine that the same force acts on the particle, but the source of the force is gravitational attraction of a mass M located at point O on the plane. Equating the forces

$$F = \frac{GmM}{x^2} = \frac{kQ^2}{4x^2},$$

we find the size of the mass

$$M = \frac{Fx^2}{Gm} = \frac{kQ^2}{4mG}.$$

Now the motion of the particle can be described by Kepler's third law.

As a preliminary calculation, let's consider a circular orbit of radius L centered on O and find the period of revolution T_0 of the particle around the mass M. We equate the centripetal force $F_c = mv^2/r$ to the gravitational force:

$$F_c = m\frac{4\pi^2}{T_0^2}L = G\frac{mM}{L^2},$$

which yields

$$T_0 = 2\pi\sqrt{\frac{L^3}{GM}}.$$

The trajectory of the particle can be considered a very elongated ellipse with semimajor axis $a = L/2$ and semiminor axis $b \ll a$ (the foci are at the point O on the plane and at the particle's initial position). To find the period T of an elliptical orbit with semimajor axis $a = L/2$, we use Kepler's third law to compare it to the period of the circular orbit with $a_0 = L$:

$$T = T_0\left(\frac{a}{a_0}\right)^{3/2} = T_0\left(\frac{L/2}{L}\right)^{3/2} = \frac{1}{2\sqrt{2}}T_0.$$

It's clear that the time it takes the particle to reach the plane is equal to half the period of revolution:

$$t = \frac{1}{2}T = \frac{1}{4\sqrt{2}}T_0 = \frac{\pi}{2\sqrt{2}}\sqrt{\frac{L^3}{GM}}$$

$$= \frac{\pi}{\sqrt{2}}\frac{L}{Q}\sqrt{\frac{Lm}{k}}.$$

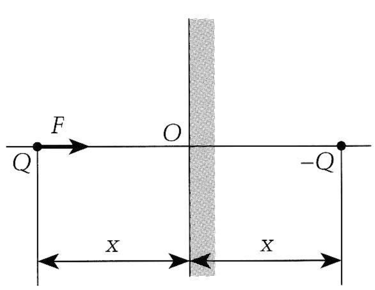


Figure 9

P188

We can neglect both the vapor pressure and the water's volume in the initial state. So the entire volume of the pressure cooker V was initially occupied by air at atmospheric pressure $P_0 = 1$ atm and temperature $T_0 = 293$ K. In the final state the pressure inside the cooker $3P_0$ consists of the air pressure and the pressure of completely vaporized water. Denoting by ρ, v and M the density, initial volume, and molar mass of water, respectively ($\rho = 10^3$ kg/m^3, $M = 18$ g/mole), then for the air pressure P_a and vapor pressure P_v in the final state we have

$$P_a = \frac{P_0 T}{T_0}, \quad P_v = \frac{\rho v R T}{MV}.$$

By the statement of the problem,

$$P_a + P_v = 3P_0.$$

From these formulas we finally get the ratio of the water's volume to that of the pressure cooker:

$$\frac{v}{V} = \frac{P_0 M(3 - T/T_0)}{\rho R T} \cong 10^{-3}.$$

P189

Heat is given off by each side of the plate. The total radiated power is

$$P = a(T_1 - T_0) + a(T_2 - T_0),$$

where a is a proportionality factor. The same power is obtained by the plate from the Sun. For a plate of double thickness we have

$$P' = a(T_3 - T_0) + a(T_4 - T_0).$$

At thermal equilibrium heat is transferred from the illuminated side to the dark side in such a way that the thermal flow is the same at any perpendicular cross section of the plate and equals the heat carried off by the air from the dark side—that is,

$$a(T_2 - T_0) = k\frac{T_1 - T_2}{d},$$

where k is a proportionality factor and d is the plate's thickness. For a plate that is twice as thick, the corresponding equation is

$$a(T_4 - T_0) = k\frac{T_3 - T_4}{2d}.$$

Algebraic manipulations give us

$$T_3 = T_0 + \frac{(T_1 + T_2 - 2T_0)(2T_1 - T_2 - T_0)}{2(T_1 - T_0)},$$

$$T_4 = T_0 + \frac{(T_2 - T_0)(T_1 + T_2 - 2T_0)}{2(T_1 - T_0)}.$$

P190

Consider the path of the ray in the prism (fig. 10 on the next page). At its back side the ray is refracted according to

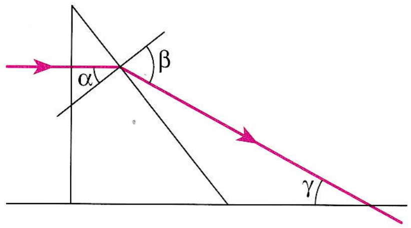


Figure 10

$$\frac{\sin\alpha}{\sin\beta} = \frac{1}{n}.$$

from which we get

$$\sin\beta = n\sin\alpha.$$

Because the angles are very small, we can write

$$\beta \cong n\alpha.$$

It's clear from the geometrical construction that

$$\gamma = \beta - \alpha \cong (n-1)\alpha.$$

The interference pattern can be observed in the overlapping region of the refracted rays leaving both halves of the prism (fig. 11). The

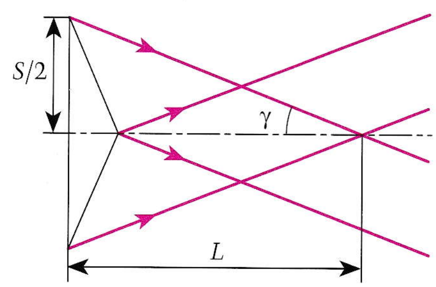


Figure 11

maximum distance where overlapping still occurs is

$$L = \frac{S/2}{\tan\gamma} \cong \frac{S}{2\gamma} \cong \frac{S}{2(n-1)\alpha} \cong 50\text{ m}.$$

Brainteasers

B186

See figure 12.

B187

Yes, the required partition is possible. For instance, consider the 18 pairs of weights "equidistant from

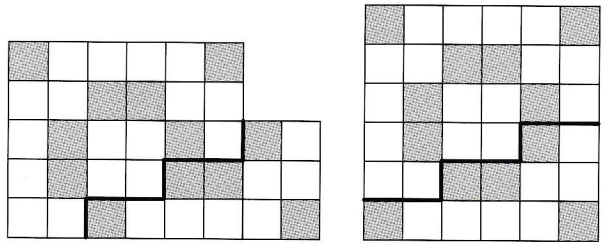


Figure 12

the ends": 1 + 101, 2 + 100, ..., 18 + 84; and 32 similar pairs for the remaining 64 weights: 20 + 83, 21 + 82, 22 + 81, ..., 51 + 52. If we take any 9 pairs from the first set and 16 pairs from the second set, we obtain the required partition.

B188

When you jump onto sand, your velocity is decreased to zero over a longer period of time, which is what we mean when we say that your fall is "cushioned."

B189

See figure 13.

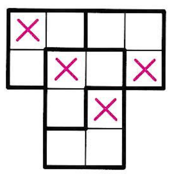


Figure 13

B190

The table is given in figure 14. Clearly 10 games were played, so the total number of points is 10. The winner, player A, lost at least one game and so couldn't win more than 3 points. But neither could A receiver fewer than 3, because otherwise the sum of the scores of all the players would be at most 2.5 + 2 + 1.5 + 1 + 0.5 = 7.5 < 10. Therefore, A lost one game and won all the rest.

	A	*B*	*C*	*D*	*E*
A		0	1	1	1
B	1		1/2	1/2	1/2
C	0	1/2		1	1/2
D	0	1/2	0		1
E	0	1/2	1/2	0	

Figure 14

Since B had no losses, A lost to B and beat the rest of the players. Notice that 3 + 2.5 + 2 + 1.5 + 1 = 10; it follows that the respective scores of B, C, D, and E were 2.5, 2, 1.5, and 1. This is possible only if, besides the win, B had three draws—with all the players except A. Subtracting the draws with B from the results of C, D, and E, we find that their scores in games among themselves are 1.5, 1, and 0.5. Clearly this is possible only with the results in the table.

Toy Store

1. (a) All seven possible strategies are shown in figures 15a–15g. (b) Shifting the 4 × 4 squares of each of the seven strategies in part (a) up and to the right four squares, we obtain a multitude of strategies for the 10 × 10 board. However, only two of them (not including equivalent rotations and reflections) are optimal. They consist of 24 shots each (fig. 16 on the next page—compare this with figures 15a and 15b). The reader may want to show that all optimal strategies for an arbitrary

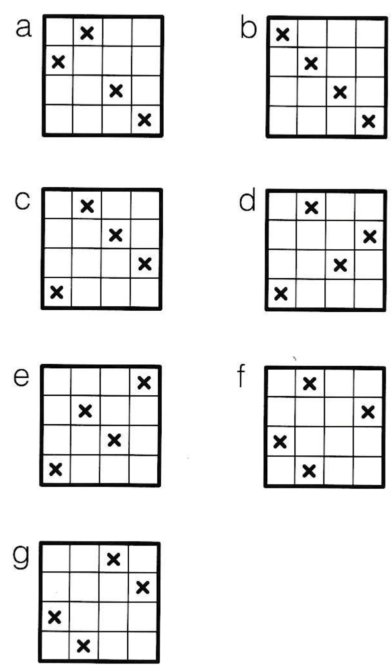


Figure 15

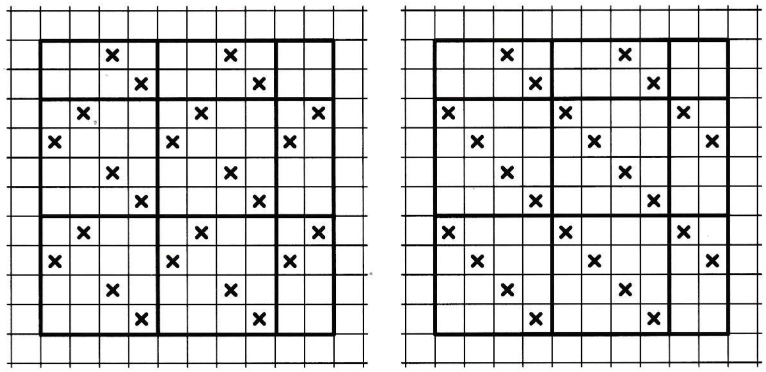


Figure 16

$n \times n$ board are obtained from translations of a certain 4×4 optimal strategy.

2. By the rules of the game, any two ships must be at least one square apart. Surround each ship with a frame 1/2 of a square wide

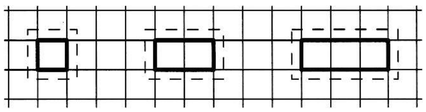


Figure 17

(fig. 17). The rectangle thus obtained will be called the *inflated ship*. It's easy to compute the total area of the seven inflated ships that we have to destroy—it's 36 unit squares. On the other hand, the area of the entire inflated board is the same 36 squares. So the inflated ships cover the inflated board without gaps and overlaps. In particular, it follows that all four corner squares are occupied by ships (otherwise the quarter-square at one of the board's corners would remain unoccupied). Now it's not difficult to list all essentially different (up to rotations and reflections) possible arrangements of the ships. There are only five such positions, as shown in figure 18.

This analysis suggests a remarkable endgame. The first four shots are made at the corners of the 5×5 portion of the board still in play. As we know, all four will find their targets. If any of these shots hits a submarine, that ship will be completely destroyed, and our opponent will have to announce this. Depending on the number of such shots, we'll find ourselves either in the position of figure 18a (three destroyed ships); 18b or 18c (two ships); or 18d or 18e (one ship). In fact, the first case, after the sunken submarines are deleted, will include two positions: the one in the figure and its reflection about the main diagonal. To distinguish between them, we shoot at a3 and c1 and find which of these squares holds the submarine. This will define the position uniquely and we can infallibly complete the "battle." In the second case, after positioning the board so as to have the two destroyed corner submarines along its left side, we'll have to choose between figure 18b, its reflection about the horizontal midline of the board, and figure 18c. This can be done by shooting at a3, c1, and c5. This will tell us which two of these three squares hold submarines and thus determine the position uniquely. Finally, in the third case we actually have four possible positions (provided the only corner submarine is at the bottom left corner): those in figures 18d and 18e and their reflections about the main diagonal. We shoot at a3 and c1. If this destroys only one submarine, we identify the position uniquely, as figure 18d or its reflection, depending on where this submarine is. If both these squares are submarines (fig. 18c), we make one more probe by shooting at b5 and e2 to determine which of the two ships at the top left and bottom right corners is the cruiser. After this—that is, after no more than eight shots—we'll have all the information we need.

This example shows that some positions in the game of Battleships require a high degree of artfulness and self-control.

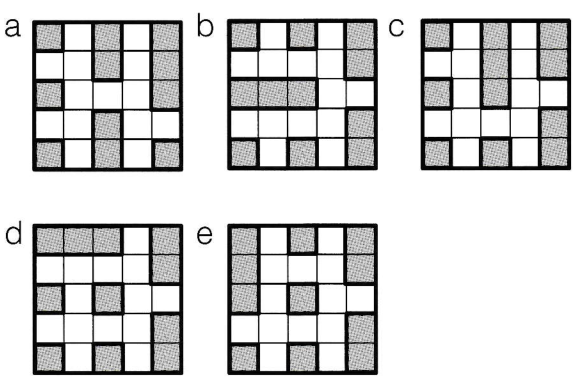


Figure 18

Correction

On page 2 of the last issue, a factor inadvertantly fell out of the equation in column 3, line 4. The expression should read "$M = 4\pi^2 R^3/GT^2$."

Index of Advertisers

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Publication title: Quantum. (2) Publication No. 008-544. (3) Filing Date: 10/1/96. (4) Issue Frequency: Bi-Monthly. (5) No. of Issues Published Annually: 6. (6) Annual Subscription Price: $39.00. (7) Complete Mailing Address of Known Office of Publication: 175 Fifth Avenue, New York, NY 10010-7858. Contact: Joe Kozak. Telephone: (212) 460-1500 ext. 303. (8) Complete Mailing Address of Headquarters or General Business Office of Publisher: 175 Fifth Avenue, New York, NY 10010-7858. (9) Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor: Publisher: Bill G. Aldridge, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000, in cooperation with Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010-7858. Editors: Larry D. Kirkpatrick, Mark E. Saul, Albert L. Stasenko, Vladimir Dubrovsky, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. Managing Editor: Timothy Weber, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. (10) Owner: National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201. (11) Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities: None. (12) The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 Months. (13) Publication Name: Quantum. (14) Issue Date for Circulation Data Below: 9/96. (15) Extent and Nature of Circulation: (a.) Total No. Copies (Net Press Run): Average No. Copies Each Issue During Preceding 12 Months, 10,408; Actual No. Copies of Single Issue Published Nearest to Filing Date, 10,100. (b.) Paid and/or Requested Circulation: (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales: Average No. Copies Each Issue During Preceding 12 Months, 2,330; Actual No. Copies of Single Issue Published Nearest to Filing Date, 2,494. (2) Paid or Requested Mail Subscriptions: Average No. Copies Each Issue During Preceding 12 Months, 6,386; Actual No. Copies of Single Issue Published Nearest to Filing Date, 6,181. (c.) Total Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 8,716; Actual No. Copies of Single Issue Published Nearest to Filing Date, 8,675. (d.) Free Distribution by Mail: Average No. Copies Each Issue During Preceding 12 Months, 270; Actual No. Copies of Single Issue Published Nearest to Filing Date, 254. (e.) Free Distribution Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months, 656; Actual No. Copies of Single Issue Published Nearest to Filing Date, 655. (f.) Total Free Distribution: Average No. Copies Each Issue During Preceding 12 Months, 926; Actual No. Copies of Single Issue Published Nearest to Filing Date, 909. (g.) Total Distribution: Average No. Copies Each Issue During Preceding 12 Months, 9,642; Actual No. Copies of Single Issue Published Nearest to Filing Date, 9,584. (h.) Copies Not Distributed: (1) Office Use, Leftovers, Spoiled: Average No. Copies Each Issue During Preceding 12 Months, 65; Actual No. Copies of Single Issue Published Nearest to Filing Date, 512. (2) Return from News Agents: Average No. Copies Each Issue During Preceding 12 Months, 701; Actual No. Copies of Single Issue Published Nearest to Filing Date, 4. (i.) (Sum of 15g, 15h(1), and 15h(2)): Average No. Copies Each Issue During Preceding 12 Months, 10,408; Actual No. Copies of Single Issue Published Nearest to Filing Date, 10,100. Percent Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 90.40; Actual No. Copies of Single Issue Published Nearest to Filing Date, 90.52. (16) This Statement of Ownership will be printed in the November/December 1996 issue of this publication. I certify that all information furnished on this form is true and complete.

Craig Van Dyck

Craig Van Dyck
Chief Operating Officer

The game of Battleships

Achieving naval superiority on a paper sea

by Yevgeny Gik

IT'S HARD TO IMAGINE A PERSON who has never played the well-known pencil-and-paper game of Battleships. In one version, each of the two players draws two 10 × 10 boards on graph paper. One of them is a "map" of the ocean area where you deploy your "fleet"; the other is used to discover your opponent's deployment. Each of the two fleets consists of ten ships: one 4 × 1 battleship, two 3 × 1 cruisers, three 2 × 1 destroyers, and four 1 × 1 submarines. The ships may occupy any squares on the grid, but they are not allowed to touch one another, not even at the corners.

After the fleets take their initial positions, the battle begins. The players take turns "shooting" at the opponent's vessels—that is, calling out the squares of the board—a3, b7, j9, and so on (the rows are denoted by the numbers 1 to 10, and the columns by the letters a through j, as in chess—see figure 1). After each of your shots, your opponent tells you whether you've hit one of the enemy vessels (if the square you called is occupied by a ship); sunk it (if that was the last untouched square of a ship, and all its other squares have been hit before); or missed (if the square was empty). In the first two cases you're allowed to shoot again, and so on until the first miss—then it's your opponent's turn. Victory is gained by the player who sinks all ten enemy vessels.

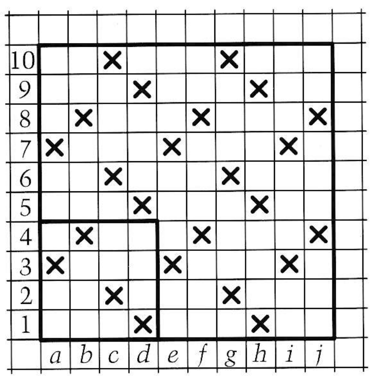


Figure 1

Usually a shot in Battleships is denoted by a dot; if it hits a ship, the dot turns into an X (and a rectangular frame is drawn around the destroyed vessel). Of course, a player will automatically place dots in squares that must be free of ships by the rules of the game—that is, the squares on a diagonal from a hit square, or all the squares adjoining a sunk ship. It goes without saying that the players can change the traditional shape and size of the board, as well as the shapes and number of ships—chess players, for example, might prefer to play on an 8 × 8 board.

Clearly success in Battleships depends, to a certain extent, on sheer luck. You might shoot at the "ocean" at random and destroy all your opponent's ships without a single miss. But it's not very reasonable to rely on chance alone. On the other hand, if you know that your partner has a habit of placing the fleet at the center of the board or perhaps at the edges, your chances of winning increase.

If we talk about the "art" of playing Battleships, two questions arise: (1) How do you shoot so as to raise the probability of hitting an enemy vessel? (2) How do you place your own ships so as to make it more difficult for your opponent to sink them?

Suppose we want to hit the enemy battleship. If we consecutively shoot first at the squares of the first row (from left to right), then at the second row, and so on, it may happen that we'll hit the battleship only on the 97th shot (if the vessel occupies the squares from g10 to j10). However, if we fire only at the squares marked in figure 1, we'll definitely hit the battleship no later than the 24th turn.

It's interesting to consider a more general situation. Suppose a $k \times 1$ ship is hiding on an $n \times n$ board. The sequence of shots that guarantees hitting this ship will be called a *strategy*; the strategy with the least possible number of shots will be called *optimal*.

One of the optimal strategies for

detecting a 4×1 battleship on the 4×4 board is highlighted at the bottom left corner in figure 1 (it consists of four shots). Optimal strategies for the $n \times n$ board are obtained by shifting this strategy four squares up and to the right. In particular, the strategy in the figure is optimal for the 10×10 board. It's clear that to hit a $k \times 1$ vessel on the $n \times n$ board, our shots must be spaced k squares apart in both directions. This means that each row (and each column) must contain approximately n/k shots in the optimal strategy. So the total number of shots is approximately n^2/k; for the battleship, this number is $n^2/4$.

Problem 1. (a) What is the number of optimal strategies for hitting a 4×1 battleship on a 4×4 board? (b) On a 10×10 board? (Strategies that differ only by a reflection or rotation of the entire board are considered the same.)[1]

Here's the modus operandi of experienced players. First, using a strategy similar to the one in figure 1, they detect the enemy battleship. When they've finished with it, they start looking for the cruisers. Shots are fired at intervals of three squares rather than four. After sinking both cruisers, they turn to the destroyers. When only the submarine is left, the unexplored squares are checked at random. Of course, the smaller ships might have been detected earlier, during the hunt for the larger vessels.

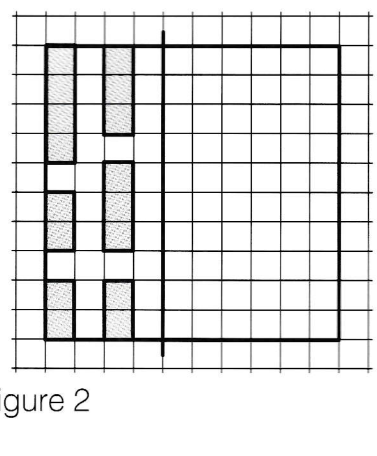


Figure 2

Thus the most difficult task is to sink the submarines. Essentially there's no strategy at all for detecting them. So in deploying their fleets, players must arrange their larger vessels as densely as possible, leaving the maximum amount of free space for the opponent to search for submarines. From this viewpoint, the most advantageous placement is shown in figure 2. Even if the opponent has destroyed all six of the larger ships (to the left of the vertical divider), the submarines must be sought in the largest possible search area (the 60 squares to the right of the divider).

Of course, chance plays a substantial role in the game of Battleships, and it's hard to avoid misses. The most interesting situations are those where a single miss spells the loss of the entire game. Let's consider one endgame of this sort.

It's shown in figure 3. At this point in the game both fleets—ours (fig. 3a) and our opponent's (fig. 3b)—have suffered equal casualties. Our deployment is already known to our opponent, and our fleet faces the danger of a continuous series of strikes that will completely destroy it as soon as it's our opponent's turn. Fortunately, it's our turn to shoot, and the fate of the game is in our hands. In this "mortal combat" we have to destroy one by one all seven enemy ships concentrated in the square a1–e1–e5–a5. The combination that wins this tense battle emerges from the following problem.

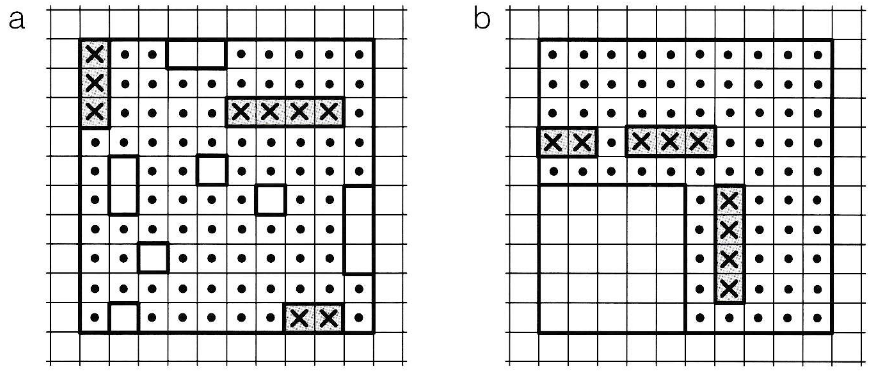


Figure 3

Problem 2. Prove that for any deployment of a cruiser, two destroyers, and four submarines on a 5×5 board, they can be sunk without a single miss.

There are many variations on the game of Battleships.[2] For instance, a move may consist of a number of shots rather than a single shot—the two sides exchange salvos, so to speak. In this version of the game the players inform each other of the general results of each move without specifying which ship was hit and on which square. The rest of the rules are the same. Each move gives a player some information about the deployment of the opponent's fleet that must be somehow processed to derive at the optimal next step.

In another version of the game each player is permitted to take as many shots simultaneously as she or he has ships "afloat" (the first move consists of ten shots at once). Again, the players exchange only general information about the damage done: the number of hits, misses, and sunk ships. When all of a player's ships have been sent to the briny deep, that player loses the right to shoot (that is, has zero shots) and the game as well.

ANSWERS, HINTS & SOLUTIONS ON PAGE 53

[1]Problems 1 and 2 were devised by V. Chvanov. See also challenge M188 in this issue, where you'll find another question about optimal strategies in Battleships, and M169 in the March/April issue, which explores the problem of the feasibility of arranging the ships in the "ocean" *one by one* but in a fixed given order.

[2]In one of them the game is treated as a puzzle (a "one-player game," as it were). A sample of such a puzzle can be found in the *Quantum* article "The World Puzzle Championship" (July/August 1996).—*Ed.*