QUANTUM

MARCH/APRIL 2001

US $7.50/CAN $10.50

Springer

NSTA

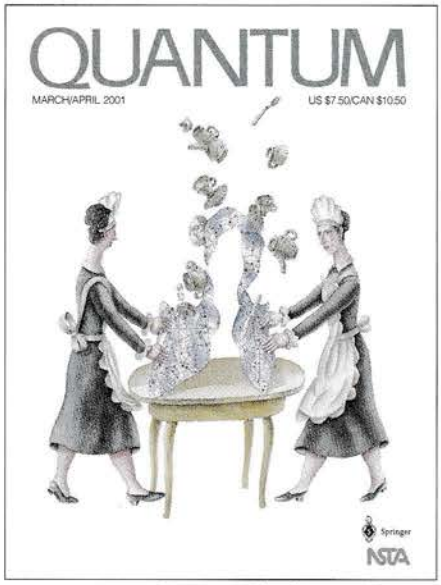
Oil on linen, 21 3/4 × 18, Chester Dale Collection, ©2001 Board of Trustees, National Gallery of Art, Washington, D.C.

Boy on the Rocks (1895/1897) by Henri Rousseau

ANYONE WHO HAS WATCHED A TODDLER STAGger across an open floor, tipping precariously from side to side, has probably wondered how children at this stage of development manage to keep upright despite being perched on such seemingly unsteady legs. Similar extraordinary balancing acts can be seen in nature, especially among rock formations where water and wind erosion has carved the landscape. Often you will find teeter-totter rock outcroppings that are so finely balanced that they sway in the breeze. To learn more about the physics behind these massive monuments of stability, turn to page 18.

QUANTUM

MARCH/APRIL 2001 VOLUME 11, NUMBER 4



Cover art by Sergey Ivanov

Just as shaking out a tablecloth will give you an indication of what was served for dinner the night before, vibrations created by gravitational waves provide scientists with clues about recent events in the Universe. Currently, scientists are devising ways to detect and decipher these waves to gain insight into back holes, the collisions of neuron stars, and other cosmic events millions of light years away. Turn to page 4 to learn more about how we'll be catching these waves of the future.

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

4 Gravitational Waves
Ripples on a cosmic sea
by Shane L. Larson

10 Geometric Gymnastics
Taking on triangles
by A. Kanel and A. Kovaldzhi

18 Rocking Cliffs
Rock 'n' no roll
by A. Mitrofanov

24 Out of Sight?
The near and far of it
by A. Stasenko

DEPARTMENTS

3 Brainteasers

17 How Do You Figure?

22 At the Blackboard I
Brocard points

28 Kaleidoscope
Many ways to multiply

30 Physics Contest
The fundamental particles

34 Looking back
From the pages of history

38 Problem primer
Exploring every angle

40 In the open air
Flights of fancy?

42 In the lab
Using cents to sense surface tension

44 Forces of nature
Lunar launch pad

48 At the Blackboard II
Electrical and mechanical oscillations

50 Answers, Hints & Solutions

54 Crisscross Science

55 Informatics
Breakfast of champions

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 11 (6 issues) will be published in 2000–2001. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; e-mail: quantum@nsta.org. **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 offices. **Postmaster:** send address changes to: *Quantum*, Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485. Printed in U.S.A.

Subscription Information:
North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $56. 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. Mail order and payment to: Springer-Verlag New York, Inc., Journal Fulfillment Services Department, PO Box 2485, Secaucus, NJ 07094-2485, USA. Telephone: 1(800) SPRINGER; fax: (201) 348-4505; e-mail: custserv@springer-ny.com.

Outside North America: Personal rate: Please contact Springer-Verlag Berlin at subscriptions@springer.de. Institutional rate is US$57; airmail delivery is US$18 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.

Advertising:
Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) M.J. Mrvica Associates, Inc., 2 WestTaunton Avenue, Berlin, NJ 08009. Telephone (856)768-9360, fax (856)753-0064; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Visit us on the internet at www.nsta.org/quantum/.

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 mission of the National Science Teachers Association is to promote excellence and innovation in science teaching and learning for all.

Just for the fun of it!

B316

On call 24-7. The working hours of a receptionist at a hotel are either 8 A.M. to 8 P.M., 8 P.M. to 8 A.M., or 8 A.M. to 8 A.M. the next day. In the first case, the break before the next shift must be not less than 24 hours; in the second case, not less than 36 hours; and in the third case, not less than 60 hours. What is the smallest number of receptionists that can provide round-the-clock operation of the hotel?

B317

Sweet primes. A bag contains 101 pieces of candy. Eric and Karlson play a game. In rotation (first Eric and then Karlson) they take from 1 to 10 pieces from the bag. When the bag becomes empty, they count the number of pieces they've taken from the bag; if these numbers are prime relative to each other, Eric is the winner; otherwise, the winner is Karlson. Who should win in this game and what must his strategy be?

B318

Ducky numbers. For many years now Baron Münchhausen has gone to a lake every day to hunt ducks. Starting on August 1, 2000, he says to his cook: "Today I shot more ducks than two days ago, but fewer than a week ago." For how many days can the baron say this? (Remember, the baron never lies.)

B319

Bench mates. Several chess players played chess in a park the whole day long. Since they had only one set of pieces, they chose the following rules: The winner of a game skips the next two games, and the loser skips the next four. How many players took part in the tournament if they managed to follow these rules? (If the game ended in a draw, the player who played white was considered the loser.)

B320

Curls when wet. Why do waves curl up on top as they approach the shore?

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

Art by Pavel Chernusky

Ripples on a cosmic sea

Cutting-edge astronomy is making waves

by Shane L. Larson

HIGH ON THE COLUMBIA plateau of eastern Washington state, a remarkable astrophysical observatory is being constructed; its twin is taking shape in the lush forests of central Louisiana. If you're peering through the dust and the tumbleweeds (or the thick mossy forest), don't expect to see any domes housing massive optical assemblies or great radio dishes tracking across the skies. These observatories are not looking for the visible light from the countless burning stars throughout the Universe, nor for the faint radio whispers of charged particles thrashing about in their hot and violent environments. These are a phenomenal new kind of observatory called LIGO: the Laser Interferometer Gravitational-wave Observatory.

Gravitational waves were one of the novel predictions of Einstein's 1916 general theory of relativity, a completely new phenomenon that was not present in the Newtonian theory of gravity prevalent up to that time. It is only now, almost a century later, that technology has become sophisticated enough to possibly detect this new radiation, and observatories like LIGO are slowly taking shape across the planet.

The modern gravitational wave observatory is a large laser interferometer (typically with arms about 0.5 to 4 km long, for current designs). They are very similar to the familiar Michelson interferometer, but on a much larger and grander scale. Scientists will carefully monitor the output of the interferometers, looking for miniscule changes in the lengths of the interferometer arms, indicating the passage of a gravitational wave.

In addition to ground-based observations, scientists at NASA and the European Space Agency are also beginning to think about the search for gravitational waves in space. They are designing a much larger interferometer, known as LISA (Laser Interferometer Space Antenna), to be launched sometime late in the next decade. The mission will consist of three spacecraft arranged in a triangular constellation, 5 million kilometers per side. These three spacecraft will orbit the Sun in a triangular configuration, just over 52 million kilometers behind the Earth in its orbit, and inclined to the Earth's orbit by 60 degrees (figure 1). By monitoring laser signals exchanged between each of the spacecraft, scientists can monitor any change in distance between the craft in an effort to detect gravitational waves.

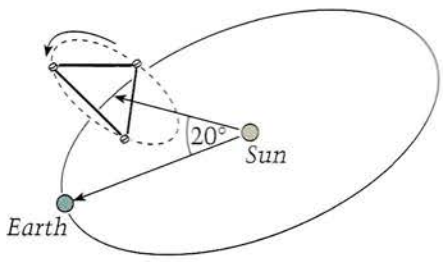


Figure 1

But what *are* gravitational waves? Why didn't we know about them before Einstein and why are they hard to detect? To understand this, we must explore the differences between Isaac Newton's theory of gravity and Albert Einstein's theory of general relativity.

Gravity according to Isaac

When Isaac Newton sat beneath the proverbial apple tree waiting for his fruitful concussion, his perception of the cosmos was built around the idea that space and time were immutable qualities of the Universe. From his perception (and indeed, from the perception of essentially all experimental evidence available at the time), space and time were fixed, absolute entities throughout the Universe.

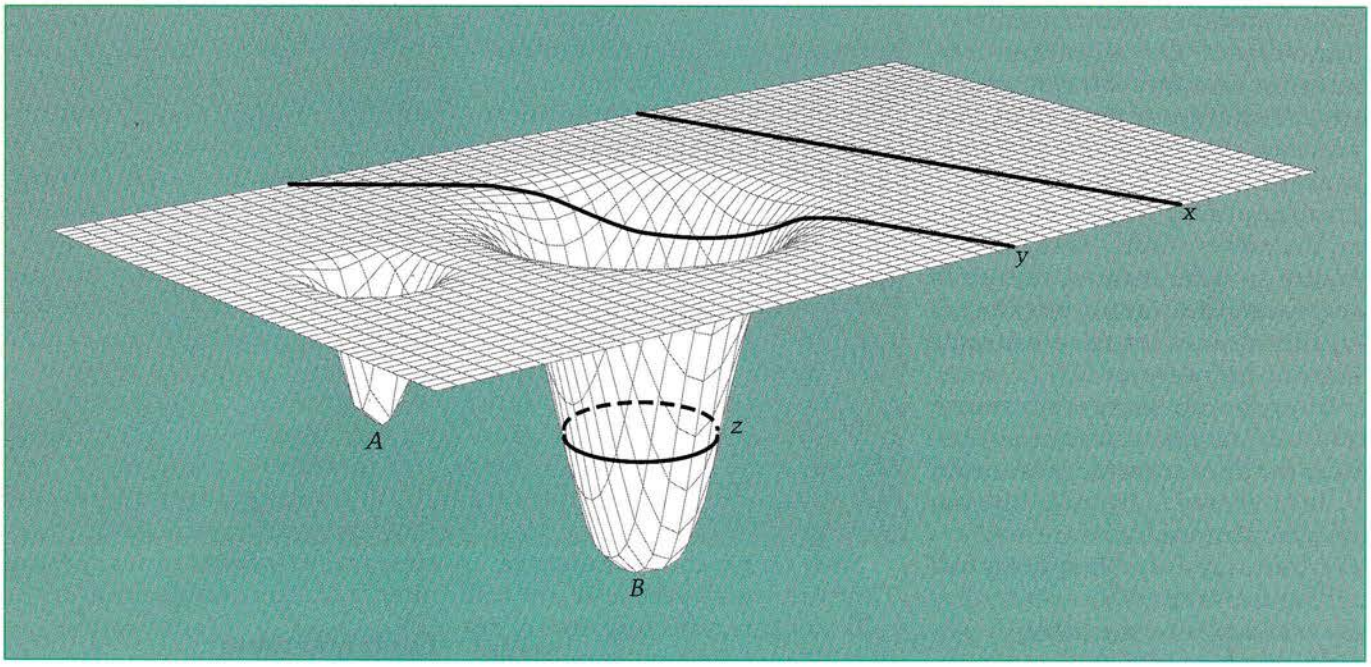


Figure 2

Newton set down his ideas about absolute space, time, and motion in 1687 in his monumental work, *Philosophiae Naturalis Principia Mathematica*, or "The Mathematical Principles of Natural Philosophy" (the *Principia* also covers Newton's ideas of *relative* space and time, which are defined with respect to human constructs; for example, time is measured in hours and minutes, and space is broken into distances or orientations with respect to Earth).

The idea that space and time were immutable, *universal* quantities allowed Newton to propose his very successful universal law of gravitation, which is familiar to us all:

$$F = \frac{Gm_1m_2}{r^2}. \qquad (1)$$

The universality of space and time allowed the application of equation (1) to any gravitational system that could be observed, including the Moon, the planets, and even the newly discovered Galilean satellites of Jupiter (Galileo had detected the four largest satellites of Jupiter with his telescope in 1610), despite the fact that these systems were far removed from the time and space that could be sampled directly in any Earth-bound laboratory.

The theory was deceptively simple in its formulation and unprecedented in its predictive power. In particular, it can be used to construct (and indeed was developed to explain) Kepler's laws of planetary motion; this can be easily demonstrated for the special case of circular orbits if one assumes equation (1) to be the centripetal force that binds a mass (for example, the mass of an asteroid, or other small test mass) in its orbit about the Sun.

An important aspect of Newtonian gravity is that the gravitational force acts instantaneously (that is, changes in the gravitational field propagate at infinite speed). If the mass of the Sun were to suddenly change, or if it were to start moving away from its location at the center of the Solar System, each of the planets would know instantly and their orbits would change in accordance with the new configuration of the gravitational field.

Gravity according to Albert

The Newtonian theory of gravity was a cornerstone of physics for more than 200 years, and even today it is extraordinarily useful and particularly well suited for many different applications, such as computing spacecraft trajectories or describing the orbits of binary stars. But when Einstein published his special theory of relativity in 1905, it immediately began to pose problems for the Newtonian theory of gravity.

Special relativity introduced many wondrous and strange predictions about the relationships between moving clocks and rulers, but one of its most important predictions is the existence of a universal speed limit: $c = 3.0 \times 10^8$ m/s. Nothing can travel faster than the speed of light. In contrast, Newton's theory of gravity allows infinite propagation speeds, clearly a violation of the much slower speed limit c.

Einstein set out to formulate a theory of gravity consistent with special relativity, and in 1916 published his *general theory of relativity*. General relativity breaks with Newtonian gravity from the outset by discarding the idea of the gravitational field in favor of a new concept: space-time geometry. Einstein's basic premise was that the motions of particles were not affected by an unseen force tugging on them, pulling them toward massive bodies. The motions of particles are determined by the geometry of the space-time around them.

A visual analogue of Einstein's remarkable idea may be seen in figure 2, which illustrates the "rubber-

sheet" model of general relativity. Space-time is "flat" when there is no mass present, as shown at the extreme right edge of the surface. The presence of a massive body curves space-time, as shown at point *A* (imagine placing a small lead weight on the rubber sheet). More massive bodies produce more curvature in space-time, deforming it more than smaller masses (imagine placing a bowling ball at point *B*).

How does the shape of space-time affect the motion of particles? Consider the three trajectories labeled *x*, *y*, and *z* shown in figure 2. You can imagine that each of these paths is the trajectory of a Ping-Pong ball rolling across the rubber sheet; they are analogous to the paths of particles (such as satellites, asteroids, and comets) in the vicinity of massive bodies. The path *x* represents the path of a particle through space-time when it is far from any mass. Such a path is called a *geodesic*. In this rubber-sheet model, geodesics are the shortest length paths between any two points. In the flat regions of the sheet, the geodesics are familiar straight-line paths.

There are other "straight-line" paths in space-time, such as the trajectory *y*. Imagine a ball rolling along *y*, which is initially parallel to *x*. When the ball encounters the curved region of space-time, the geodesic path dips into the curved region, and reemerges along a new direction that is diverging from *x*. This path is also a geodesic because it is the straightest trajectory for the ball through the region of high curvature. No external forces acted on the ball to alter its trajectory. The trajectory was altered only by virtue of the fact that the ball rolled on a curved surface.

Now think about the curve *z*. This path is also a geodesic; the ball rolls along its trajectory, free of external forces pushing or pulling on it, its course determined only by the curvature of the space around it.

Each of the three paths in figure 2 is analogous to familiar particle trajectories described in terms of a central potential,

$$V = -\frac{GM}{r}, \qquad (2)$$

where *M* is the mass of the central source of the potential. The path *x* is that followed by a particle far from any source of gravitational attraction, the path *y* is that of a particle scattering off a gravitational potential, and the path *z* is that of a particle in orbit about a larger mass.

This way of thinking about general relativity can be summarized in the two-line mantra of the modern gravitational theorist, popularized by Misner, Thorne and Wheeler in their classic text *Gravitation*: "Matter tells space how to curve; space tells matter how to move."

The idea that the "gravitational field" is simply curvature of space-time will be integral to our physical picture of a gravitational wave.

Gravitational waves

The existence of a cosmic speed limit is at great odds with Newtonian gravity, which allowed signals to propagate at infinite speed. By imposing the constraints of special relativity on a theory of gravity, we suddenly find a myriad of new phenomena we can experimentally search for in nature—phenomena that we did not know existed be-

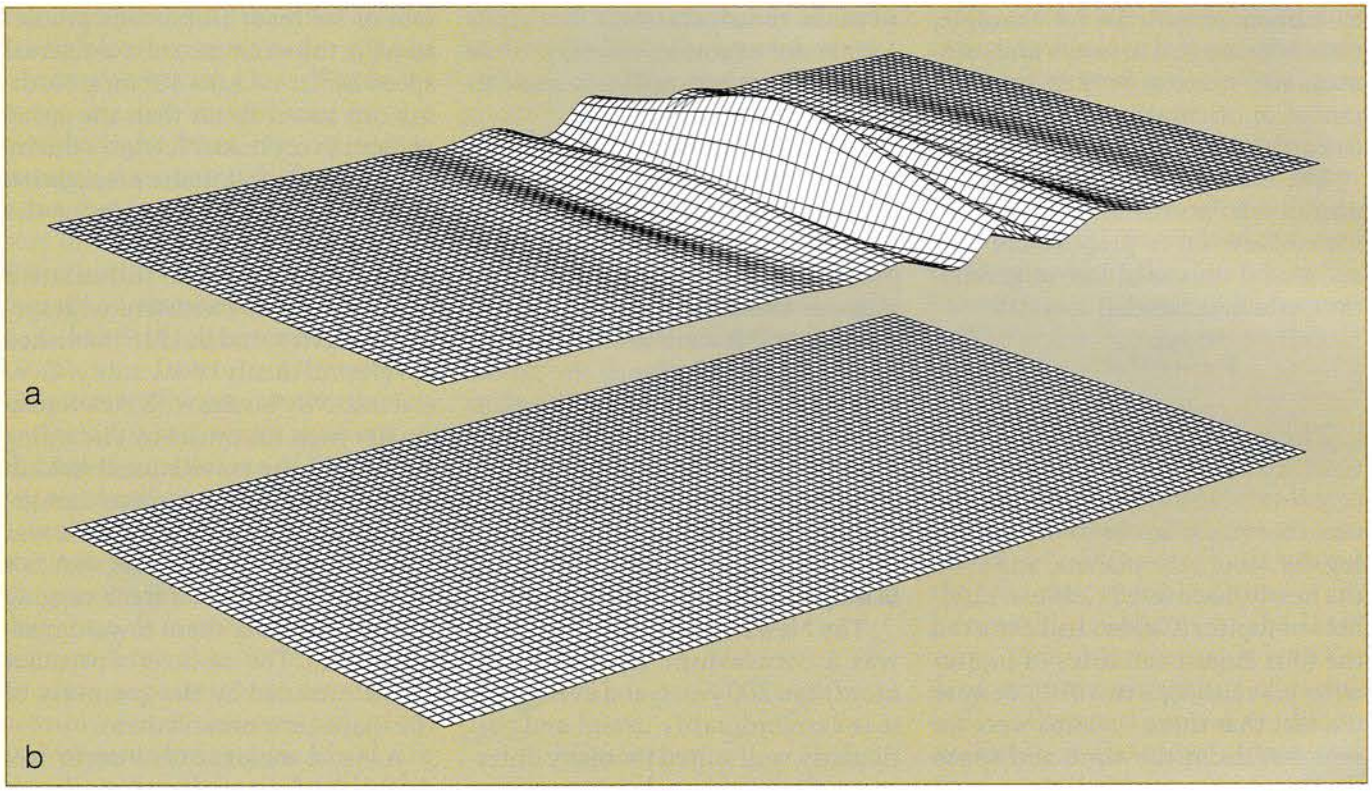


Figure 3

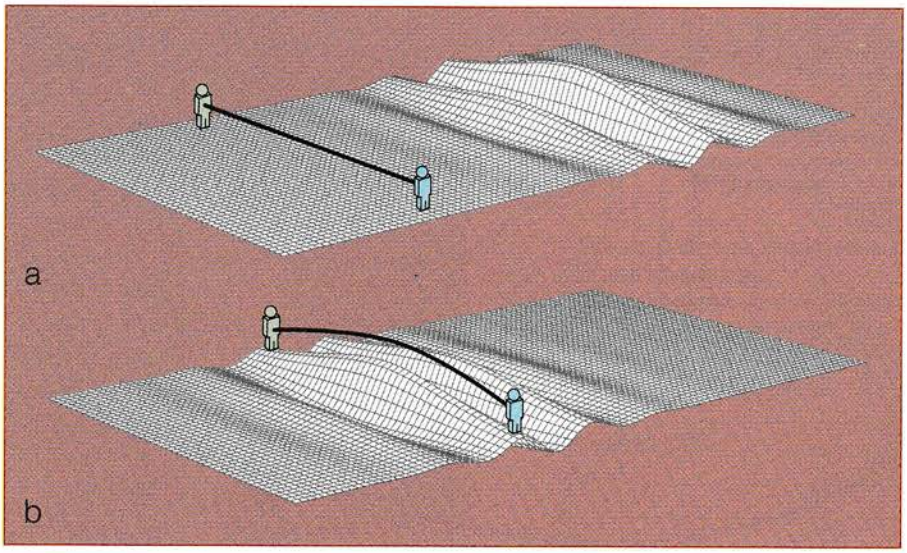


Figure 4

cause they simply cannot be explained with Newtonian physics. One example is the famous "bending of light" by a gravitational field, which Einstein put forth as a test of his new theory of gravity. Eddington's measurement of the deflection of starlight by the Sun during the total eclipse of 1919 confirmed the predictions of general relativity and made Einstein a worldwide celebrity.

To understand how waves are treated by modern relativistic theory, recall our Einsteinian description of gravity as curvature of space-time. If the analogue to the gravitational field is curvature, then changes in the gravitational field are analogous to changes in the curvature of space-time. When changes in curvature propagate, moving through space-time, they are called gravitational waves. Figure 3 shows a model of gravitational waves in the context of the rubber-sheet analogue outlined in the section above.

Like the waves we are more familiar with, gravitational waves have an amplitude (usually denoted h), a wavelength λ, and a frequency f, which are related to the propagation speed c:

$$c = \lambda f. \qquad (3)$$

It is no accident that the propagation speed is written as c; general relativity predicts that gravitational waves travel at the speed of light.

Detection

How does one go about detecting gravitational waves? To do this, you must develop a way to measure the changes in the space-time curvature. We can imagine a detector of gravitational waves in the rubber-sheet model we developed above. Consider the two people in figure 4a. They are hanging out in an essentially flat space-time, shining a flashlight back and forth at each other, and timing how long it takes the beam to traverse the distance between them. This time is a measure of the *proper distance* between them. Unbeknownst to them, a curvature wave is approaching, and it will affect the results of their experiment.

In figure 4b, the wave is upon our space-time experimenters. Because the wave has changed the curvature of the space-time between them, it takes a different amount of time for the photons to travel back and forth; our intrepid young experimenters can measure this time difference, thus detecting the wave!

Interferometers detect gravitational waves in much the same way, by comparing the distance along two different directions in spacetime. Using laser light, the beams in two different directions are interfered with each other. When a gravitational wave passes by, the lengths of the interferometer arms change, and so the interference pattern made by the two laser beams shifts.

The quantity measured in a gravitational wave observatory is called the *strain* and is defined as

$$s = \frac{\Delta l}{l}, \qquad (4)$$

where Δl is the change in proper length the gravitational wave produces between our two experimenters and l is the unperturbed length, before the wave is upon them. The strain can be approximately related to the amplitude of the wave by $s \sim h/2$.

Laser interferometers aren't the only way to detect gravitational waves. One could imagine that the two experimenters in figure 4 aren't shining a flashlight back and forth, but rather are holding a long metal bar between them. When the gravitational wave passes by, it stretches the bar a little. The bar snaps back to its original shape after the wave passes by, and as a consequence begins to "ring" (that is, it begins to vibrate). The frequencies that the bar can see depend on its length. Roughly speaking, the bar is sensitive to waves that have a frequency corresponding to its normal modes of vibration:

$$f = n\frac{v}{2l}, \qquad (5)$$

where v is the speed of sound in the bar, l is the length of the bar, and n is an integer indicating the mode. The amplitude of these oscillations depends on the strain induced in the bar by the gravitational wave.

These types of detectors are called "bar-detectors" and were first pioneered by Joseph Weber at the University of Maryland in the 1960s. The most sensitive bar-detector in operation today is called ALLEGRO, and is operated by Louisiana State University.

Sources

Gravitational waves are created from the dynamical motions of mass. Any asymmetric acceleration in a massive system will generate gravitational waves (to be precise, a

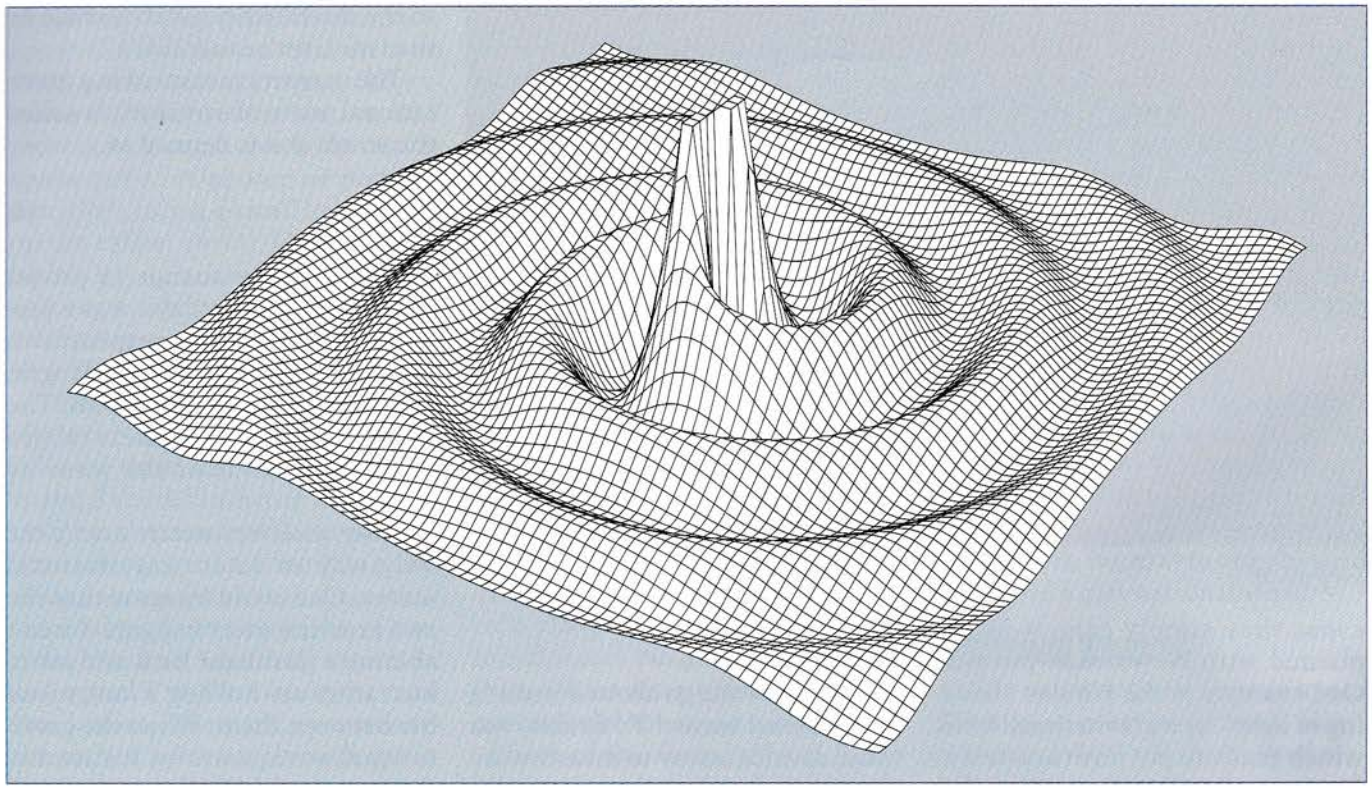


Figure 5

system will emit gravitational radiation if it has a non-zero *quadrupole moment*). Symmetric motions, such as radial pulsations in spherical stars, will not generate gravitational waves.

Purely symmetric systems are, of course, an ideal in physics and very unlikely to exist in nature. A survey of common astrophysical systems shows that the cosmos is replete with asymmetric dynamical systems, from active galactic nuclei and spiral galaxies on the largest scales to supernovae and common binary stars on smaller scales.

Binary systems in general are expected to be one of the most important sources of continuous astrophysical gravitational waves. Interesting targets for study include binary stars, neutron-star binaries, black hole binaries, or combinations of any of the three. Throughout most of their lives, binary systems evolve slowly. If they are in approximately circular orbits, gravitational waves are emitted with a frequency that is twice the orbital frequency:

$$f = 2f_{\text{orb}}. \qquad (6)$$

These waves are said to be *monochromatic* in analogy with single frequency visible light. Figure 5 shows the amplitude of the gravitational waves generated by a typical binary system. The stars generating the radiation lie at the center of the figure. Gravitational waves are thrown off as a result of their orbital motion, and propagate out through the vast sea of space-time until they come gently lapping up on the shores of Earth.

The (dimensionless) amplitude of gravitational waves radiated by a binary system, as measured at the Earth, can be estimated by the formula

$$|h| = \frac{4G^2 m_1 m_2}{c^4 aR}, \qquad (7)$$

where m_1 and m_2 are the masses of the binary components, G is Newton's constant, a is the semi-major axis of the binary orbit, R is the distance from the Earth to the binary, and c is the speed of light. By rewriting a using Kepler's third law

$$a^3 = \tau^2 \frac{G(m_1 + m_2)}{4\pi^2}, \qquad (8)$$

equation (7) may be expressed in terms of the orbital frequency (which is related to the gravitational wave frequency by equation (6)):

$$|h| = \frac{2\pi^{2/3}(2G)^{5/3}}{c^4 R} \frac{m_1 m_2}{(m_1 + m_2)^{1/3}} f_{orb}^{2/3}. \qquad (9)$$

Indirect, astrophysical evidence for the existence of gravitational waves exists from monitoring the famous Hulse-Taylor binary pulsar, PSR 1913+16, over the past 25 years. By monitoring the pulsar's orbit, it was discovered that the orbital period was gradually shrinking, implying the two pulsars are slowly spiraling together. The rate at which the orbital period is changing is *precisely* the amount that general relativity predicts through the emission of gravitational radiation. Gravitational waves from the orbital motion of the pulsars carry away orbital energy, causing the orbit to shrink. This discovery earned Joseph Taylor and Russell Hulse the 1993 Nobel Prize in physics.

We can estimate the amplitude of the gravitational waves from this pulsar when they arrive at the Earth

using equation (9). This binary has two stars, each of mass $m \sim 1.4M_{sun}$, an orbital period of about 7.75 hours (implying $f_{orb} \sim 3.58 \times 10^{-5}$ Hz), and lies at a distance of $r \sim 5$ kpc from Earth. This gives a dimensionless amplitude of $|h| = 6.4 \times 10^{-23}$, a very small amplitude indeed! This is many orders of magnitude lower than the minimum amplitude detectable by LISA at this frequency.

Is it possible for us to detect sources of gravitational radiation much closer to the Earth? For instance, binaries are expected to be good sources of gravitational waves; might we expect to see gravitational radiation from the orbits of the Galilean satellites (Io, Ganymede, Callisto, and Europa) about Jupiter? Consider equation (9) again, and let's apply it to the case of Io, the innermost of the Galilean satellites. Jupiter has a mass of 1.90×10^{27} kg, and Io has a mass of 8.94×10^{22} kg. The orbital period of Io is 1.77 days, giving an orbital frequency of $f_{orb} = 6.54 \times 10^{-6}$ Hz. At their closest approach, Jupiter and the Earth are 6.29×10^{8} km apart. Using equation (9) with these values gives an amplitude of $|h| = 1.4 \times 10^{-24}$. Despite being significantly closer to Earth than the binary pulsar, the gravitational radiation from Io is much weaker. It is not expected that detectors such as LIGO or LISA will detect any gravitational radiation from any source within our own Solar System.

What we hope to learn

Gravitational waves are a completely new way of looking at the Universe. When the first radio telescopes were built, we learned a tremendous amount about distant astrophysical systems because radio waves bring us different information than ordinary light. Similarly, we hope that by observing the Universe in gravitational waves, we should learn different things than we would by looking in ordinary light. In particular, we should be able to observe the collisions of massive black holes, the collisions of neutron stars, stars falling into the black holes at the centers of galaxies, and supernova explosions.

The most prevalent type of source will be close binaries such as those described above. Early on, close binary systems will have relatively small orbital frequencies. Frequencies in the range of roughly 10^{-5} to 10^{-1} Hz should be accessible to LISA. Late in their lives, binary systems tend to evolve rapidly, the components spiraling toward one another. As they spiral together, the frequency ramps up rapidly and the binary "chirps." Ultimately, these binary systems coalesce to form a single object. This high frequency inspiral, chirp, and coalescence should be observable by LIGO at frequencies from about 10 to 1000 Hz.

The process of coalescence will be a dynamic and violent one, and scientists expect it to produce copious amounts of gravitational radiation. By studying these gravitational waves, it is hoped we will gain our first direct observations of what happens during the collision of two massive bodies and how the final object wobbles, stretches, and vibrates before settling down into its final state. Short, transient pulses of gravitational radiation, such as supernovae, are known as *burst sources*. Predicting what bursts of radiation from violent events might look like using a gravitational wave observatory such as LIGO or LISA is a problem that is at the forefront of modern theoretical physics, and is being studied using advanced numerical simulations on fast-computing systems. Whether or not we will be able to detect burst sources (they are much harder to detect than inspiraling binaries) will depend on precisely how strong the burst of radiation from an explosive event is, and how far away from Earth it is.

The future

Because of the weak nature of gravitational waves, it is possible that our initial searches with LIGO will not detect any gravitational radiation. This is largely due to limitations in technology, but as computing power, laser technology, and our understanding of gravitational waves improves, we'll be able to build better gravitational wave observatories. Plans to upgrade LIGO to LIGO II are already in place, and should make the detection of gravitational waves a routine occurrence.

Space-based observatories such as LISA are literally guaranteed to see nearby interacting white dwarf binary stars. The closest of these, a star called AM CVn, is a helium cataclysmic variable about 100 parsecs away in the constellation of Canes Venatici, and can be seen in small telescopes. Since stars like AM CVn can be observed with ordinary telescopes, we know a tremendous amount about the masses and orbits of these binary systems. Therefore, we know what the gravitational wave signal should look like, and should be able to detect such stars almost immediately after LISA becomes operational.

The study of gravitational waves from the Universe at large promises to produce a revolution in astrophysics as spectacular as the revolution brought on by the advent of X-ray, radio, and γ-ray astronomy. Unlike photons, gravitational waves propagate very readily through regions of dense gas and dust. Gravitational waves will be generated by the mysterious "dark matter" that seems to pervade much of the cosmos, and should have been generated in the earliest moments after the Big Bang. By studying this remarkable new type of radiation, astrophysicists will, for the first time, be able to probe the dense cores of galaxies, see the inspiral and collision of neutron stars millions of light years away, and study the region very near the event horizons of black holes. Gravitational wave astronomy promises to be one of the hottest areas of research as we move into the 21st century. ◙

Shane L. Larson *is a NASA EPSCoR postdoctoral research associate at Montana State University and the Jet Propulsion Laboratory, where he works on issues related to sources of gravitational radiation and the design of space-based observatories like LISA.*

Taking on triangles

In search of answers between the lines

by A. Kanel and A. Kovaldzhi

HERE'S A PROBLEM THAT was published in our sister magazine *Kvant* in the early seventies:

Given are n straight lines in general position in a plane (that is, no three of them pass through the same point and no two are parallel to each other). They divide the plane into several parts. Prove that among these parts there are at least (a) n/3 triangles, (b) (n – 1)/2 triangles, and (c) n – 2 triangles.

We want to come up with the exact number of triangles. This problem was stated as early as 1870; its statement is simple and appealing. Yet it's so difficult that it had remained unsolved for over a hundred years. It draws you in; you think you're on the verge of a solution. But every "Eureka!" just adds to the list of subtle errors.

The first solution was obtained in 1979 by the well-known mathematicians Grünbaum and Sheppard. In this article, we present a shorter, elementary solution found by A. Kanel. We could actually present it in three sentences, but such a barebones approach would contribute little to the reader's understanding of the essence of the problem. For this reason, we try to trace the steps leading to the solution and uncover the key in the idea. Sometimes we will take a side trip, solving a completely different problem with a technique that will unlock the original one.

Simplified statement

The following problem was proposed at the Moscow Mathematical Olympiad in 1972.

Problem 1. *We are given* 3,000 *straight lines in the plane such that no two of them are parallel and no three meet in a point. The plane is cut into pieces along these lines. Prove that among these pieces there are at least (a)* 1,000 *triangles, and (b)* 2,000 *triangles.*

Solution. It's clear that item (a) of this problem coincides with item (a) of the problem formulated at the beginning of the article for n = 3,000; and item (b) is a strengthening of item (b) of that problem, since 2,000 > (3,000 – 1)/2.

Let's begin with a key idea. We will prove that for every line, there exists at least one triangular piece adjacent to it (that is, a triangle with a side on this line). If we manage to prove this fact, this will prove item (a), since each triangle is adjacent to three lines.

For each line, let's try to find a triangle adjacent to it that is not intersected by other lines. Here we use an elegant idea: Consider the point of intersection of other lines that is the nearest to the line under consideration (figure 1). We leave it to the reader to prove that this point can be chosen as a vertex of the desired triangle.

To prove item (b), it's sufficient to find, for each line, one more triangle adjacent to it. This seems simple; indeed, every line divides the plane into two half-planes. Perhaps we can choose one triangle on each half-plane to correspond to each line. Unfortunately, if we choose a certain line, it may happen that all the other lines will intersect on only one

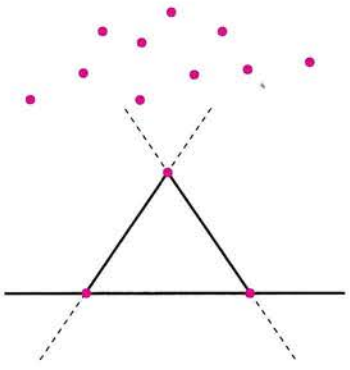


Figure 1. *The nearest intersection point.*

Art by Ekaterina Silina

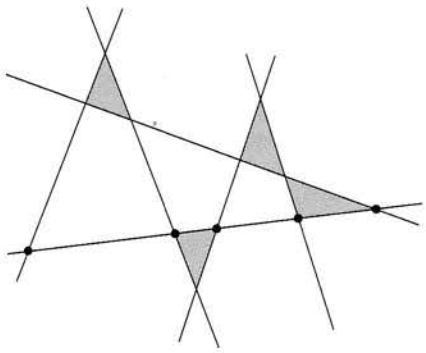


Figure 2. *Every triangle leaves a smaller triangle when intersected by a line.*

of its half-planes. How often can such a "bad" case occur? Let's make an estimate.

It's clear that there can be three bad lines. Fortunately, this is the worst case (when the total number of lines is three). If $n > 3$, it turns out that at most two lines are bad. Here is a proof of this fact.

Assume that there are three bad lines. Draw these three lines and one more arbitrary line. In any case we examine, new intersection points appear on both sides of one of the bad lines. Indeed, the fourth line intersects all three given lines, and one of the three intersection points lies between two others. We have arrived at a contradiction, and this proves our assertion.

Thus, for $n > 3$, there exist not more than two lines adjacent to only a single triangle: All the other lines are adjacent to at least two triangles. Now we can obtain an estimate of the number of triangles. It's clear that this number is not less than $(2n - 2)/3$. For $n = 3{,}000$, this yields $1{,}999\ ^1/_3$. But the number of triangles is an integer. Thus, there are at least 2,000 triangles.

At the same time we have proved a stronger assertion than item (b) of the original problem, since $2(n - 1)/3 > (n - 1)/2$.

Here are several exercises related to the problem just solved.

Exercises

1. We are given n planes in general position, in space (that is, any four of them form a tetrahedron). The planes divide the space into several parts. Prove that among these parts there are at least (a) $n/4$ tetrahedrons (for $n \geq 4$), (b) $(2n - 3)/4$ tetrahedrons (for $n \geq 5$).

2. (This problem exploits the idea of a "nearest point"). We are given n straight lines ($n \geq 3$) in the plane such that any two of them intersect and at least three lines pass through each intersection point. Prove that all the given lines meet at a point.

3. We are given n planes in space such that any three of them have a common point and at least four planes pass through each intersection point. Prove that all the given planes have a common point.

Exact estimate—subtle errors

In this section, we consider the basic problem:

We are given n straight lines in general position in a plane (that is, any three of them form a triangle). They divide the plane into several parts. Prove that among these parts there are at least n – 2 triangles and that this estimate is exact.

Exercise 4. Draw n lines in general position such that they form $n - 2$ triangular parts.

The observation that *a triangle cannot be destroyed* was at the basis of early attempts at proving the exact estimate. What we mean by this is that if we have a certain number of lines in the plane, and draw a new line, it will divide any triangle it intersects into two parts, one of which is again a triangle. (Is a similar assertion true for a tetrahedron and an intersecting plane?)

Here is one elegant, but erroneous, "solution": Suppose we have n lines in general position, and manage to find $n - 2$ triangular pieces. If we draw a new line, we will not destroy any of these triangles, and the "closest point" to the new line provides one more triangle. It remains to prove that there exist $n - 2$ triangles. Indeed, consider an arbitrary line. The other lines intersect it at $n - 1$ points, which form $n - 2$ segments. The lines that pass through the endpoints of these segments form $n - 2$ triangles (figure 2).

Exercise 5. Find the error in this "solution."

An attempt to use induction

Since no triangle can be destroyed, it seems reasonable to use induction—for example, to prove that adding one more line increases the number of triangles.

This is the line of reasoning led to many errors. The fact is that the underlying assertion is wrong: Adding a line doesn't necessarily mean that a triangle is added!

Exercise 6. Draw several lines in general position such that removing one of them does not decrease the number of triangles.

Attempts were made to prove that there exists a line such that removing it decreases the number of triangles (in which case induction could be used). However, the problem ultimately fell to other methods. The question of whether such a line always exists remains open.

So adding and removing lines was of no use. We'll try to find another line of reasoning.

Don't go to extremes to avoid extremes

The notion of general positions results from a desire to avoid considering degenerate cases. The very word "degenerate" has overtones of pathology, of something that must not happen. However, it is the idea of degeneracy that forms a basis for our solution to the problem on triangles. We first take a small side trip, and consider a new problem, not closely related to this one. Its solution will illustrate the idea of "going to extremes."

Problem 2. *In a convex pentagon, every diagonal cuts off a triangle. Prove that the sum of the areas of these triangles is greater than the area of the pentagon.*

Solution. It's rather difficult to deal with an arbitrary pentagon. It may be possible to "simplify" the situation—not to consider a particular case (which can also be useful), but rather to reduce the general solution to a specific one.

Let's move a vertex of one of the triangles along a line parallel to its base (figure 3). Then the area of this triangle and of the pentagon remain

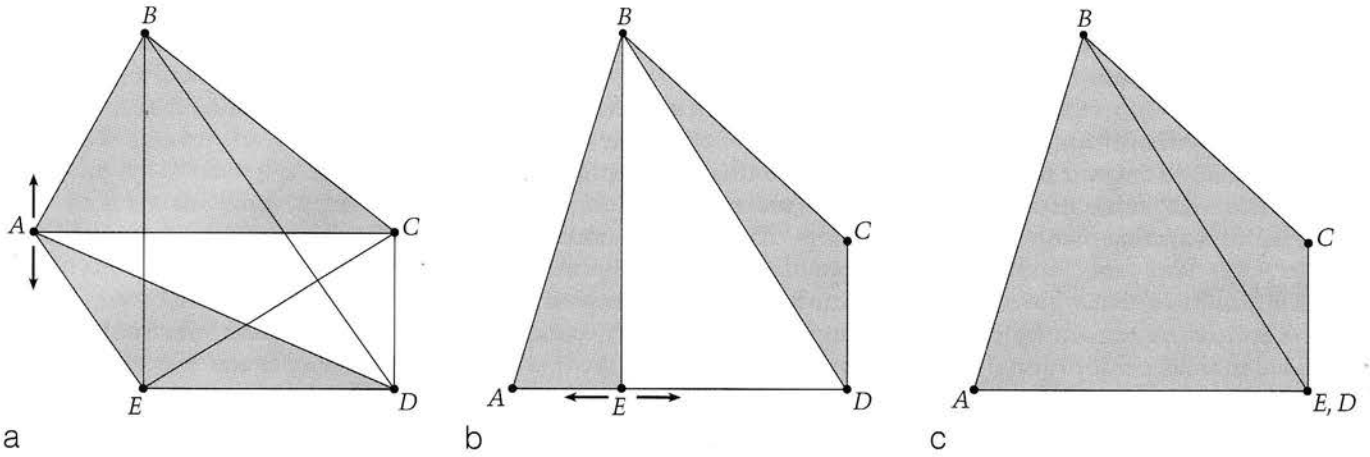


Figure 3. *Degeneration (simplification) of a pentagon.*

unchanged, but the areas of the neighboring triangles vary.

The key idea is as follows: if we move a vertex of a triangle along a straight line, its area either keeps increasing, keeps decreasing, or stays the same. Indeed, the base remains the same, and the altitude either keeps increasing, keeps decreasing, or remains the same. We can say a bit more: *If the vertex moves at a constant rate along a line, then the area of the triangle also varies at a constant rate.*

Therefore, the sum of the areas of all the triangles varies at a constant rate as well. It would be nice if the sum decreased, because we hope to prove the inequality for the new pentagon, in this case, it would be true for the original pentagon as well.

But what shall we do if the total area of the triangles increases? It's simple—*we just move the vertex in the opposite direction* and the area will decrease.

How far should we move the first vertex? We certainly want the pentagon to remain convex. Let us move the vertex as far as we can, without letting the pentagon lose this property. Then one of the angles of the pentagon (the one at E in the diagram above) will have become 180°. In this extreme position, the pentagon has become a simpler figure: a quadrilateral.

Now we repeat the reasoning for the vertex with the straight angle. In its extreme position, it will coincide with one of the neighboring vertices.

We can move the vertices further until the pentagon becomes a triangle (with two double vertices or one triple vertex). But it's not necessary, as triangles EAB and BCD already cover the entire pentagon.

Thus the proof for the original pentagon is reduced to a particular (degenerate!) case, and one that is quite easy to see.

Remark. The vertices of the pentagon can be moved along an arbitrary line. Then all the areas will vary linearly (i.e. at a constant rate), and we need only to ensure that the difference of the total area of the triangles and the pentagon (which also varies linearly) decreases.

Similar reasoning helps solve a number of other problems.

Exercises

7. Any three neighboring vertices in a convex hexagon form a triangle. Is the sum of these triangles always greater than the area of the hexagon?

8. A triangle is cut from a parallelogram. Prove that its area doesn't exceed half of the area of the parallelogram.

9. Given an arbitrary convex polygon, we select three points on its perimeter and look at the area of the triangle they determine. Prove that the triangle of largest area is formed by three of the vertices of the polygon.

10. A convex solid is placed inside a cube, and it turns out that the projection of this solid on any face of the cube covers this face. Prove that the volume of the solid is not less than a third of the volume of the cube.

11. If we are free to select any n points on a line segment, what is the maximal sum of the distances between these points, taken in pairs?

12. Prove that the maximum of the linear function $f(x, y, z) = ax + by + cz + d$, for all points (x, y, z) lying inside a convex polyhedron, is attained at a vertex of this polyhedron.

13. Ali Baba arrived at a cave where there are gold, diamonds, and a trunk. If the trunk is filled with gold it weighs 200 kg; if it's filled with diamonds, it weighs 40 kg; the empty trunk is weightless. A kilogram of gold costs 20 dinars and a kilogram of diamonds costs 60 dinars. How much money can Ali Baba get for the treasure if he can carry only 100 kilograms?

We will now return to our original problem, having drawn inspiration from this "method of extremes."

In solving all these problems, three ideas proved useful: (1) reducing the general case to a particular case by moving some elements of the figure; (2) choosing ways to move the figure at a constant rate, so as to improve the situation; (3) considering extreme cases, which are often degenerate.

Sometimes, in what follows, we will not look at the process of moving the figure in detail, but rather analyze only the end results of the process. Indeed, sometimes there is an enormous number of possible ways to move the figure, but the end result of any such movement can be analyzed simply.

Thus, in what follows we will

(1) move the lines such that their mutual arrangement is most convenient,

(2) use linear motion—that is, translation at a constant rate, and

(3) analyze extreme cases of the line's mutual arrangement.

A plan of action

The following reasoning is not included in the final solution. However, it is worth considering for three reasons. First, it shows how we have found the solution; second, similar reasoning can be used in solving many well-known problems; and third, it's useful to see how the solution is "cleaned up." This section will express the situation more intuitively. Later, we'll make our reasoning rigorous, and, finally, provide a short formal proof.

Where do triangles live?

All of our attempts at a systematic description of the arrangements of our lines have failed. The arragements are simply too numerous. Let's try moving the lines instead. To achieve linearity, we'll move them parallel to themselves. What can happen during such a motion?

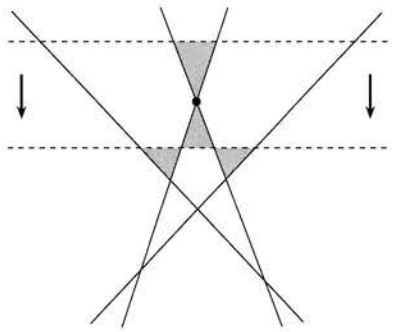


Figure 4. *As the broken line moves, a triangle is destroyed and other triangles are generated.*

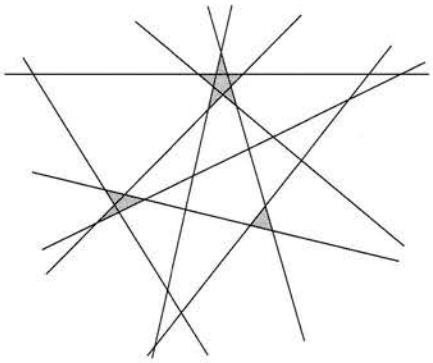


Figure 5. *Foci.*

Until the lines reach the points of intersection of other lines, the overall picture doesn't change. But as soon as several intersection points merge, a "catastrophe" occurs—one or several triangles disappear. As the line moves further, reorganization occurs, the results of which are difficult to foresee (figure 4).

We won't create a catastrophe. Rather, we'll stop moving the lines just before a catastrophe occurs, when triangles are not yet created or destroyed, but have become very small. We will call the arrangement of the lines just before the intersection points merge a *focus*. Thus a focus is formed by a number of lines whose intersection points lie in a small area, almost a point, and this area is not intersected by other lines (figure 5).

For our analysis, it's especially important that in the neighborhood of a focus we are dealing with a miniature decomposition of the plane, with a smaller number of lines. In other words, foci can be treated as little isolated "worlds" inhabited by minute triangles. Thus, if we manage to decompose the picture into foci, the triangles will be easier to count. Indeed, if we introduce an induction hypothesis (which we will do shortly), we can even assume that the problem has been solved for a smaller number of lines.

Returning to induction

Let us proceed by induction. Our induction hypothesis will be that any k lines, $k < n$, divide the plane into pieces among which there are at least $k - 2$ triangles. Thus if we analyze a focus that involves $k < n$ lines, we can be sure that there are at least $k - 2$ triangles in the neighborhood of this focus.

For example, assume that by moving the lines we managed to gather them in two foci with a common line, one involving lines from 1 through k and the other from k through n (the kth line is common for both focuses—see figure 6).

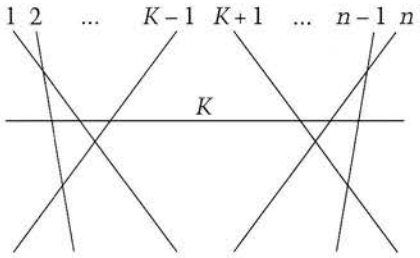


Figure 6. *Two foci.*

Then the total number of triangles in both foci is not less than $(k - 2) + (n - k + 1 - 2) = n - 3$. To make the inductive step, it remains to find one more triangle outside the foci. This can be done; however, can we be sure that it's possible to gather all the lines in just two foci? Unfortunately we can't, and if there are many foci, a great variety of situations arise. What can we do in this situation?

The boundaries of the possible

Let's analyze how foci are formed. At first the lines can be moved independently, but if a focus is already formed, it must be preserved. That is, the lines involved in the focus can only be allowed to move together, as a sort of "porcupine" (we must not allow the lines in the focus to move one relative to each other, or else a catastrophe might occure which would destroy the treatment of the focus as a miniature decomposition of the plane). Foci are "pseudo-points," and so can be augmented by new lines or joined with one another to form new "pseudo-points." Preserving the foci will impose ever stricter constraints on the movements of the lines, until it becomes impossible to move them any more.

In this process we must avoid contracting the whole picture into a single focus (so that we can use induction on the number of lines). For example, it is sufficient to fix two intersection points, so that they cannot move and become part of the same focus.

Now we will try to understand how the rates of motion of the lines involved in a focus must be related. Two lines can be moved independently if the speed of the others is

adjusted accordingly (consider the case of three lines). In other words, preserving a focus involving k lines requires $k - 2$ constraints on the speeds. However, by assumption, there are $k - 2$ triangles in this focus—exactly the same number as the number of constraints. This is a key to the solution.

Let's try to estimate the number of triangles in the final state (when the lines can't be moved any more). First, we fix a line and two intersection points on it—that is, three lines. We have $n - 3$ lines left free, and we will move them as long as it's possible. To preserve the foci, $n - 3$ constraints are required. By the induction hypothesis, the number of triangles in every focus is not less than the number of constraints. Thus, there are at least $n - 3$ triangles.

The finishing touch

It remains to find one more triangle—just one, somewhere among the foci. But it isn't clear how it can be found. Let's think this over: If we need one more triangle, why search for it? Isn't it possible to arrange things so that it exists *from the very beginning*? After all, it doesn't matter which three lines are fixed at the beginning of the process. So we can *fix three lines that already form a triangle!* Thus the problem is solved.

Constructing the solution

Our work is far from complete: We must clean up our reasoning, improve it, and make it rigorous. It's important to reveal the relationship between the triangles and foci, in particular, to understand why the number of triangles in a focus is not less than the number of constraints required to preserve this focus.

Exercise 14 (This exercise exploits the idea of a focus.) Moving a single line, prove that there exists a triangle adjacent to it.

Thus we have decided to preserve a triangle from the very beginning, which guaranteed that no collapse would occur and produced a triangle outside all the foci. But isn't it possible to preserve *all triangles*, granting them all equal rights? Then no foci will occur as the lines move (indeed, there exists a contracting triangle in any focus).

If the initial arrangement of lines is not rigid, a focus will inevitably occur during a certain motion, which leads to a contradiction. Thus, preserving all the triangles guarantees that the arrangement of lines is rigid, and it remains to prove that it is impossible to guarantee rigidity by preserving fewer than $n - 2$ triangles.

Recall how we preserved a focus: We allowed it to move as a whole. The simplest focus is a tiny triangle. We treat an ordinary triangle in a similar way: We allow it to move while preserving its size. (Verify that fixing $n/3$ triangles can lead to losing the mobility of the lines.)

However, now we haven't even fixed the initial triangle. So to avoid uninformative parallel translations of the entire arrangement, we fix *two lines*. (By the way, the position of a focus is also determined by two lines.)

Every triangle, as well as the simplest focus, yields one constraint on the rates of motion of the lines. Therefore, we obtain as many constraints as we had triangles that are preserved.

Thus we need to choose $n - 2$ speeds, which we will call parameters. If there are not enough triangles (fewer than the number of parameters), preserving their sizes cannot guarantee rigidity (figure 7).

Why can we be sure that if the arrangement is not rigid, a focus can be obtained? The reason is that, as in the problem about the pentagon, we can move the lines in the opposite direction—that is, change the direction of all the motions. Therefore, we can direct one of the lines to the intersection point of the fixed lines, and then a focus will inevitably occur. (It's interesting that while constructing the solution, induction dropped out, just as the scaffolding is removed when a building is constructed.)

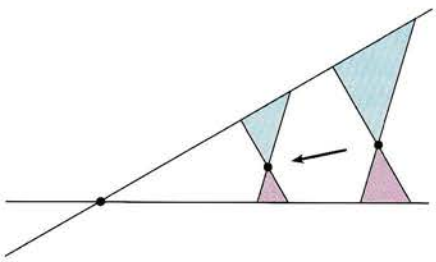


Figure 7. *It's possible to contract the red triangle while preserving the blue one.*

Summing up

First, we fixed two lines and allowed all the others to move at constant rates so as to preserve the size of all the triangles. Then, if the number of triangles turns out to be less than $n - 2$, the rates can be chosen to be nonzero (this fact will be proved later). Changing, if necessary, the direction of all the motions, we can create a focus where the uncounted triangle will be found. The contradiction obtained solves the problem.

Refining our reasoning

To obtain a rigorous proof, we must refine our intuitive reasoning—first of all, our arguments concerning rigidity. We translate them into algebraic language where the rates are interpreted as unknowns and the constraints as equations. The mobility of the lines means that there exists a nonzero solution to the system of equations.

Anyone who has dealt with systems of equations knows that the general rule is as follows: If the number of equations equals the number of unknowns, the system has a finite number of solutions; if the number of equations is greater than the number of unknowns (an overdetermined system), there are no solutions; and if the number of equations is less than the number of unknowns (an underdetermined system), the system has an infinite number of solutions. The last consideration is what we need for our problem.

Unfortunately, these considerations are valid only "as a rule"—that is, not always. For example, the system

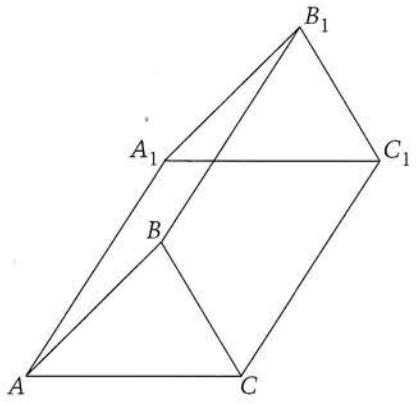


Figure 8. *Three parallelograms.* $S_{AA_1BB_1} + S_{BB_1C_1C} = S_{AA_1C_1C}$.

$$\begin{cases} x+y+z=1, \\ x+y+z=2 \end{cases}$$

has no solution, although there are fewer equations than unknowns. However, we will use one particular kind of system of equations. If all the unknowns in the system appear to the first power, and if the right-hand side of all the equations is zero, we say that the equations are *linear* and *homogeneous*. For such a system, the following theorem holds:

Theorem. *Any underdetermined system of m linear homogeneous equations with n unknowns (m < n)*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0, \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0 \end{cases}$$

has infinitely many solutions.

This theorem can be proved by induction, sequentially eliminating the unknowns by substitution. The proof can be found in any textbook on linear algebra.

Let's translate the relationships of the lines' rates of motion into the language of linear equations. The speed of a line is interpreted as the rate at which it moves from its initial position. It's easy to see that the intersection point of two lines that move at constant rates also moves at a constant rate, and the sides of the triangle change at a constant rate. It is also not hard to see that the sides of the triangles formed by three lines moving at a constant rate also change at a constant rate. From this we conclude that *the condition that the size of a triangle is preserved can be written as a linear homogeneous equation for the speeds of the lines.*

We prove this proposition geometrically. When a triangle is translated, three parallelograms are formed (figure 8), and the area of the larger one equals the sum of the areas of the other two. The area of the parallelogram is the product of the length of the triangle's side and the magnitude of the line's shift.

We assume that the direction the line is moving in is positive as the area of the triangle increases. Then the condition of the equality of the parallelograms' areas can be written as a linear homogeneous equation:

$$a_1h_1 + a_2h_2 + a_3h_3 = 0,$$

where a_1, a_2, a_3 are the sides of the triangle and h_1, h_2, h_3 are the shifts of the lines. A similar equation relates the speeds of the lines.

Thus we can write the condition that all the triangles are preserved as a system of linear homogeneous equations. Linearity has turned out to be very useful.

Exercise 15. Prove that, in the coordinate system (x, y), the equation of a line that moves at a constant speed v can be written as

$$x \sin \alpha - y \cos \alpha = c + vt,$$

where t stands for time and α is the slope of the line relative to the x-axis.

Rigorous proof

In conclusion, we give a rigorous proof in which all the details and turns of thought are hidden, as it is common in mathematical journals. Such mathematical texts are like rebuses or computer programs without comments. The Russian mathematician V. I. Arnold once remarked that their meaning, like the meaning of parables, is explained to students only in private.

Assume that the number of triangles k in the decomposition is less than $n - 2$. Let d be the minimum side of the triangles, let $v_1, \ldots, v_n$ be the speeds of the lines in perpendicular directions, and let $v_1 = v_2 = 0$.

The condition that we preserve the size of all the triangles is equivalent to a system of k linear homogeneous equations in the speeds v_i $(i = 3, \ldots, n)$. By the theorem above, this system has a nonzero solution.

We can assume (changing the direction of time, if necessary) that a certain line l_i moves in the direction of the intersection point of the lines l_1 and l_2. There exists a moment in time (the "catastrophe") when three or more lines will pass through the same point. Let t be the first such moment. Then, at the moment

$$t_1 = t - d/(2 \max v_i)$$

there exist three lines that form a nonintersected triangle with a side less than d. This fact contradicts the condition that the size of all the triangles is preserved, and this contradiction completes the proof.

Remarks

1. The condition of the problem for triangles can be extended for space of any dimension; in particular, if n planes in general position are given in three-dimensional space and they divide the space into several parts, there are at least $n - 3$ tetrahedrons among these parts. (Why is the number of triangles $n - 2$, but the number of tetrahedrons is $n - 3$? What would the number be for the one-dimensional case?)

2. Analyzing the solution, we see that the requirement that the lines be in general position can be relaxed: If among n lines in the plane any two intersect and *not all of them pass through the same point*, then there are at least $n - 2$ triangles among the pieces the plane is broken into.

The main difference in the proof for this variant is that points of multiple intersection can be destroyed in the process of moving the lines, and then new triangles appear. However, these triangles increase at a constant rate, and they are easily distinguished from the desired triangle, which is contracting into a point.

We invite the reader to construct the complete proof. ■

Challenges

Physics

P316

Weight on a string. A weight of mass m is suspended by an elastic, weightless string with a "spring constant" k. The maximum force the string can sustain is T. The weight is lifted to height x above the equilibrium position and then dropped from this height. At what minimum x will the string break?

(V. Kharitonov)

P317

Wire cuts ice. A wire loop with a weight attached is put on a block of ice (figure 1). Gradually the wire cuts the ice. This is because the pressure exerted by the wire decreases the melting point of the ice, which begins to melt under the wire and freeze above it. However, if the metal wire is replaced by a nylon thread of the same or smaller diameter, it will cut the ice but very slowly. Why? Do this experiment on your own. (I. Slobodetsky)

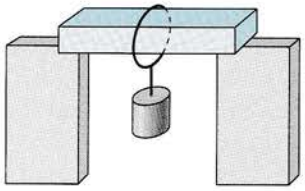


Figure 1

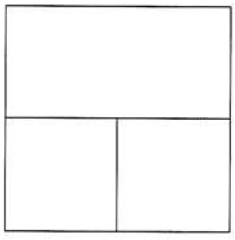


Figure 2

P318

Vessel with a partition. A tall vertical vessel with a square cross section, separated by vertical partitions into three sections (figure 2), was filled with various liquids to the same height. The large compartment contained hot soup (+65°C), while the two small sections were filled with warm stewed fruit at 35°C and cold Russian kvass at 20°C. The external walls are thermally isolated from the environment. All internal partitions are the same thickness and are made of material with a rather small thermal conductivity. After a while the soup cooled 1°C. Assuming all the liquids to be water (from the thermal viewpoint), find the temperatures of the kvass and stewed fruit. Note that the volumes of kvass and stewed fruit are the same, while the volume of soup is twice that of either the kvass or the fruit. (A. Zilberman)

P319

Electrical maxima in an LC circuit. A switch S is closed in the LC circuit shown in figure 3. Find the maximum value of the current in this circuit and the maximum voltage across the capacitor. (P. Zubkov)

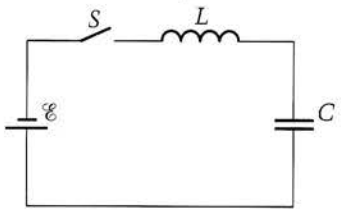


Figure 3

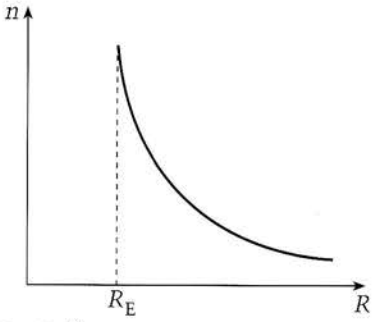


Figure 4

P320

Sparkling planet. During liftoff, astronauts observed a thin shining layer at altitude H_0 caused by total internal reflection of light in the atmosphere. They had a good reference book in the spacecraft, which contained a graph (figure 4) of the dependence of the refractive index n on the distance to the Earth's center (R_E stands for the radius of the Earth). How can one obtain the value of H_0 from the graph?

(O. Batishchev)

Math

M316

Composite after eight. A sequence of natural numbers $a_1, a_2, \ldots, a_n, \ldots$ is such that, for any $k \geq 1$,

$$a_{k+2} = a_k a_{k+1} + 1.$$

Prove that for $k \geq 9$ the number $a_k - 22$ is composite. (S. Genkin)

CONTINUED ON PAGE 43

Rock 'n' no roll

"... there are a lot of strange things in the mountains."
—Alexander Green

by A. Mitrofanov

"IT SHOULD ALSO BE REmembered that it's not an uncommon thing to encounter so-called 'rocking cliffs' in these places. This is a very curious phenomenon: a piece of cliff has acquired the conditions for stable equilibrium. It's usually situated on a stone platform, and if one shoves it, it returns to its initial position (like one of those inflatable toys with a weight at the bottom). These rocks sometimes weigh thousands of tons, but they respond to a push from a person of average strength. They cannot fall over, unless of course you blow them up with dynamite ..." This quotation is taken from *The Rocking Cliff* by the Russian writer Alexander Green. It's the sad tale of a poor hunter who was offered a huge sum of money to topple a huge stone pillar that rocked near its equilibrium position. Despite all his efforts, the hunter couldn't accomplish the task (although he continued trying) and went insane.

Let's try to figure out why such "rocking cliffs" are so stable.

For an object to be in equilibrium, we know that two conditions must be met:

(a) the vector sum of all the forces acting on the body must be zero; and

(b) the algebraic sum of all the torques relative to an arbitrary axis must also be zero.

However, not every equilibrium state is stable. For example, a needle affected only by gravity and a normal force doesn't stand vertically on a table. Nevertheless, if the needle is oriented strictly vertically, both equilibrium conditions (a) and (b) are met. The point is that a slight shift from the vertical position produces torques that topple the needle. In contrast, a brick stands firmly on each of its faces. We can make the brick smaller proportionally (that is, preserving its shape) as much as we want, and the smaller bricks will be in equilibrium on the table. However, it's much more difficult for a brick to be in equilibrium on a convex curved surface (say, on a football) in comparison with a flat or concave surface. Therefore, the conditions for an object to be in stable equilibrium depend on the object's shape (more precisely, on the shape of its base) and on the shape of the supporting surface.

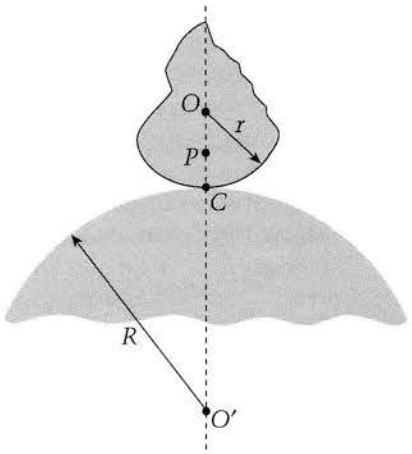
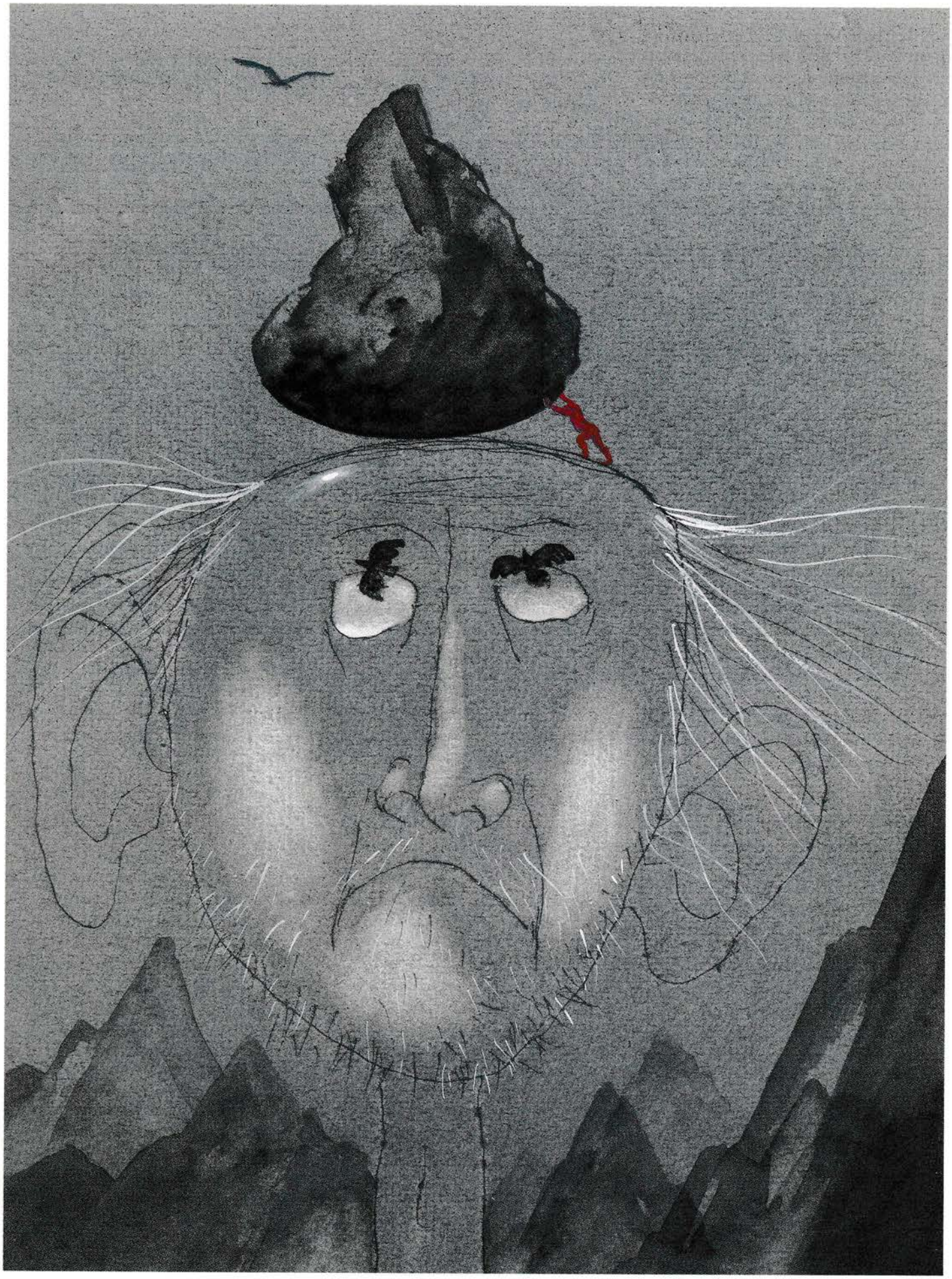


Figure 1

To deduce the criterion of stability, let's again consider the rocking cliff, or a boulder, and assume that both the boulder and its supporting surface, the ground, are spherically shaped at the point of contact. Let's assume that both the boulder and the ground have been eroded by wind and water. As a result, their surfaces are smooth and have no protruding points. In this case, the contact region between the two objects has narrowed to a point. Figure 1 shows the cross section of the stone and the ground in the vertical plane passing through the contact point C. Here O and O' are the centers of the spherical surfaces of the boulder and the ground in the contact area, while r and R are their radii. Above all, equilibrium requires that the boulder's center of gravity P lie on the vertical line OO'.

Art by Pavel Chernusky



Clearly both conditions (a) and (b) are met in this case. Let's consider what would happen if the boulder were given a slight push from its initial equilibrium position.

Let's say that our little push has put the boulder in the position shown in figure 2. Here Q is the intersection of the line PO with the vertical line passing through point A—the new contact point of the boulder with the ground. If point P lies to the left of the vertical line AA', the gravitational torque will "try" to restore the boulder to its initial position. This means that the equilibrium position of the boulder is stable.

Thus, if $CP < CQ$, the equilibrium position is stable. Let's consider how the values CP, R, and r are related in this case. The triangle OAQ (figure 2) yields

$$\beta = \frac{CA}{r} = \frac{C'A}{r} = \alpha\frac{R}{r},$$

because the angles are small. According to the law of sines,

$$\frac{OQ}{\sin\alpha} = \frac{r}{\sin(\pi-(\alpha+\beta))} = \frac{r}{\sin\left(\alpha+\alpha\frac{R}{r}\right)}. \qquad (1)$$

We are interested only in small deviations from the equilibrium position. When we say "small deviation" we assume that the distance "traveled" by the contact point on the ground (that is, the arc $C'A$ and thus the equal arc CA) is small compared to the radii r and R. And this means that the angles α and β are small:

$$\alpha \ll 1$$

and

$$\beta = \alpha\frac{R}{r} \ll 1.$$

As you probably know, the sine of a small angle is equal, to a high degree of accuracy, to the angle itself (if, of course, we measure angles in radians, which is what we're doing here). Therefore, equation (1) can be written as follows:

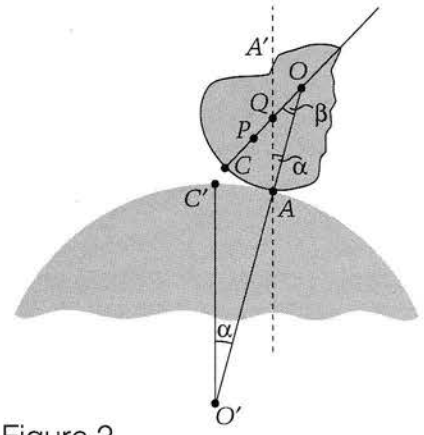


Figure 2

$$\frac{OQ}{r} = \frac{1}{1+R/r},$$

or

$$OQ = \frac{r^2}{R+r}.$$

Since

$$CQ = r - OQ = \frac{Rr}{R+r},$$

the condition for stable equilibrium of the boulder $CP < CQ$ is transformed to

$$CP < \frac{Rr}{R+r}. \qquad (2)$$

If the ground is concave and has a radius R, the condition of stable equilibrium will look like this:

$$CP < \frac{Rr}{R-r}. \qquad (3)$$

(Try to deduce this condition on your own.)

Now we should note an important feature of the problem. Assume that the boulder is in stable equilibrium. In this case, deflection of the boulder from its equilibrium position produces a torque that resists this deflection because the boulder has "thrown its weight behind" the new supporting point (figure 2). To keep the boulder at the new equilibrium position, one must apply an external force such that its torque relative to the new contact point is equal in magnitude and opposite in direction to the torque due to gravity. The magnitude and direction of the external force are determined by condition (a).

Now we see that one must perform some mechanical work to produce even a small deflection of the boulder from the equilibrium position. This work is spent on increasing the potential energy of the boulder. Therefore, the potential energy of the boulder is minimal at stable equilibrium—in other words, in this state the center of gravity is the lowest. This feature makes it possible to deduce condition (2) in another way: by considering how the center of gravity moves when there is a small deflection (see exercise 1). These two different ways of solving the problem are equivalent. If the boulder is slightly deflected from the equilibrium position, it will move back and travel beyond the equilibrium position due to its inertia. After a while the boulder will again pass through the equilibrium point. In short, the boulder will oscillate about its stable equilibrium position.

If a small deflection of an object from the equilibrium position lowers the center of gravity, the equilibrium of the object is unstable. The slightest deflection produces a gravitational torque that acts in the same direction and "tries" to increase the deflection. The object falls over.

There are cases when the displacement of an object from the equilibrium position doesn't change the height of the center of gravity relative to the supporting point. This kind of equilibrium is called neutral. For example, a homogeneous ball on a horizontal plane is in neutral equilibrium. Note that when the equilibrium is neutral, and if the object and the supporting surface are spherical in the contact area, the following equation holds:

$$CP = \frac{Rr}{R+r}. \qquad (4)$$

This equation is also valid for a homogeneous ball on a horizontal plane whose "radius" is infinite: $R \to \infty$. By the way, this condition is necessary, but will it be sufficient to yield neutral equilibrium? The answer is no. To prove this fact, we can show at least one example where condition (4) is met but the

equilibrium is neither neutral nor stable. Consider a spherical object placed on top of a fixed sphere of the same radius. The ball is not homogenous, and its center of gravity is located at half its radius, so $CP = r/2$. It's not difficult to show (do it on your own) that the equilibrium of such ball on the sphere is not stable, even though condition (4) is strictly satisfied. For any finite angle of tilt of the ball from the equilibrium position, the center of gravity drops, so it rolls down off the sphere (see problem 2).

Note that in deducing the stability criterion we considered only small displacements of the object from the equilibrium position and in our calculations took into account only the linear terms that are proportional to α. Under the conditions of this linear approximation, an object or a system of objects may remain in neutral equilibrium. However, if we consider the problem more carefully and take into consideration terms of higher orders (proportional to α^2, α^3, and so on), the equilibrium position may be unstable, as can readily be seen in practice.

Now we can better understand what the "rocking cliff" is: This is a vertically standing boulder with a low center of gravity or large radius of curvature at its base. Deflecting such a boulder (within certain limits, of course—see exercise 2) induces oscillations near its equilibrium position. And so a rocking cliff is simply a stone pendulum.

Clearly, it's not a simple matter for a person of moderate strength to rock a huge stone pillar. It's not just that the rocking cliff has a very large mass so that a very large force must be used to give it an appreciable acceleration. The support under the boulder has become deformed due to the boulder's weight, which leads to reactive forces preventing further displacement from the vertical (equilibrium) position. Nevertheless, rocking cliffs—or at least rocking (not to mention rolling) stones—exist in nature. Maybe you've seen them with your own eyes.

Now let's look at some examples in greater detail. We won't need to journey to a mountain area, but in essence they're the same as the rocking stones.

Example 1. In a uniform ball the center of gravity coincides with its geometric center, so the ball is unstable on a convex surface. However, if the "top" is cut from such a ball, it can rest in a stable manner on the top of a convex surface (see exercise 3).

Example 2. The amusing children's toy Weeble Wobbly resembles the cut ball in example 1. A piece of lead or steel hidden near the spherical base gives the toy a surprising stability.

Did you know that Weeble Wobbly had (and surely has) many relatives? "There were once five and twenty tin soldiers, all brothers, for they were the offspring of the same old tin spoon. Each man shouldered his gun, kept his eyes well to the front, and wore the smartest red and blue uniform imaginable."

Such were the steadfast tin soldiers from the famous fairy tale of Hans Christian Andersen (1805–1875). Why they were so steadfast? The reason is clear: no matter how one might try to tip them over, they always returned to the vertical position, ever alert. When a box packed with these soldiers was opened, they all jumped up as if by command. Every soldier was attached to the flat side of a lead hemisphere and was remarkably stable, standing at attention "forever."

Example 3. It's known that a uniform ellipsoid of revolution (elongated spheroid)—or, in other words, the body formed by the curve $(x/a)^2 + (y/b)^2 = 1$ rotating about its major axis X ($a > b$)—cannot stand vertically on a flat surface. The reason is that the radius of curvature of the ellipsoid's top (that is, the radius of the sphere that approximates the ellipsoid's surface at the apex) is b^2/a, while in this body the center of gravity is located at height a. Therefore, the criterion of stability (2) is not met.

However, what will happen to the stability if the shape of the ellipsoid is modified slightly? The Danish mathematician Pit Hein invented a body formed by rotation of the curve

$$\left(\frac{x}{a}\right)^{2.5} + \left(\frac{y}{b}\right)^{2.5} = 1.$$

In this formula the negative values of x and y are replaced by the corresponding absolute values. There is a happy choice of combinations of height a and width b (say, 5 and 4 cm, respectively) for which Hein's so-called super-ellipsoid rests in stable equilibrium when set on any of its poles.

It can be shown that this property of stability characterizes any uniform body formed by rotation of the curve $(x/a)^n + (x/b)^n = 1$ about the x-axis, provided that $n > 2$ and $a > b > 0$. For large n this body looks like a cylinder with rounded upper and lower bases—isn't that just the thing for a model of a rocking stone?

The problem of the stability of objects has occupied the minds of many outstanding scientists over the years. In 1644 Evangelista Torricelli (1608–1647) formulated a criterion of stable equilibrium of two bodies in a gravitational field. Later Christian Huygens (1629–1695) generalized it for a system of several bodies (Torricelli's principle). In 1788 Joseph Lagrange (1736–1813) proved a theorem that formulates the sufficient condition for equilibrium of a system of bodies. Later Peter Dirichlet (1805–1859) produced a more rigorous proof of this theorem. According to the Lagrange–Dirichlet theorem, if the potential energy of an isolated system is minimal at the equilibrium position, this equilibrium is stable. ■

Brocard points

by V. Prasolov

EVERY TRIANGLE HAS MANY interesting points: the intersection of its medians, the intersection of its altitudes, the centers of the circumscribed and inscribed circles, and so on. In this article I consider two such points—namely, the Brocard points.

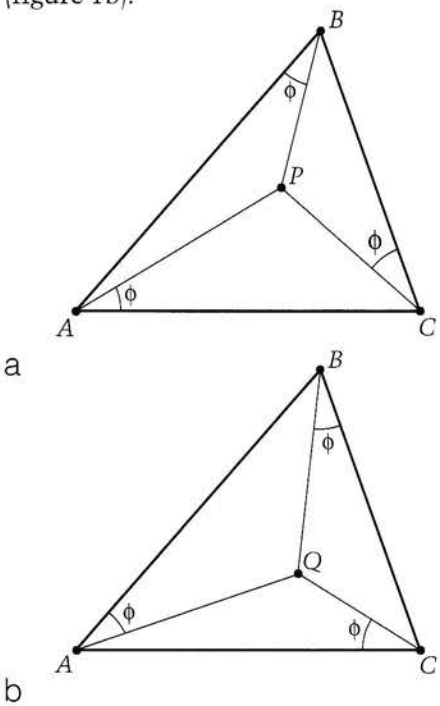
A point P inside a triangle ABC is called a *first Brocard point* if

$$\angle PAC = \angle PCB = \angle PBA$$

(figure 1a). A point Q inside a triangle ABC is called a *second Brocard point* if

$$\angle QAB = \angle QBC = \angle QCA$$

(figure 1b).



a

b

Figure 1

Before discussing the properties of these points, let's prove that, for every triangle, there exists exactly one first Brocard point and exactly one second Brocard point. We begin with the first point.

We begin by constructing the first Brocard point. In triangle ABC let $\angle BAC = \alpha$, $\angle ABC = \beta$, and $\angle ACB = \gamma$. The key is to construct triangles A_1BC, AB_1C, and ABC_1 outside triangle ABC, so that the three new triangles are all similar to the original triangle, and so that the angles are as shown in figure 2. (The reader should stop and prove that this is indeed possible.)

Our plan is to show that the circles circumscribed about these new triangles all pass through the same point, which has the first Brocard property. To this end, let P be the intersection point (other than C) of the circumcircles of A_1BC and B_1AC. We remember that a quadrilateral is inscribed in a circle if and only if its opposite angles are supplementary. This gives us

$$\angle BPC = 180^\circ - \alpha,$$

$$\angle APC = 180^\circ - \beta,$$

$$\angle APB = 360^\circ - \angle BPC - \angle APC$$
$$= 360^\circ - (180^\circ - \alpha) - (180^\circ - \beta)$$
$$= \alpha + \beta = 180^\circ - \gamma,$$

which means that P lies on the circumcircle of triangle AC_1B. The three circumcircles intersect at point P.

Now we draw PA, PB, and PC, and, noting equal inscribed angles, note that $\angle PAC = \angle PB_1C = \gamma - \angle ACP = \angle PCB$. We can show similarly that $\angle PCB = \angle PBA$. Thus P qualifies as a first Brocard point.

But is it unique? We can show that it is by reversing our reasoning. We invite the reader to check that our reasoning can be reversed. Suppose P' is a point such that $\angle P'AC = \angle P'CB = \angle P'BA$. Then $\angle ACB = \gamma = \angle P'CA + \angle P'CB = \angle P'AC + \angle P'CA = 180^\circ - \angle AP'C$. This implies that P' is on the circle passing through A, B_1, and C. Similarly, P is on the circle passing through A_1, B, and C, and thus P' coincides with P. This means that the point P, satisfying the first Brocard property, is indeed unique.

We note in passing that if we draw PA_1, PB_1, and PC_1, then $\angle APB_1 = \angle ACB_1 = \alpha$. Similarly, $\angle CPA_1 = \gamma$, and $\angle B_1PC = \beta$. Hence $\angle APA_1 = \alpha + \beta + \gamma = 180^\circ$, so A, P, and A_1 are collinear. Similarly, B, P, and B_1 are collinear, as are C, P, and C_1.

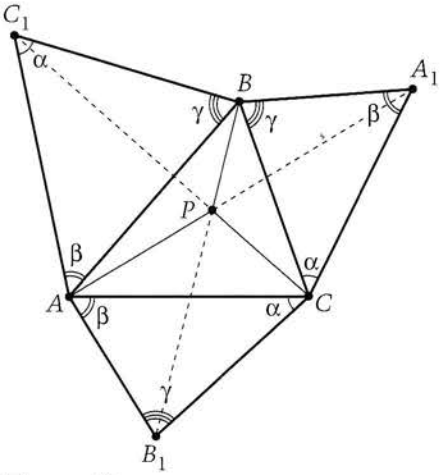


Figure 2

Problem 1. Construct triangles A_1BC, AB_1C, and ABC_1 similar to triangle ABC on its sides such that segments AA_1, BB_1, and CC_1 meet at the second Brocard point.

Thus we have proved that there exists exactly one first Brocard point in every triangle. So we're ready to discuss some properties of Brocard points. If we draw lines AO, BO, and CO through the center O of the circumscribed circle of triangle ABC, these lines intersect the circle at points A_1, B_1, and C_1 such that triangles ABC and $A_1B_1C_1$ are congruent (they are symmetric about the point O). The Brocard points possess a similar property.

Problem 2. (a) Let P be the first Brocard point. Lines AP, BP, and CP intersect the circumscribed circle of triangle ABC at points A_1, B_1, and C_1, respectively. Prove that $\triangle ABC = \triangle B_1C_1A_1$.

(b) Formulate and prove a similar proposition for the second Brocard point.

If we draw perpendiculars OA', OB', and OC' from the center O of the circumscribed circle of triangle ABC to its sides, points A', B', and C' will in fact be the midpoints of the sides, and it is not hard to see that $\triangle ABC$ is similar to $\triangle A'B'C'$. The Brocard point P possesses a similar property.

Problem 3. Perpendiculars PA', PB', and PC' are drawn from the first Brocard point P to the sides of triangle ABC. Prove that $\triangle ABC$ is similar to $\triangle B'C'A'$.

In a certain sense, Problem 3 becomes redundant after Problem 2 has been solved. The fact is that the following proposition holds.

Problem 4. Let lines AX, BX, and CX intersect the circumscribed circle of triangle ABC at points A_1, B_1, and C_1, respectively. Let also A', B', and C' be the projections of X onto the sides of triangle ABC. Prove that $\triangle A_1B_1C_1$ is similar to $\triangle A'B'C'$.

We now consider angle $\phi = \angle PAC = \angle PCB = \angle PBA$. We can express ϕ in terms of the angles of triangle ABC. To this end, we erase all the construction lines in figure 2 (see figure 3), and draw A_1K from point A_1 to line AC, then, from the triangles AKA_1, A_1CK, we have

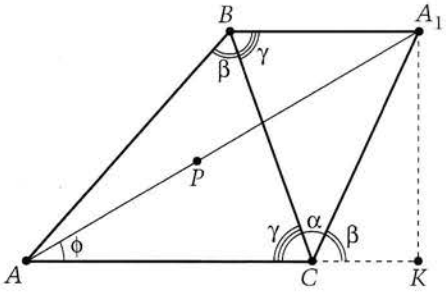


Figure 3

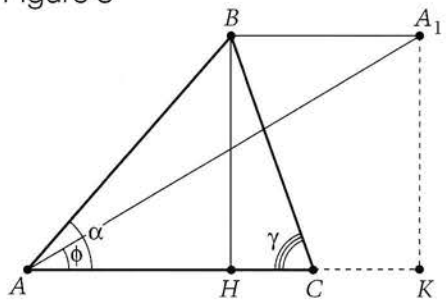


Figure 4

$$\cot\phi = AK/A_1K = AC/A_1K + CK/A_1K = AC/A_1K + \cot\beta.$$

Now BA_1 is parallel to AC, hence A_1K is equal to the altitude of triangle ABC drawn to side AC. If BH is this altitude (see figure 4), then

$$\frac{AC}{A_1K} = \frac{AC}{BH} = \frac{AH}{BH} + \frac{HC}{BH} = \cot\alpha + \cot\gamma.$$

Hence $\cot\phi = \cot\alpha + \cot\beta = \cot\gamma$.

For the second Brocard point, we obtain the same expression. Angles in the range from 0° to 180° are equal if and only if their cotangents are equal. Thus we obtain the same angle ϕ for the second Brocard point. This angle is called the *Brocard angle*.

Problem 5. (a) Prove that the Brocard angle ϕ is less than or equal to 30°.

(b) A point M is given inside triangle ABC. Prove that one of the angles ABM, BCM, or CAM does not exceed 30°. (This problem was given at the International Mathematical Olympiad in 1991.)

Let's discuss in more detail the fact that the angles for the first and second Brocard points are equal. Let P and Q be these points. Reflect the lines AP, BP, and CP in the bisectors of angles A, B, and C, respectively. Then the lines obtained meet at point Q. However, this property is not unique to the Brocard points: For any point X not on the circumscribed circle of triangle ABC, when lines AX, BX, and CX are reflected in the bisectors of the corresponding angles, the resulting lines meet at a point. We won't discuss this remarkable fact here—it deserves a separate article. We only note the following.

Problem 6. Let O be the circumcenter of triangle ABC. Prove that the lines symmetric to lines AO, BO, and CO about the bisectors of angles A, B, and C, respectively, pass through the point of intersection of the triangle's altitudes.

We leave it to you to solve the following problems concerning the properties of the Brocard angles and points.

Problems

7. Let P be the Brocard point of triangle ABC, and let R_1, R_2, and R_3 be the radii of the circumscribed circles of triangles ABP, BCP, and CAP, respectively. Prove that $R_1R_2R_3 = R^3$, where R is the radius of the circle circumscribed about triangle ABC.

8. Let Q be the second Brocard point of triangle ABC, let O be the center of its circumscribed circle, and let A_1, B_1, and C_1 be the centers of the circles circumscribed about triangles CAQ, ABQ, and BCQ, respectively. Prove that $\triangle A_1B_1C_1$ is similar to $\triangle ABC$, and that O is the first Brocard point of triangle $A_1B_1C_1$.

9. Prove that one can construct a triangle $A_1B_1C_1$ from the medians of triangle ABC, and that the Brocard angles of both these triangles are equal.

10. An equilateral triangle ABC is given with its center at a point O. Prove that the Brocard angle of the triangle formed by the projections of an arbitrary point X on the sides of triangle ABC depends only on the length of OX. ■

The near and far of it

Limitations of optical instruments

by A. Stasenko

MR. LUND CAME TO THE telescope and began to look at the Moon.

"Don't you see the pale spots moving near the Moon?"

"Good gracious, sir! Call me an old coot if I can't see these spots! What are they?"

"These are spots that can be seen only with my telescope. Time's up! Leave it be."

A half-hour later, Mr. William Chatterly, John Lund, and the Scotsman Tom Snipe were flying off to the mysterious spots on eighteen balloons.

Readers wishing to learn more about Mr. Chatterly can read his remarkable treatise, "Did the Moon exist before the Flood? If yes, why didn't it sink?" By the way, this book also describes how the author lived in the Australian swamps for two years, where he subsisted on crayfish, ooze, and crocodile eggs ... and devised a microscope very much like our ordinary microscope.

—Anton Chekhov, *Flying Islands* (a parody of Jules Verne)

Of all the devices invented by physicists, two have attained widespread fame: the telescope and the microscope. One is aimed at the depths of the Universe, while the other allows us to see little things that are literally under our nose. Let's take a quick look at how these devices work.

From the point of view of geometric optics, the telescope is pretty straightforward. It consists of two coaxial lenses with focal lengths F_{obj} (for the objective) and F_{eye} (for the eyepiece), respectively (figure 1). Let's aim this device at a pair of stars that are near each other. The rays from each star are nearly parallel. From the definition of focal length, the objective focuses the light at points B and K on its focal plane (figure 1). But in a telescopic system this plane coincides with the focal plane of the eyepiece, so after passing through the eyepiece, the rays from each star will also be parallel. Denote the angle between incident rays 1 and 2 (from the two stars) by α and the angle between the emerging (refracted) rays 1′ and 2′ by β. It's easy to see the "trick" of a telescopic system. The rectangular triangles OBK and $O'BK$ show that their common side is

$$BK = F_{obj} \tan \alpha = F_{eye} \tan \beta,$$

from which we get

$$\frac{\tan \beta}{\tan \alpha} = \frac{F_{obj}}{F_{eye}} \approx \frac{\beta}{\alpha}. \qquad (1)$$

This approximate equation is valid for the small angles characteristic of most optical devices.

At first glance, equation (1) opens unlimited possibilities for increasing the magnification of a telescope: We just need to use an objective with the longest possible focal length (this explains why refracting

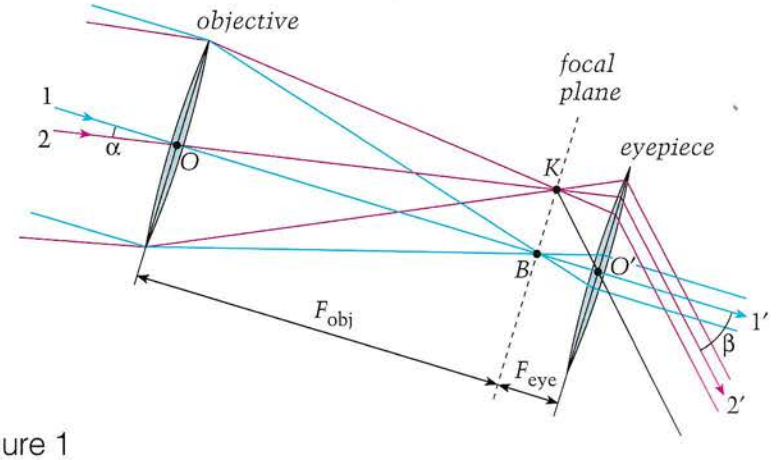
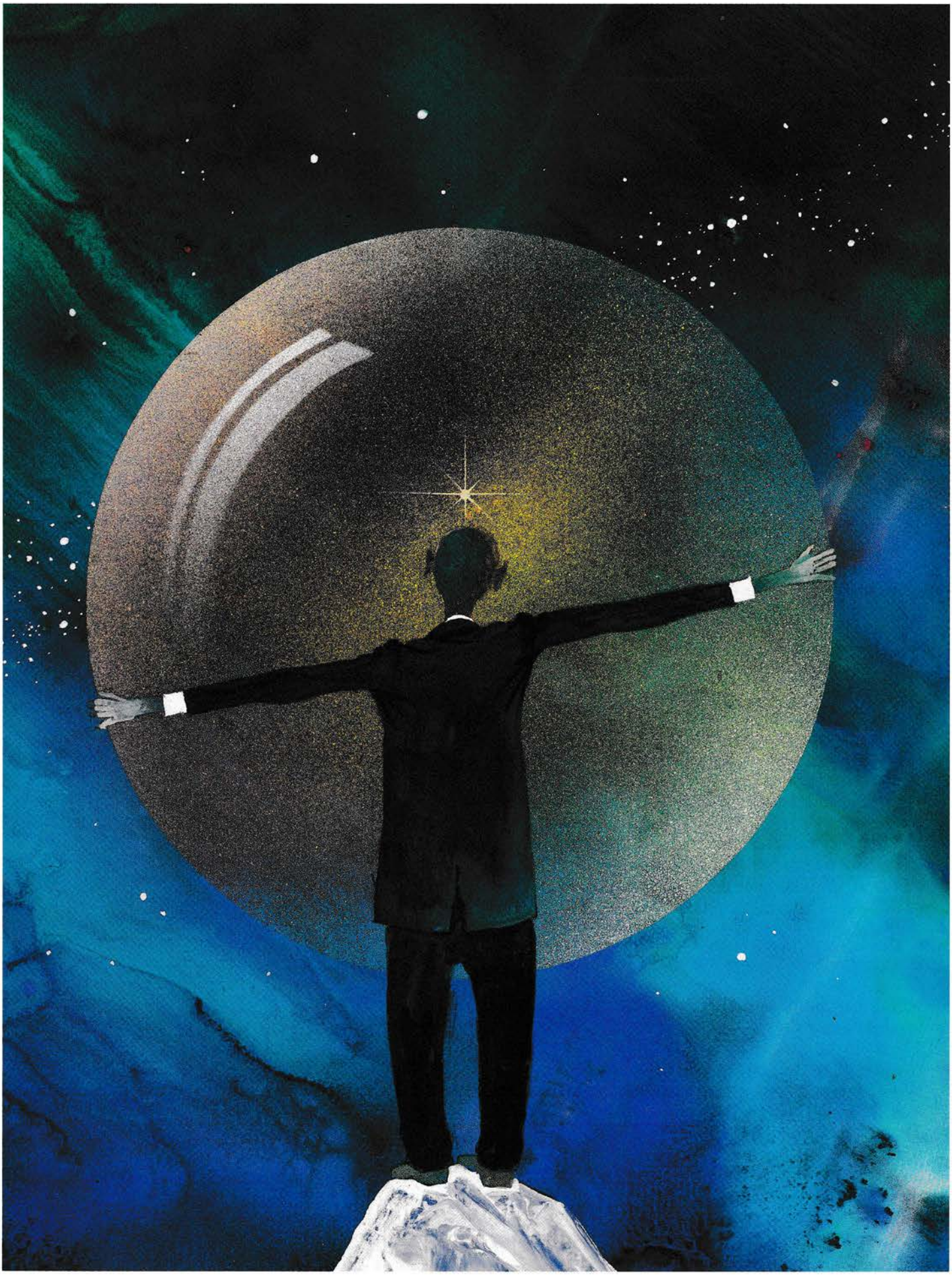


Figure 1

Art by Pavel Chernusky



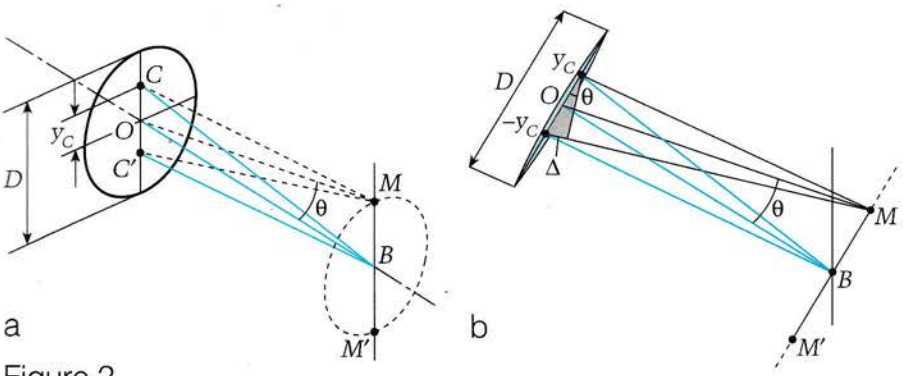


Figure 2

telescopes have such a large base length) and an eyepiece with the shortest possible focal length.

Unfortunately, a characteristic parameter of light foils our plans: the wavelength λ. And how could it be otherwise? A beam of light is composed of electromagnetic waves with wavelengths in the range $0.4\ \mu m \le \lambda \le 0.8\ \mu m$. And any wave passing near an obstacle diffracts. In addition, according to the Huygens–Fresnel principle, any portion of the primary wave (say, that located on the plane of the objective) can be considered a source of secondary waves that interfere with one another wherever they meet—for example, on the focal plane of the objective.

Let's use this principle to approximately determine how much the light emitted by a star is diffracted by the objective of a telescope. We divide the objective into two parts (figure 2a) and consider them to be the sources of the secondary waves. The distance between points C and C' is about half the objective's diameter D, so the difference in the paths of the waves arriving at point M is approximately equal to

$$\Delta = \frac{D}{2}\sin\theta,$$

which can be obtained from the small triangle drawn in figure 2b. The resulting interference is defined by this difference. For example, at point B (and along the entire optical axis OB) we have $\theta = 0$ and $\Delta = 0$. Therefore, the waves amplify each other along the optical axis. If we place a screen in the plane normal to the optical axis and passing through the focal point (this is called the focal plane), we'll see a bright spot.

We can obtain a better formula for the path difference at point M by assuming that points C and C' are the centers of mass of each half of the objective. It can be shown experimentally (by cutting a semicircle of cardboard and balancing it on a knife's edge) that the center of mass of a semicircle is located at the height

$$y_C = \frac{4}{3\pi}\frac{D}{2}$$

above its diameter. Thus the path difference Δ of two spherical waves emitted from points C and C' at an angle θ to the optical axis is

$$\Delta = 2y_C\sin\theta = \frac{8}{3\pi}\frac{D}{2}\sin\theta. \quad (2)$$

Now let's shift the observation point at which we examine the interference up or down in the focal plane. It's important to find the angle $\theta_{1\,min}$ for which the path difference is $\Delta_{1\,min} = \lambda/2$, so that the waves annihilate each other. Equation (2) yields

$$\sin\theta_{1\,min} = \frac{3\pi}{8}\frac{\lambda}{D} = 1.18\frac{\lambda}{D} \approx \theta_{1\,min}.$$

Of course, a correct application of the Huygens–Fresnel principle requires us to take into account and add (that is, integrate) all the elementary waves emitted from small areas of the primary wave. In this process we would need Bessel functions, which in the case of axial symmetry are analogous to "conventional" sines and cosines that describe a number of similar one-dimensional problems (such as the vibration of a guitar string). This solution gives the following formula:

$$\sin\theta_{1\,min} = 1.22\frac{\lambda}{D}. \quad (3)$$

Our simple approximation differs from the strict theoretical result by a mere fourhundredths—not bad. But why are we so interested in this angle? Because it corresponds to the radius $BM = F_{obj} \cdot \theta_{1\,min}$ of the first dark ring surrounding the bright spot (the image of the star) in the focal plane of the objective. It turns out that this image is far from being an infinitely small point, as geometrical optics says it should be. This means the second star, located at an angular distance α from the optical axis, also produces a bright spot on the focal plane, and the problem is how far this new image must be from the first one so we can tell one star from the other. The great physicist Sir John Rayleigh (1842–1919) suggested a simple criterion for resolution:

$$\alpha \gtrsim \theta_{1\,min}. \quad (4)$$

If this requirement is not met, the images of both stars will be fused even at the focal plane of the objective. Try as we might, we won't be able to separate them.

Now let's turn our eyes from the heavens and squint into a microscope. Using the thin-lens approximation, we can plot the images of the object formed by the objective and the eyepiece (figure 3). It's important that the object be located

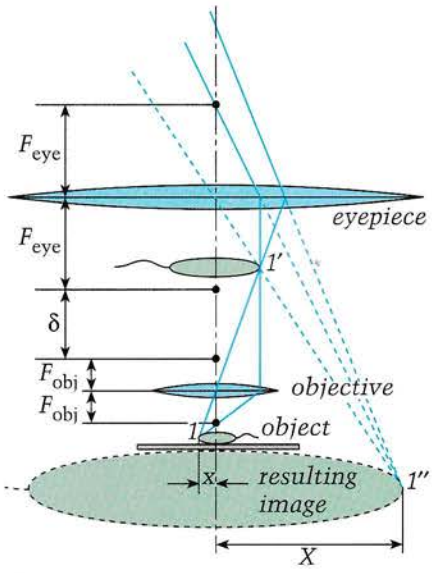


Figure 3

outside the focal point of the objective to form the real image l' and that this real image be located between the eyepiece and its focal point to produce the final virtual image l''.

Geometrical optics yields the following formula for the magnification of our simple microscope (figure 3):

$$\frac{X}{x} = \frac{D_0 \delta}{F_{eye} F_{obj}},$$

where Δ is the distance between the foci of the objective and the eyepiece and D_0 is the near point of the eye—that is, the smallest distance at which the eye can focus. The magnification of a microscope can be very large. For example, for reasonable values $F_{obj} = 2$ mm, $F_{eye} = 15$ mm, $\delta = 160$ mm, and $D_0 = 250$ mm, the magnification is $X/x = 1{,}333$.

It would seem that this isn't the limit—we could increase the magnification by improving the quality of the lenses (through better polishing) and by eliminating their defects—aplanatism, astigmatism, chromatic and spherical aberration, distortion, and so on. But the wavelength λ returns to play its tricks again!

The theory of the resolving power of microscopes was developed by Ernst Abbe (see the article in the July/August 2000 issue of *Quantum*). Abbe came up with the idea of looking at a diffraction grating under a microscope (figure 4). What is the minimum amount of information that can be obtained about this grating? Well, first of all we can try to determine its period d.

We know that when light of wavelength λ passes through a diffraction grating, it produces a pattern of light with a set of diffraction maxima. If the light hits the grating at some angle θ_0, the directions to these maximums are determined by the equation

$$\Delta_m - \Delta_0 = d \sin \theta_{m\,max} - d \sin \theta_0 = m\lambda. \quad (5)$$

A microscope will provide information about the period d if at least two beams arrive at its objective, which correspond to two adjacent

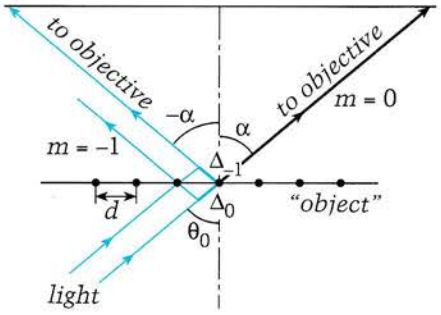


Figure 4

maxima of the diffraction pattern (say, the maxima with $m = 0$ and $m = -1$). This limiting case with $\alpha = \theta_0$ and $\alpha = \theta_{-1\,max}$ is shown in figure 4. Note that the period d of the screen to be resolved is very small—on the order of microns. Compared to such a tiny size, the objective of the microscope and its focal length (several millimeters) are so large that the objective should actually be drawn far outside the margins of this magazine's pages—at a distance of several meters. Therefore, the objective is shown schematically by the dashed line, and the rays traveling to it from the screen are drawn almost parallel.

For the case of two neighboring maxima, equation (5) yields $2d \sin \alpha = \lambda$, where α is the angular aperture. Thus, at the given wavelength of incident light, the smallest period of the screen that can be resolved and seen with a microscope is $d_{min} = \lambda/(2 \sin \alpha)$.

We can improve things a bit by placing a transparent medium with refractive index n (say, a drop of some liquid) between the screen and the objective. As a result, the difference between the paths will be increased, because in this medium the speed of light and the wavelength are smaller by a factor of n, which gives us

$$d_{min} = \frac{\lambda}{2n \sin \alpha}.$$

Now let's compare the resolving power of a telescope and a microscope. To improve this important parameter in both optical instruments, two opposing requirements must be met: For the telescope, $\lambda/D \cong \alpha_{min}$ should be as small as possible, while for the microscope, $\lambda/d \cong 2n \sin \alpha$ should be as large as possible.

Now it's clear why telescopes are constructed with the largest possible diameter of the input "pupil" (objective), while microscopes have the smallest possible focal length of the objective (to get $\sin \alpha$ as close as possible to 1). In addition, the space between the objective and the examined object is filled with some liquid that has the largest possible refractive index n (this is the so-called *immersion technique*).

What have scientists and engineers achieved in their attempts to increase the resolving power of optical devices? The largest diameter of an optical telescope is $D \sim 6$ m. Equations (3) and (4) yield $\lambda_{min} \sim 10^{-7}$ for the "average" wavelength of visible light $\lambda \sim 0.6$ μm. Assuming the radius of the Universe to be $R \sim 10^{26}$ m, the minimum distance between two resolvable points at its "boundary" must be

$$l_{min} \sim R\alpha_{min} \sim 10^{19} \text{ m}.$$

To estimate the resolving power of a microscope, we assume $\sin \alpha \leqslant 1$ and $n \cong 1.6$ (the refractive index of an aniline). In this case, equation (6) yields

$$d_{min} \gtrsim \frac{\lambda}{4} \sim 0.1\,\mu\text{m} = 10^{-7} \text{ m}.$$

Now we know the characteristic limits for the functionality of these wonderful optical instruments. To further improve our ability to observe both very distant and very small objects, we must rely on other techniques (such as X-ray astronomy or electron microscopy). ■

Quantum on light diffraction and interference:

A. Eisenkraft and L. D. Kirkpatrick, "Color Creation," November/December 1997, pp. 32–33.

V. Surdin and M. Kartashev, "Light in a Dark Room," July/August 1999, pp. 40–44.

A. Stasenko, "Physical Optics and Two Camels," September/October 1999, pp. 44–47.

Many ways t

WHEN I WAS A KID, I HAD a hard time remembering the multiplication table. Of course, some of it was easy—2 times 2 is four, 5 times 5 is 25—but 7 times 8 or 9 times 6 just wouldn't settle in my noggin. Here's how I calculated such difficult products: 5 times 8 gives 40, and 2 times 8 gives 16, which adds up to 56. But my teacher wanted quick answers, so I had to learn the table by heart. I wasn't the only one to suffer from the multiplication table. (And some people never get over it, to judge from some of my students at the Moscow Institute of Physics and Technology, whose grasp of the table is rather shaky.)

Eventually I learned the multiplication table. It ended up being pretty useful, and not just to my teacher: Using this table, I could multiply any numbers I came across, and pretty quickly, too.

For many years I was convinced that it was impossible to multiply quickly without using the multiplication table. This certainty became even stronger when I learned the different multiplication methods used in India, in China, and in Europe during the Renaissance.

One day I came across an old Russian method used by peasants about two hundred years ago and saw that it didn't require any knowledge of the multiplication table. All one needed to know was how to multiple and divide by two and to add numbers. Here's how they did it.

Let's write the numbers on a single line, one on the left and the other on the right. We'll divide the left number by two and double the one on the right. We'll put the results

13	17
~~6~~	~~34~~
3	68
1	136
	221

Figure 1

in a column as shown in figure 1. When an odd number is divided by two, we'll discard the remainder. When we obtain 1 on the left-hand side, we eliminate all the rows in which the left-hand side contains an even number. We then add up all the remaining numbers on the right-hand side. The result obtained is the product of the two original numbers!

Naturally, I didn't believe in this method at first. I began experimenting and multiplied 13 by 17 using this method. The answer was correct: 221. Then I changed the order of the factors and did the multiplication again. The answer was again correct (figure 2).

17	13
~~8~~	~~26~~
~~4~~	~~52~~
~~2~~	~~104~~
1	208
	221

Figure 2

I couldn't believe my eyes. It was like one of those mathematical tricks where incorrect operations give a correct answer (see, for example, figure 3).

$$\frac{\cancel{16}}{\cancel{64}} = \frac{1}{4}, \quad \frac{\cancel{49}}{\cancel{98}} = \frac{4}{8} = \frac{1}{2}.$$

Figure 3

However, the method turned out to be quite correct. Before reading further, try to multiply several pairs of numbers by this method to convince yourself that it's valid.

And now I'm going to show that this method always yields a correct answer.

In earlier issues of our magazine, articles were published that discussed the binary system of notation (For example, "Number Systems" by I. M. Yaglom in the July/August 1995 issue of *Quantum*). In this system, any number is represented as a sequence of ones and zeros—for example, $32 = 100{,}000_2$, $13 = 1101_2$, $17 = 10{,}001_2$. The subscript 2 shows that it is written in binary notation. These representations are interpreted as follows:

$$32 = 100{,}000_2 = 1 \cdot 2^5 + 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0,$$

$$13 = 1101_2 = 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1,$$

$$17 = 10{,}001_2 = 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1.$$

Under each digit of the binary representation of 13, let's write the corresponding number from the left-hand column obtained when multiplying by the "Russian peasant method," and do the same for the number 17 (figure 4). Do you see a

1	1	0	1
1	3	6	13

1	0	0	0	1
1	2	4	8	17

Figure 4

pattern? Yes, you're right. If a certain place in the binary notation is occupied by the digit 1, then an odd number is written under it; otherwise, this number is even. Try to prove this fact.

Now I'll reformulate the peasant rule of multiplication. On the right-hand side, we write the numbers equal to the second factor multiplied by 2 raised to a power that is 1

s to multiply

less than the number of the row in which the number is written. Then the result is multiplied by 1 if the corresponding number in the left-hand column is odd, and by 0 if it is even. To make this clearer, I suggest that we add one more column between the first two and write in it the remainder upon division of the corresponding number on the left-hand column by 2 (figure 5). Thus, in

13	1	17
6	0	17×2
3	1	17×2^2
1	1	17×2^3
		221

Figure 5

essence, the peasant method of multiplication suggests that the right and middle columns are multiplied row by row, and the results are combined. In our example of multiplying 13 by 17, we have

$$17 \cdot 1 \cdot 1 + 17 \cdot 0 \cdot 2 + 17 \cdot 1 \cdot 2^2 + 17 \cdot 1 \cdot 2^3$$
$$= 17(1 + 0 \cdot 2 + 1 \cdot 2^2 + 1 \cdot 2^3)$$
$$= 17 \cdot 1101_2 = 17 \cdot 13.$$

Multiplying the factors in reverse order, we have

$$13 \cdot 1 \cdot 1 + 13 \cdot 0 \cdot 2 + 13 \cdot 0 \cdot 2^2 + 13 \cdot 0 \cdot 2^3 + 13 \cdot 1 \cdot 2^4$$
$$= 13(1 + 0 \cdot 2 + 0 \cdot 2^2 + 0 \cdot 2^3 + 1 \cdot 2^4)$$
$$= 13 \cdot 10001_2 = 13 \cdot 17.$$

We see that the peasant method of multiplication is based on representing one of the factors in binary form. Isn't that simple and elegant?

Now let's see how well this method works for larger numbers—say, 567 and 3,984. At the left of figure 6 we give the peasant method of multiplying these numbers, and at the right we give the conventional method. We see that in the conventional method, less addition is required, but the summands are obtained in a more complex way. With the peasant method, what we gain in simplifying the calculations we lose in time, so that the conventional method is perhaps better.

567	3984
283	7968
141	15936
~~70~~	~~31872~~
35	63744
17	127488
~~8~~	~~254976~~
~~4~~	~~509952~~
~~2~~	~~1019904~~
1	2039808
	2258928

$$\begin{array}{r} 567 \\ \times\ 3984 \\ \hline 2268 \\ +\ 4536 \\ +\ 5103 \\ +\ 1701 \\ \hline 2258928 \end{array}$$

Figure 6

"No!" say those who are not on good terms with the multiplication table. "With the peasant method you don't have to memorize any tables, and that's certainly worth something!" Well, I can present the following argument in favor of the multiplication table: It would be inconvenient to take a pen and a piece of paper or look for a copy of the "Pythagorean table" (figure 7) to calculate how much 8 pies would cost at 70 cents each.

1	2	3	4	5	6	7	8	9	10
2	4	6	8	10	12	14	16	18	20
3	6	9	12	15	18	21	24	27	30
4	8	12	16	20	24	28	32	36	40
5	10	15	20	25	30	35	40	45	50
6	12	18	24	30	36	42	48	54	60
7	14	21	28	35	42	49	56	63	70
8	16	24	32	40	48	56	64	72	80
9	18	27	36	45	54	63	72	81	90
10	20	30	40	50	60	70	80	90	100

Figure 7

This table really is ancient: The Pythagoreans used it more than 2,000 years ago. (Recently, the Tashkent mathematician A. Azamov noticed an interesting property of this table: If we choose four numbers from the table whose positions form the vertices of a square, and there is a number at the center of the square as well, then the number in the center is the arithmetic mean of the numbers at the vertices. For example, for the numbers that are highlighted in figure 7, we have $42 = (25 + 48 + 63 + 32)/4$.)

Over several thousand years of mathematical study, many methods of doing multiplication were invented. Around the turn of the 16th century the Italian mathematician Luca Pacioli presented eight different methods in his treatise on arithmetic. Here I'll describe two that I find most interesting.

In the first method, called *a small castle*, the digits of the upper number are multiplied by the lower number one by one, beginning with the most significant one, and written in a column (figure 8).

$$\begin{array}{r} 3984 \\ \times\ 567 \\ \hline 1701000 \\ +\ 510300 \\ +\ 45360 \\ +\ 2268 \\ \hline 2258928 \end{array}$$

Figure 8

CONTINUED ON PAGE 49

The fundamental particles

by Larry D. Kirkpatrick and Arthur Eisenkraft

THE SEARCH FOR THE FUNdamental building blocks in nature has gone on for more than two thousand years. Aristotle felt that all the materials around us were composed of varying quantities of four basic *elements*—earth, fire, air, and water. In hindsight, this may seem simplistic. How can the myriad of properties of these materials arise from only four basic elements that don't share these properties? Is this any more strange than believing that everything is composed of chemical elements? For instance, hydrogen—a very flammable substance, and oxygen—a gas required for combustion, combine to form water, which is used to put out fires!

With the number of chemical elements exceeding one hundred, it was quite comforting to discover a hundred years ago that atoms were composed of three more basic particles—electrons, protons, and neutrons. For instance, the most common neutral atom of carbon is composed of six electrons, six neutrons, and six protons. Combinations of only three basic particles determine the myriad of chemical properties that exist in nature.

Beginning in 1932, scientists discovered many new "elementary particles" such as the muon, the pion, and the kaon. In 1932 the first of the "antiparticles" was discovered. The positron is just like an electron, except that it has a positive charge. Over the next few decades more than a hundred new particles were discovered, leading to renewed efforts to find a simpler set of fundamental building blocks.

The elementary particles are grouped into two families—the *leptons* and the *hadrons*. The lepton family has six members: the electron, a heavy electron known as the muon, a still heavier electron known as the tau, and three neutrinos, one associated with each of these "electrons." It is currently believed that these six particles are truly fundamental, as they do not show any internal structure.

Name	*Symbol*	*Charge*	*Other*
Down	d	$-1/3$	
Up	u	$+2/3$	
Strange	s	$-1/3$	Strangeness = −1
Charm	c	$+2/3$	Charm = +1
Bottom	b	$-1/3$	Bottomness = +1
Top	t	$+2/3$	Topness = +1

Table 1

The other elementary particles are composites. Murray Gellmann and George Zweig hypothesized that these particles are combinations of more fundamental particles known as *quarks*. Theorists believed (and still believe) that there should be one quark for each lepton. Therefore, it was very comforting to discover six quarks. In order to account for the known hadrons, the quarks must have baryon number $1/3$, spin $1/2$, and the properties given in table 1.

Let's look at an example of how this works. The proton is a *baryon* with a spin of $1/2$ and charge of +1. All baryons are composed of three quarks—hence the assignment of baryon number $1/3$ to the quarks. Each quark has a spin of $1/2$. If two of these are aligned in one direction and the third is aligned in the opposite direction, they can combine to form a particle with a spin of $1/2$ if the quarks have no orbital angular momentum. This leaves us with obtaining the correct charge. We see that the combination uud has a charge of +1. We could conceivably use some of the other quarks but the proton is the lightest baryon and we expect it to be composed on the lightest quarks, the up and down quarks.

The antiproton is an antibaryon and is composed of three antiquarks. Antiquarks have the same properties as the quarks, but many of these properties have the opposite sign, in particular, the baryon number, the charge, and the "other" properties in table 1. The composition of the antiproton is just like that of the proton except that all of the quarks are replaced by the corresponding anti-

Art by Tomas Bunk


NEUTRONS
ELECTRONS
PROTONS
MUONS
PIONS
KAONS
QUARKS
FUNDA-
MENTAL
BUILDING
BLOCKS
Tomas Bunk
LE MUSEE
QUANTUM
PATAPHYSIQUE
NEW ROCHELLE
1-11-01
NEW YORK

Name	*Symbol*	*Baryon*	*Spin*	*Charge*	*Strangeness*
Neutron	n	+1	$^1/_2$	0	0
Pi minus	π^-	0	0	−1	0
K zero	K^0	0	0	0	+1
Lambda zero	Λ^0	+1	$^1/_2$	0	−1
Antineutron	$\bar{n}$	−1	$^1/_2$	0	0
Xi minus	Ξ^-	+1	$^1/_2$	−1	−2

Table 2

quark. Therefore, the composition of the antiproton is $\bar{u}\bar{u}\bar{d}$, where the overbar indicates an antiquark. This yields a baryon number of −1, a spin of $^1/_2$, and a charge of −1.

The members of the other subfamily of hadrons are known as *mesons*. Mesons are composed of a quark and an antiquark, which yields a baryon number of zero. For example, a positively charged pion has a spin of a zero and a charge of +1. Verify for yourself that $u\bar{d}$ has the correct properties. To get a spin of a zero, the spins of the two quarks must be aligned in opposite directions.

The positively charged kaon is a meson with a spin of zero, a strangeness of +1, and a charge of +1. Its composition is $u\bar{s}$. The neutral pion is an interesting case as it is composed of two combinations, $u\bar{u}$ and $d\bar{d}$.

A. What combinations of up, down, and strange quarks make up the particles in table 2?

There is one problem that we've not mentioned. The omega minus (Ω^-) is a baryon with a spin of $^3/_2$ and a strangeness of −3. The only combination of three quarks that gives the correct properties is sss. In order to get the spin of $^3/_2$ all three of the spins need to point in the same direction. This is the root of the problem. The Pauli exclusion principle says that no two quarks in a hadron can have the same set of properties—that is, none of the quarks can have the same set of *quantum numbers*. But, in the Ω^-, the three quarks have the same set of quantum numbers.

This led to the idea that there must be another quantum number that describes the quarks. This quantum number is known as *color*, and has three values. But, in this case, the values are not numbers. They are called *red*, *green*, and *blue*. If we think of these quantum numbers as combining like colored lights, all hadrons must be *white*. Therefore, one of the strange quarks in the Ω^- must be red, another green, and the third blue.

The antiquarks have the complimentary colors; the antired quark is *cyan*, the antigreen quark is *magenta*, and the antiblue quark is *yellow*. Therefore, the positive pion is a combination of

$$u_{red}\bar{d}_{cyan} + u_{green}\bar{d}_{magenta} + u_{blue}\bar{d}_{yellow}.$$

B. What combinations of up, down, and strange quarks make up the particles in table 3?

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.

Name	*Symbol*	*Baryon*	*Spin*	*Charge*	*Strangeness*
Neutron	n	+1	$^1/_2$	0	0
Antineutron	$\bar{n}$	−1	$^1/_2$	0	0
Pi minus	π^-	0	0	−1	0
Xi minus	Ξ^-	+1	$^1/_2$	−1	−2
Delta plus plus	Δ^{++}	+1	$^3/_2$	+2	0
Antilambda	$\overline{\Lambda^0}$	−1	$^1/_2$	0	+1

Table 3

Curved reality

In the September/October issue of *Quantum*, we began the exploration of shapes found in nature. These shapes included the straight lines of falling objects, the parabolas of trajectories, the circles of charged particles trapped in a magnetic field, the ellipses of planets orbiting the Sun, the hyperbolas of alpha particles scattering off a nucleus, and the cycloid traced out by a rolling wheel.

In Part 1 of the problem, we asked you to derive an equation for a trajectory that has a frictional force proportional to the velocity and to sketch the paths for different values of the proportionality constant b. It is probably best to begin with the one-dimensional problem of an object falling with a retarding force. The total force on the ball is

$$F = mg - bv,$$

where the constant b depends on the size and shape of the object and the viscosity of air. Putting this force into Newton's second law, we have

$$m\frac{dv}{dt} = mg - bv.$$

Separating the variables and assuming that the initial velocity is zero, we obtain

$$\int_0^v \frac{dv}{v - (mg/b)} = -\frac{b}{m}\int_0^t dt.$$

We now integrate and solve for v:

$$v = \frac{mg}{b}(1 - e^{-bt/m}).$$

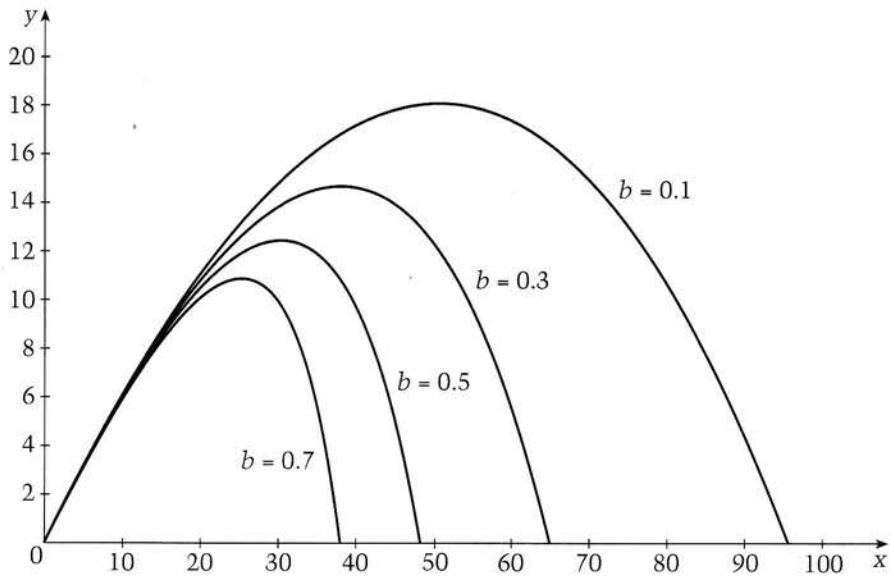


Figure 1

As t gets large, the velocity approaches mg/b, which is called the *terminal velocity*.

In the two-dimensional problem, we realize that the initial velocity will not be zero and that when the object is ascending, the air resistance opposes the rise and when the object is descending, the air resistance opposes the fall. Since the velocity has a horizontal component, there will be an air resistance in the horizontal direction as well as in the vertical direction.

The appropriate equations for an object traveling in the xy-plane are:

$$m\frac{d^2x}{dt^2} = -b\frac{dx}{dt}$$

and

$$m\frac{d^2y}{dt^2} = -mg - b\frac{dy}{dt}.$$

Assuming that the object starts from the origin, the solutions of these equations for velocity and position are:

$$v_x = v_{x0}e^{-bt/m},$$

$$x = \frac{mv_{x0}}{b}(1 - e^{-bt/m}),$$

$$v_y = \left(\frac{mg}{b} - v_{y0}\right)e^{-bt/m} - \frac{mg}{b},$$

$$y = \left(\frac{m^2g}{b^2} + \frac{mv_{y0}}{b}\right)(1 - e^{-bt/m}) - \frac{mg}{b}t.$$

In the same way that we eliminate t and combine the equations to find that trajectories with no air resistance travel in parabolas, we find that the equation for the trajectory with air resistance is

$$y = \left(\frac{mg}{bv_{x0}} + \frac{v_{y0}}{v_{x0}}\right)x - \frac{m^2g}{b^2}\ln\left(\frac{mv_{x0}}{mv_{y0} - bx}\right).$$

We can graph this equation for different values of b using a graphing calculator or a spreadsheet.

As we can see from the graph in figure 1, the trajectory approximates a parabola for low air resistance (b = 0.1) but for larger air resistance (b = 0.7), the path falls off more rapidly than a parabola. If you crumple a piece of paper and toss it at an angle, you will see that the path is quite similar to that of our theoretical calculation.

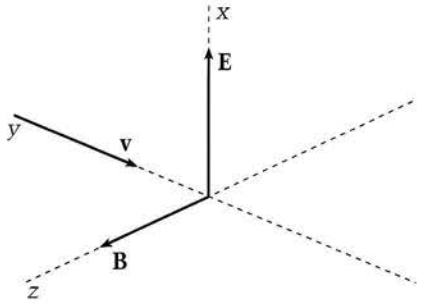


Figure 2

Part 2 of our problem asks about the path of a charged particle entering a region of crossed electric and magnetic fields. If the particle were to travel in a straight line without any deflection, then the speed must be equal to E/B. We can show this quite simply by asserting that for the net force on the particle to be zero, the magnitude of the electric force must equal the magnitude of the magnetic force:

$$qE = qvB,$$

and therefore

$$v = \frac{E}{B}.$$

In part B, we asked what would happen if the particle entered the region of crossed fields in the opposite direction. In this case the particle would no longer travel undeflected since the direction of the electric force would be the same but the direction of the magnetic force would be reversed.

Part C asks what the path of the particle would be if it traveled at a speed other than E/B. Qualitatively, we can look at a laboratory frame that moves at a speed of E/B. In this frame, the particle will move in a circle. We can show this by looking at the general force equations for the particle, changing reference frames, and analyzing the resulting equations.

Using the orientations indicated in figure 2, we have

$$a_x = \frac{q}{m}E + \frac{q}{m}v_yB$$

and

$$a_y = -\frac{q}{m}v_xB.$$

Changing to a reference frame that moves with speed E/B in the negative y-direction requires the following equations:

$$v_x = v'_x$$

and

$$v_y = -\frac{E}{B} + v'_y,$$

CONTINUED ON PAGE 37

From the pages of history

by A. Varlamov

MANY NEWSPAPERS AND magazines have a column where they reprint stories that appeared in their publication a hundred years ago. There the reader is entertained with interesting or even bizarre (to modern eyes) things that went on in the "good old days." *Quantum* has a long way to go before its centenary, so it's a little premature to use such a headline. Still, let's imagine what a magazine like ours would have written a hundred years ago.

At that time the phenomena of superconductivity and superfluidity were not yet discovered, and nobody was aware of the possibility of constructing lasers, thermonuclear reactors, artificial satellites, and rocket ships. So it may seem at first glance that, for inquisitive students in the year 1901, there wasn't anything to read about. And yet they did read, and with great success, because at that very time a generation of remarkable physicists was coming of age—a generation that created modern physics.

Recently I came across an old, beat-up copy of Bruno Donath's *Physikalisches Spielbuch für die Jugend* (*spiel* = to play, *Jugend* = youth—I'll leave it to you to decipher the rest), which was published at the end of the 19th century. In those days the study of physics was based mainly on laboratory demonstrations, which were both engaging and instructive.

Presented below are descriptions of some of the engaging experiments presented in this wonderful book.

A fountain that spurts on command

The design of such a fountain is shown in figure 1. A retort *A* (with a volume of about 1 liter) has two openings, one on top and one on the side. It's placed on a support *B* located near a reservoir *C*, which collects water. Both openings are stopped with rubber plugs in which the curved tubes *a* and *b* are inserted. Tube *b* is the main outlet of the fountain, while tube *a* (whose end almost reaches the bottom of the reservoir) plays the role of an improvised stopcock.

If tube *a* is open, air enters the retort, and water freely runs out of it through tube *b*, thereby raising the level of water in the reservoir. As soon as the water level reaches the opening of tube *a*, the influx of air is stopped and water doesn't run any more—the fountain dries up!

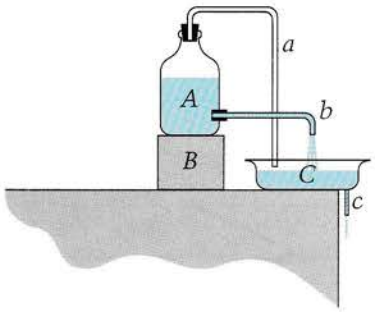


Figure 1

In order to revive it, one must open the end of pipe *a*. Pipe *c*, located at the bottom of the reservoir, serves just this purpose. If its diameter is smaller than that of tube *b*, the flow of water into the reservoir is greater than the flow of water out while the fountain is in operation. This means the opening of tube *a* will be closed at some moment and the fountain will immediately stop working. When enough water has run out of the reservoir, the fountain will start working again.

If we hide the internal mechanism of our fountain from the spectators—for example, by pasting paper around the retort—we can use the demonstration as a trick. What we need is to reverse the cause and effect. Although in reality the behavior of the fountain doesn't depend on what you say, you can peek at the opening of pipe *a* and say, at the appropriate moments: "Fountain, spurt! Fountain, stop!"

Speaking doll

This experiment is based on the wave nature of sound. To perform it, we need to prepare two concave spherical mirrors with a radius of curvature of about one meter.

Since these mirrors must reflect sound instead of light, there is no particular need to polish their surfaces. It's enough to make the size of the wrinkles and irregularities far less than the wavelength of a sound wave. The characteristic wave-

Art by Vasily Vlasov

lengths of human speech range from dozens of centimeters to several meters, so the irregularities of the mirror's surface must not be greater than 1–2 cm. (See if you can come up with a similar estimate of acceptable roughness for optical mirrors.)

It isn't hard to make a sound mirror. Take a piece of cardboard and cut it into identical acute isosceles triangles (the more the better). Construct a pyramidal surface with these triangles by stitching the pieces together (figure 2). Longer tri-

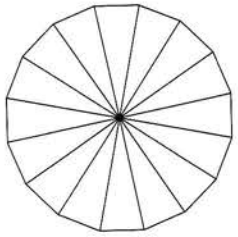


Figure 2

angles will yield a wider mirror. For our experiment it's sufficient to use triangles with lateral side lengths of about 30–40 cm.

Moisten the stitched cardboard to make it soft and stretchable, and then press it against a large flat plate. Finally, flatten the irregularities and finish the surface with a template, which you can also make yourself. Here's how. Draw an arc of radius 1 m on a piece of cardboard about 70 cm long and about 30 cm wide so that the entire length of the cardboard piece is covered (figure 3).

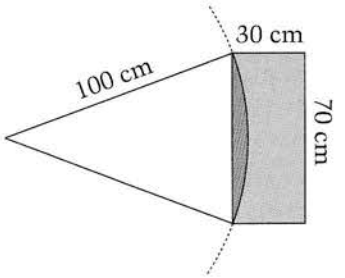


Figure 3

Now carefully cut out the segment of the circle obtained—there's your template. Insert the template into the mirror and form its surface so that you can freely rotate the template without catching on any irregularities. Now put the mirror in the shade to dry.

For our experiment we need two mirrors that are as close to identical as possible. Hang them up on opposite walls in two adjacent rooms, one facing the other with a door in between. The distance between the mirrors can be as much as 10 m. However, the mirrors must be strictly lined up or the experiment won't work. To check this, place an analog watch at the focal point of one mirror (situated at a distance equal to half the radius of the mirror from the mirror's apex). Its ticking should be loudest at the focal point of the second mirror.

After you've tuned the acoustic system, put a doll or statuette at the focal point of one mirror and tell your friends that this doll can answer any question whispered into its ear. You should hide the second acoustic mirror from the audience (for example, screen the doorway with a cloth or brightly illuminate the room containing the doll while leaving the adjacent room darkened).

Of course, you cannot perform the trick alone—your able assistant must be near the distant hidden mirror. In the adjacent room your helper will hear everything that is whispered into the doll's ear and will answer the questions. The answers will be heard near the doll, creating the illusion of a speaking creature.

This isn't hard to explain (when the time comes to explain it). If the source of a sound is at the focal point of one mirror, the sound waves will reflect from it and produce a sound beam traveling parallel to the axis of our system. When this wave arrives at the second mirror, it will be concentrated at the focal point of this mirror.

Even better results can be achieved by your partner with a megaphone instead of a hidden mirror. The megaphone can be used both to intercept the questions and to whisper the answers. In this case, the effect of the trick will be even stronger, because the megaphone not only amplifies the sounds, it also distorts them.

Singing goblet

You can produce various musical sounds with a thin-walled goblet, and not just by tapping it. How? Read on!

Wash your hands with hot water to remove the oils from your fingers. Dip your finger into some water and carefully run it over the rim of the goblet, wetting your fingers repeatedly. At first the goblet will give off unpleasant, squeaky sounds. However, when the brim is well rubbed, the sound will be more pleasant and tuneful. By changing the pressure of your finger, you can vary the tone of the goblet's song. In addition, the pitch of the tone depends on the size of the goblet, thickness of its wall, and amount of liquid in it.

By the way, not every glass can produce pleasant melodic sounds, so your search for a suitable goblet might be long and vexatious. The best melodic (nonsqueaky) sounds are generated by very thin, high-quality goblets in the shape of a paraboloid with a long, thin stem. The pitch can be changed by adding liquid to the glass—more liquid yields a lower pitch.

Here's another curious property: If the water level rises to the middle of the glass, ripples appear on the liquid's surface, generated by the vibration of the goblet's wall. The ripples will be most pronounced near your moving finger.

It's interesting that Benjamin Franklin (1706–1790)—who in addition to his fame as a statesman was known worldwide for his discovery of atmospheric electricity and inventions such as the lightning rod and the modern chimney—created a strikingly original musical instrument based on the singing glass phenomenon. He took a set of well-polished glass cups, drilled holes through their centers, and attached them to a shaft, spacing them equally. This system was rotated by a pedal drive (similar to that of a sewing machine). Touching the rotating cups with wet fingers produced a wide range of sounds, from a fortissimo to a hushed whisper.

It's hard to imagine how this wonderful musical instrument sounded, but those who heard it said the harmony of sounds had a tremendous effect on both the player and the audience. In 1763 Franklin presented this musical wonder to an English woman by the name of Davis. She demonstrated this instrument in many European countries before it disappeared forever without a trace.

A mirror that doesn't confuse "left" and "right"

Take two ordinary flat mirrors without frames and set them at right angles to each other with their edges touching and their reflecting surfaces inside. Now look into such a compound mirror along the bisector. You'll see your own image. Close your right eye—the image will do the same. Lift your left arm up—the image will repeat your movement and raise its left arm. You see, this optical device is a "perfect" mirror—it doesn't confuse "left" and "right." How does it manage it?

The answer is simple. You can't observe the image of your left side in the left mirror, because the law of reflection says that it reflects the beams not to you but to its neighboring mirror. The right mirror produces the final image of the left part of your body, which consequently will be located on the right. In other words, the mirrors "exchange" images, so the final image results from two reflections instead of one produced by a common mirror. Therefore, the left side of an object will be naturally transformed into the left part of the image, and vice versa.

Now let's gradually increase the angle between the mirrors. At first, the middle part of your face with your nose will disappear; then you'll see your own ears only, and at some angle the image will disappear entirely. However, when the angle between mirrors reaches about 180°, the familiar image of your face will appear in the flat mirror.

The experiment can be carried out in reverse order. Obviously the chain of events will be reversed: First your face widens, then your nose swells, your mouth stretches, a third eye appears at the top of your nose... and so on. Try to understand the strange behavior of the image and plot the path of the light rays in this mirror system for various angles between the mirrors. ■

Quantum on amusing physics:

A. Eisenkraft and L. D. Kirkpatrick, "The Clamshell Mirrors," March/April 1992, pp. 49–50; September/October 1992, p. 26.

A. Eisenkraft and L. D. Kirkpatrick, "Fun with Liquid Nitrogen," March/April 1994, pp. 38–40.

S. Kuzmin, "Spinning in a Jet Stream," September/October 1994, pp. 49–52.

I. I. Mazin, "Physics in the Kitchen," September/October 1997, pp. 54–56.

N. Paravyan, "Jingle Bell?" November/December 1997, p. 27.

N. Paravyan, "Amusing Electrolysis," May/June 1998, pp. 41–42.

P. Kanaev, "Suds Studies," July/August 1998, pp. 47–48.

S. Krotov and A. Chernoutsan, "Cold Boiling," January/February 1999, p. 33.

V. Mayer, "Modeling a Tornado," May/June 2000, pp. 42–43.

CONTINUED FROM PAGE 33

where the primed variables refer to the moving reference frame.

Substituting these expressions into the acceleration equations, we get

$$a'_x = \left(\frac{qB}{m}\right)v'_y, \quad a'_y = \left(-\frac{qB}{m}\right)v'_x. \quad (1)$$

These are the equations for a circle (the acceleration is perpendicular to the velocity and the velocity is constant).

If we choose a particle having a velocity other than E/B, the path of the particle will be a cycloid, the same curve as a point on a bicycle wheel rolling with uniform speed E/B. This can be shown by assuming that the particle is at the origin and at rest in the laboratory frame. The particle will begin to accelerate due to the electric force in the x-direction. Once it has a velocity in the x-direction, it will then be deflected by the magnetic force. It will travel in a cycloid finally coming to rest again at a point on the y-axis where its motion will be repeated.

We can prove this mathematically. By "guessing" the solutions to equation 1, we can then transform the solution back to the laboratory frame and analyze the final equations. Our "guess" at the solutions of equation 1 satisfying the initial condition that the speed in the x-direction is 0 and the velocity in the y-direction is E/B are

$$v'_x = \frac{E}{B}\sin\left(\frac{qB}{m}\right)t,$$

$$v'_y = \frac{E}{B}\cos\left(\frac{qB}{m}\right)t.$$

We can take the derivative of these equations and substitute back into equation 1 and see that they are solutions. Transforming these equations back into the laboratory reference frame, we arrive at

$$v_x = \frac{E}{B}\sin\left(\frac{qB}{m}\right)t,$$

$$v_y = \frac{E}{B}\cos\left(\frac{qB}{m}\right)t - \frac{E}{B}.$$

Integrating these equations, substituting $\omega = qB/m$ and $R = E/\omega B$, and using the initial conditions that $x = y = 0$ at $t = 0$ yields:

$$x = r(1 - \cos \omega t),$$

$$y = r(-\omega t + \sin \omega t).$$

These are the equations of a cycloid as shown in the article. The resulting path of the particle will be a cycloid having loops, cusps, or ripples, depending on the initial conditions and on the magnitude of E. ■

PROBLEM PRIMER

Exploring every angle

by Boris Pritsker

HOW MANY SOLUTIONS does a problem have? Why is it important to find different solutions to a single problem? Why aren't we satisfied sometimes with the solutions we've already found, and it is worthwhile to look for another one?

In finding multiple solutions to a problem, we can develop a deeper understanding of the subject matter. Searching for multiple solutions to a given problem helps one develop problem-solving ability and flexibility. The process can also stimulate imagination and creativity, as well as build a broad base of experience from which to draw in solving more difficult problems.

In this article we pose one interesting construction problem and several approaches to its solution.

Problem. *Given an angle with an inaccessible vertex O, arbitrary points K and N on different sides of the angle, and a point M that is interior to the given angle, construct the line passing through points O and M.*

Before considering any solutions to the problem, I'd like to emphasize that point O must not be used—this point is inaccessible. What I find attractive in this particular problem is that one might have to solve it in real life, perhaps in constructing a building, in geological field work, or in surveying (geodesy).

Solution 1. Construct the perpendicular to ON through M and denote the point of intersection by A (figure 1). Construct the perpendicular to OK through M and denote the point of intersection by B. The right triangles OAM and OBM have a common hypotenuse OM; therefore, points A, M, B, and O lie on the circle whose center is at the midpoint of OM and whose radius is $^1/_2OM$. We now find the circumcenter of the triangle ABM, which coincides with the center of circumscribed circle of triangles OAM and OBM. It is the point of intersection of the perpendicular bisectors of AM and BM—point H. Thus MH is the desired line.

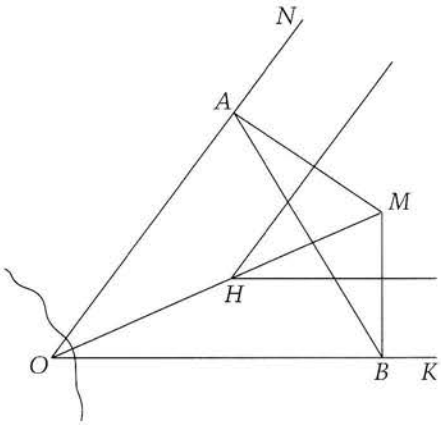


Figure 1

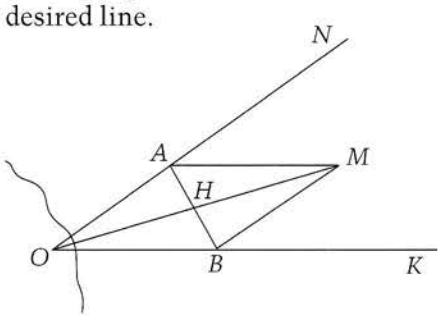


Figure 2

Solution 2. Construct $MA \parallel OK$ and $MB \parallel ON$. Then $OAMB$ is a parallelogram (figure 2). The diagonals of a parallelogram bisect each other. So if we find the midpoint of the diagonal AB (point H), it must be also the midpoint of the diagonal OM. Line MH is the desired line.

Solution 3. Construct $MA \parallel OK$ and $MB \parallel ON$. Draw line $N'K'$ parallel to AB through M (figure 3).

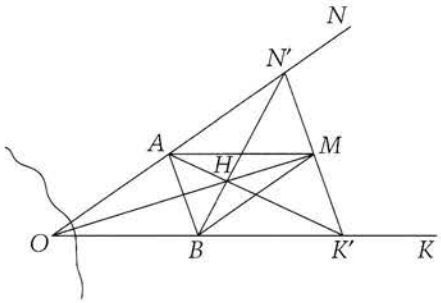


Figure 3

Then we can show that M is the midpoint of segment $N'K'$. Indeed, $AN'MB$ is a parallelogram, so $AB = MN'$, and $AMK'B$ is a parallelogram, so $AB = MK'$. Thus $MN' = MK'$. Similarly, by choosing another pair of parallelograms, we can show that A is the midpoint of ON' and B is the midpoint of OK'. Thus $K'A$ and $N'B$ are medians of triangle $ON'K'$. The point of their intersection H is the centroid of that triangle, and its third median OM must also pass through H. Thus MH is the desired line.

Solution 4. First we prove a lemma:

Lemma: *In any trapezoid, the midpoints of the bases are collinear*

with the point of intersection of the diagonals, and also with the point of intersection of the two non-parallel sides.

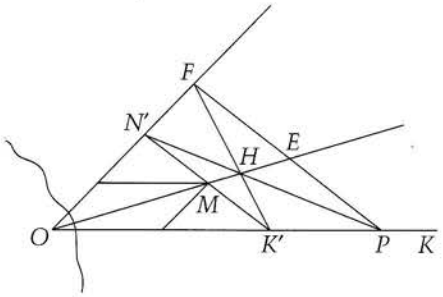


Figure 4

Proof: Let the trapezoid be $K'N'PF$ (see figure 4), and let H and O be the intersection of the diagonals and of the nonparallel sides, respectively. Suppose OH intersects base $N'K'$ at point M. We use a succession of similar triangles.

From similar triangles $K'MO$ and PEO, we have

$$\frac{K'M}{PE} = \frac{OM}{OE}.$$

From similar triangles MON' and EOF, we have

$$\frac{OM}{OE} = \frac{MN'}{EF}.$$

From similar triangles $K'HM$ and FHE, we have

$$\frac{K'M}{EF} = \frac{MH}{EH}.$$

From similar triangles $MN'H$, EPH, we have

$$\frac{MH}{EH} = \frac{MN'}{PE}.$$

Multiplying together these inequalities, and canceling some terms, we find that $K'M^2/(PE \cdot EF) = MN'^2/(PE \cdot EF)$, which implies that $K'M = MN'$, or M is the midpoint of $K'N'$. The proof that E is the midpoint of FP proceeds analogously.

Now we turn to the solution of our problem. Using the method of solution 3, we construct segment $N'K'$ through point M such that M is its midpoint, and its endpoints are on the sides of the given angle (figure 4). Draw an arbitrary segment FP, parallel to $N'K'$ with its endpoints on the sides of the given angle. Then $FN'K'P$ is a trapezoid. If H is the intersection of $N'P$ and FK', then our lemma says that line HM will pass through point O.

Solution 5. Construct an arbitrary triangle MAB with vertices A and B on sides ON and OK, respectively, of the given angle (see figure 5). Pick any point C on segment OA, and

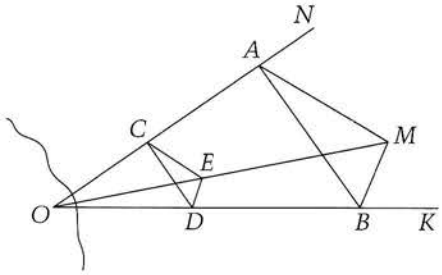


Figure 5

draw CD parallel to AB (where D lies on line OB). Draw the line parallel to AM through C and also the line parallel to MB through D. Suppose these two lines meet at point E. We will show that line ME passes through point O, solving our problem.

Indeed, triangles COD and AOB are similar, so $CD : AB = OD : OB$. Triangles CDE and ABM are similar, so $CD : AB = DE : BM$. Thus $OD : OB = DE : BM$. By construction, $\angle ODE = \angle OBM$. Now triangles ODE and OBM are similar because they have an angle in common, and the sides which include this angle are in proportion. This means that $\angle DOE = \angle BOM$. Since points O, D, and B are collinear, it follows that points O, E, and M must also be collinear, and ME is the required line.

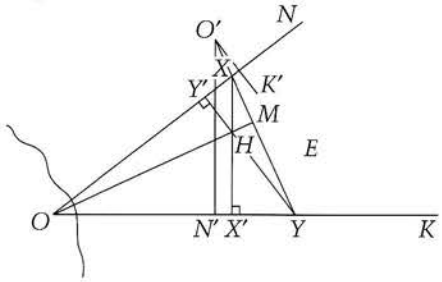


Figure 6

Solution 6. This solution uses rotation by 90° about point M (figure 6). Suppose such a rotation takes the given angle NOK into the angle $N'O'K'$, where N' is on ray OK. (Note that this rotation can be performed without knowing where point O is: We merely rotate the accessible portions of the angle's sides.) Then OM is perpendicular to MO'. Suppose line $O'M$ intersects ray ON at point X, and OK at point Y. Consider triangle OXY, and draw its altitudes XX' and YY'. Since OM is perpendicular to XY (which is the same line as MO'), OM is the third altitude of the triangle. Since the altitudes of triangle OXY intersect at point H, we can draw line MH and it will pass through point O. Thus MH is the desired line.

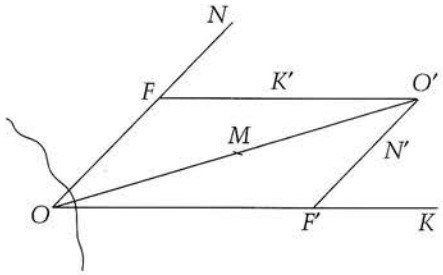


Figure 7

Solution 7. This time we rotate by 180° about point M (figure 7). Angle NOK (even if we cannot access its vertex) is taken into some angle $FO'F'$, where F is on ray ON and F' on ray OK. (Again, we can construct this parallelogram without accessing point O.) Then $OFO'F'$ is a parallelogram, and point M is the midpoint of diagonal OO'. Hence line $O'M$ passes through point O and solves our problem.

In conclusion, it's worthwhile to compare the solutions offered here. In my opinion the most elegant is solution 1, but solutions 2 and 7 are easier to follow and require fewer steps to construct; they also use more elementary ideas. Solutions 3, 4, 5, and especially 6 may seem more difficult. However, they are also very useful in developing thinking skills and creativity.

In discovering new solutions to these problems, we may be led to find interesting generalizations. We can, for example, combine the results of solutions 1, 3, and 6 to show that the circumcenter, centroid, and orthocenter of any triangle are col-

CONTINUED ON PAGE 41

Flights of fancy?

by V. Drozdov

Art by Sergey Ivanov

ALTHOUGH THE USE OF bows and arrows as weapons stopped long ago, bows and arrows still remain favored toys of children and are used in special hunts and Olympic competitions. It was recently reported that an arrow shot from a distance of 536 meters hit the bull's-eye. Compare this value with 50 meters, which is an approximate range for an arrow shot from a homemade bow.

Let's evaluate the capacity of a military bow applying the laws of physics. All the data that are needed to characterize the bow may be found in the literature. Usually bows are subdivided according to the force F needed to draw a bow full length. In the British classification there are bows of small, medium, and large tension. In modern units this subdivision corresponds to forces of 648, 864, and 1079 N. As a rule, the length l of an arrow is 60–100 cm, while its diameter d varies from 0.5 to 1.2 cm.

Let's assume that a bow is an elastic body that obeys Hooke's law. Conservation of energy tells us that

$$\frac{mv_0^2}{2} = \frac{F(l-h)}{2},$$

where v_0 is the initial speed of an arrow with mass m, h is the maximal bending deflection of the bow (for definiteness we assume $h = l/2$).

Clearly, the flight of an arrow is greatly affected by air resistance (otherwise, why would an arrow have a fletching?). Since it is not a simple matter to calculate air resistance precisely, we shall give only a rough estimate that will not differ greatly from the experimental data.

Let's evaluate the maximum range L with the help of the following reasoning. Physics textbooks show that in the absence of air resistance the maximal range of a body is v_0^2/g (prove this on your own). The maximal altitude that can be attained by a body thrown vertically upward at the same speed is $v_0^2/(2g)$, which is half the distance. Let's assume that the same ratio is valid when air resistance is taken into account. This is a reasonable hypothesis, because air resistance affects both the vertical and the horizontal motion.

First, consider the motion of an arrow shot vertically upward. The height of its trajectory can be determined from energy conservation

$$mgH - \frac{mv_0^2}{2} = W,$$

where W is the work performed by air resistance during the upward flight of the arrow. When the speed of the arrow is large (although far less than the speed of sound), the air resistance is proportional to the square of the speed: $F_r = kv^2$. Since we are only estimating the range, we need not calculate the work W precisely. Assume that

$$W = -F_{r\,mean}H,$$

where

$$F_{r\,mean} = \frac{kv_0^2}{2}$$

is the "mean" air resistance.

We then find that

$$H = \frac{v_0^2}{2g + \frac{k}{m}v_0^2}.$$

Accordingly, the maximal range of the arrow is

$$L = \frac{v_0^2}{g + \frac{k}{2m}v_0^2}.$$

It can be easily shown that although the speed affects both the numerator and the denominator, the maximal range increases monotonically with the initial speed (this confirms the validity of our estimation).

The proportionality factor k is usually written as

$$k = \frac{c}{2}\rho S,$$

where ρ is the density of the medium, S is the maximal cross-sectional area of the body in the plane perpendicular to its velocity, and c is a dimensionless factor (as a rule, $c < 1$). Note that the proportionality of the resistance to the product ρSv^2 can be established by dimensional analysis. Note also that the air resistance is frequently mentioned in our discussions in *Quantum* (see the references below).

It is clear from experience that if the velocity of the arrow is parallel to its length, the air resistance is minimal and $S = \pi d^2/4$. In contrast, if the velocity of the arrow is perpendicular to its length, the air resistance is maximal, and $S = ld$. During the flight of the arrow, the angle between the arrow and its velocity constantly changes. What value of S should be used in the calculations? We allow for these factors by setting $c = 1$ and $S = \pi d^2/4$.

In this connection, we note that the head of the arrow has two functions. It is not only a "warhead" but it also increases the range by decreasing the angle between the arrow and its velocity (this can be easily proved experimentally).

For our calculations we use the parameters for three oak arrows (table 1): a short arrow (l = 60 cm, d = 0.5 cm, arrowhead mass 3 g) used with a small-tension bow; medium-length arrow (l = 80 cm, d = 0.85 cm, arrowhead mass 4 g) used with a medium-tension bow; and long arrow (l = 100 cm, d = 1.2 cm, arrowhead mass 5 g) used with a strong-tension bow. We see that the range discussed at the beginning of this article is quite realistic. ■

	short arrow	*medium arrow*	*long arrow*
m (g)	11.24	35.76	84.13
v_0 (m/s)	131.5	98.3	80.1
L (m)	885	655	509.6

Table 1

Quantum on aeromechanics and aerobraking:

A. Mitrofanov, "Against the Current," May/June 1996, pp. 22–29.

A. Eisenkraft and L. D. Kirkpatrick, "A Physics Souffle," July/August 1997, pp. 30–33.

A. Mitrofanov, "Satellite Aerodynamic Paradox," January/February 1999, pp. 18–22.

A. Stasenko, "Gliding Home," March/April 1999, pp. 21–23.

CONTINUED FROM PAGE 39

linear. These three points form the so-called *Euler line*. The reader may enjoy proving that the centroid divides the distance from the orthocenter to the circumcenter in the ratio 2:1. It would be interesting to find several different proofs of this fact.

We learn from solutions 3 and 6 how to construct the segment whose endpoints are on the sides of the given angle, passing through the given point inside the angle such that this point bisects the segment, and how to construct the line passing through a given point inside a given angle and perpendicular to the line connecting the vertex of the angle and the given point. I believe it would be a good exercise to find other solutions to these problems. ■

Using cents to sense surface tension

by Mary E. Stokes and Henry D. Schreiber

WHEN IS A CUP OF WATER really full? Suppose you fill a cup with water so that it appears full, and you think the cup will hold no more. Then, add pennies one at a time to the cup. You discover that you can add quite a few pennies before the water overflows. The cup was not as full as you initially thought.

Experiment 1

Fill a 3-ounce plastic cup with water until the surface of water is level with the rim of the cup. Carefully add one penny at a time until the water overflows. Record the number of pennies that you added.

Imagine the water as consisting of an inconceivably large number of tiny molecules. Further, imagine each water molecule being strongly attracted to other water molecules. Consequently, the molecules exist as an interconnected network, not as isolated molecules. The attraction among water molecules is, in fact, so great that the water's surface tends to repel anything that tries to break into their network.

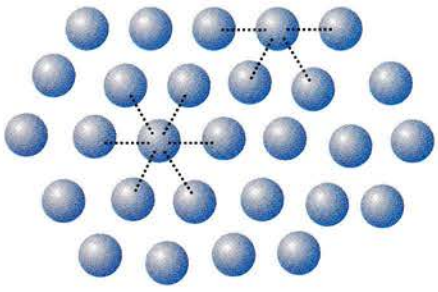


Figure 1. *Intermolecular attractions at the surface versus those in the interior of a liquid.*

A water molecule in the interior of this network experiences strong intermolecular attractions in all directions, as shown in figure 1. However, a water molecule at the surface is different. It is attracted only to neighboring molecules at the surface and to those molecules beneath the surface. The net inward force on the surface molecules results in *surface tension*, or a surface "skin" on the water. The formal definition of a liquid's surface tension is the work required to increase the surface area by one unit. Alternatively, if you move a line segment perpendicular to itself in order to increase the surface area, you can envision the surface tension as the force opposing the line's movement.

The greater the attraction of one molecule to another in the liquid, the greater the liquid's surface tension. Water molecules will do almost anything to keep one or more molecules from breaking loose from the surface. Other liquids, of course, have different surface tensions, which are used as a measure of the intermolecular attractions among the molecules of the liquid.

Experiment 2

Fill a 3-ounce plastic cup with isopropyl alcohol[1] until the liquid is level with the rim of the cup. As in Experiment 1, carefully add one penny at a time to the cup until it finally overflows. Compare the number of pennies used in this case to the number used for water.

Prepare a solution by dissolving 40 grams (2 tablespoons) of table salt in 160 mL (3/4 cup) of water. Fill another plastic cup with this solution until the solution is level with the cup's rim. Add one penny at a time until the solution overflows, and record the number of pennies used.

Note the extent that the liquid bulges up above the cup's rim in each one of these three systems (water, isopropyl alcohol, and salt water). The greater the liquid's surface tension, the more the liquid will bulge up and the more pennies the cup held before overflowing.

Add a drop of liquid detergent to the penny-filled cup of water from your first experiment. What do you observe?

You can make a crude calibration of the liquid's surface tension (g) by the number of pennies needed to overflow the cup. Plot the value of the surface tension (from the table below) on the *y*-axis versus the number of pennies added before the cup overflowed on the *x*-axis.

Liquid	γ (dyne/cm), 20C	*Number of pennies*
salt water (20%)	79	
water	73	
isopropyl alcohol (90%)	22	

[1]Isopropyl alcohol (rubbing alcohol) can be purchased as a 91% solution at most drug stores. To protect the surface of the table from the overflowing liquids, you may wish to put a plate underneath each of the cups.

Is this relationship linear? Prepare different mixtures of isopropyl alcohol and water, and measure their surface tensions by seeing how many pennies are required to overflow the respective cups. Determine whether or not the surface tension is a linear function of the concentration. Finally, measure the surface tensions of other liquids (for example, vinegar or soap solutions) found in your kitchen.

Why are the surface tensions of isopropyl alcohol and salt water different from that of water? How does detergent affect water's surface tension? Does this explain some of the properties of detergent for washing clothes?

In a sense, molecules of a liquid are stuck together by molecular "glue." A liquid like isopropyl alcohol has much weaker "glue" than water, making it easier to penetrate its surface. Because these surface molecules stick together, it is then easy to see why a liquid's surface resists penetration.

Aquatic life evolved around water's surface tension, as debris resting on pond surfaces provides shelter and nutrients. In addition, water's high surface tension allows some insects to walk on water. Even though the insects are denser than water, when their weight is spread across outstretched legs, it does not exert enough pressure to exceed water's surface tension.

Experiment 3

Fill three cups once again so that each is level-full with the water, 91% isopropyl alcohol, and salt water solution, respectively. Use tweezers to carefully place a paper clip on the surface of the water. You should find that, indeed, even though the paper clip is more dense than water and will dent the water's surface, it will not penetrate the surface of water.

Predict, then determine, whether or not you can float the paper clip on the surface of the other two liquids. As the paper clip rests on the surface of water, add a drop of liquid detergent and observe what happens.

All liquids, in the absence of other forces, tend to minimize their surface area. Because a sphere has the lowest ratio of surface to volume, freely suspended volumes of liquids assume a spherical shape. Once again, this is a result of the intermolecular attractions of the liquid's molecules. For example, when you place water on a freshly waxed car, the water beads into near-spherical droplets instead of spreading over the entire surface. Water molecules would rather stick to themselves than to the wax. However, water molecules do have strong interactions with glass and fabric surfaces. Consequently, water molecules will spread over, or wet, these surfaces. Whether a liquid will form a droplet on a surface or wet a surface depends on whether the liquid's surface tension is greater than the attractive forces between the liquid and the other surface.

Experiment 4

Determine the shape of a drop of various liquids on different surfaces. Use a medicine dropper to place one drop of each liquid (water, salt water, and isopropyl alcohol) on a piece of wax paper, glass, and aluminum foil. Sketch the shape of the drop, and measure the angle of the drop with respect to the surface as shown in figure 2. Can you relate the magnitude of a liquid's surface tension to its droplet shape? How might detergent affect water's ability to wet a surface?

Each of these experiments has illustrated that the surface molecules of a liquid are, in a way, different from the molecules in the bulk of the liquid. The difference results in an inward pulling force on the surface molecules, resulting in a surface tension for the liquid. Furthermore, throughout these experiments, you have envisioned liquids as consisting of lots of submicroscopic molecules with varying degrees of intermolecular attractions. *You have imagined molecules!* And you have used this molecular model to understand your observations of the surface tension of liquids. ◙

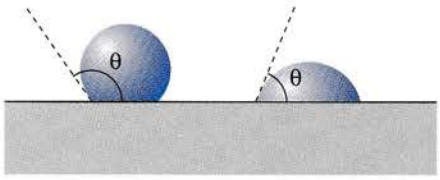


Figure 2. *Droplets of different liquids on a surface. The one on the left does not wet the suface while the one on the right wets the surface.*

Henry D. Schreiber *is professor and head of the Department of Chemistry at the Virginia Military Institute.* **Mary E. Stokes** *is a junior at Rockbridge County High School.*

CONTINUED FROM PAGE 17

M317

Hyperbolic equilateral. The points $M(x_0, y_0)$ and $N(-x_0, -y_0)$ are given on the hyperbola $y = 1/x$. The points are symmetric about the coordinate origin. A circle centered at M is drawn through the point N. This circle intersects the hyperbola at three other points. Prove that these points are the vertices of an equilateral triangle. (V. Senderov)

M318

N for n. Prove that, for any natural n, the number

$$2^{2^n} + 2^{2^{n-1}} + 1$$

has at least n different prime divisors. (N. Vasilyev and V. Senderov)

M319

Shifting stones. A circle is divided into n sectors. Some of the sectors are occupied by stones; the total number of stones is $n + 1$. The arrangement of the stones is then transformed according to the following rule: Two arbitrary stones located in the same sector are chosen and moved to the two next sectors on the left and on the right. Prove that after a certain number of such transformations at least half of the sectors will be occupied by stones. (N. Konstantinov and N. Vasilyev)

ANSWERS, HINTS & SOLUTIONS ON PAGE 50

Lunar launch pad

by A. Stasenko

VOLCANOES ARE VERY INteresting natural objects—majestic and terrifying. Here's what one encyclopedia says: "Krakatoa is an active volcano in the Sunda Strait between the islands of Java and Sumatra. It is 813 meters high. In August 1883 this volcano erupted with exceptional violence. The explosion destroyed more than half of the volcanic island, and it was heard more than 3,000 km away. It generated a huge tidal wave (tsunami) that killed more than 36,000 people on the shores of Java and Sumatra. The volume of ejected material was about 19 km^3. Launched to an altitude of 80 km, the volcanic ashes were suspended in the air for several years."

A book on cosmic gas dynamics has this to say about the possibility of asteroids being created by volcanoes: "Of particular theoretical interest is the action of routine volcanic explosions, which can result in spectacular catastrophic eruptions similar to that of Krakatoa. Such eruptions are not exceptions—they are the natural result of physical and chemical processes inside the Earth. The high speeds of the ejected gas, which undoubtedly exceed several kilometers per second, explain the great heights attained by the column of ejected material, sometimes reaching 60 km. ...In some cases, when the initial speed of the material reaches 11 km/sec, it is expelled beyond the limits of Earth's gravity."

"...and, lo, there was a great earthquake; and the sun became black as sackcloth of hair, and the moon became as blood; and the stars of heaven fell unto the earth, even as a fig tree casteth her untimely figs, when she is shaken of a mighty wind. And the heaven departed as a scroll when it is rolled together; and every mountain and island were moved out of their places."

—The Revelation of St. John the Divine

So, could it be possible that a volcano has given birth to a satellite of the Earth or Sun? Let's see.

Let's assume that a randomly torn-off piece of basalt, lava, or some other volcanic material is moving upward through the vertical shaft of a volcano (figure 1 on page 46). True to form as physicists, we'll also assume a perfectly cylindrical shaft and a spherical projectile. To top it off, both of them have the same radius R. The projectile is accelerated by the pressure P of the volcanic gases, which is far greater than atmospheric pressure (according to the estimates in the aforementioned book, the pressure in a volcanic explosion is about a hundred thousand atmospheres).

At the "muzzle" of the volcano—that is, the crater—the projectile reaches its greatest speed v_0, after which the projectile will be slowed by Earth's gravity and air resistance. Note that we don't dare utter the proverbial words "neglect air resistance," because the initial speed of the volcanic projectile must be greater than the escape velocity

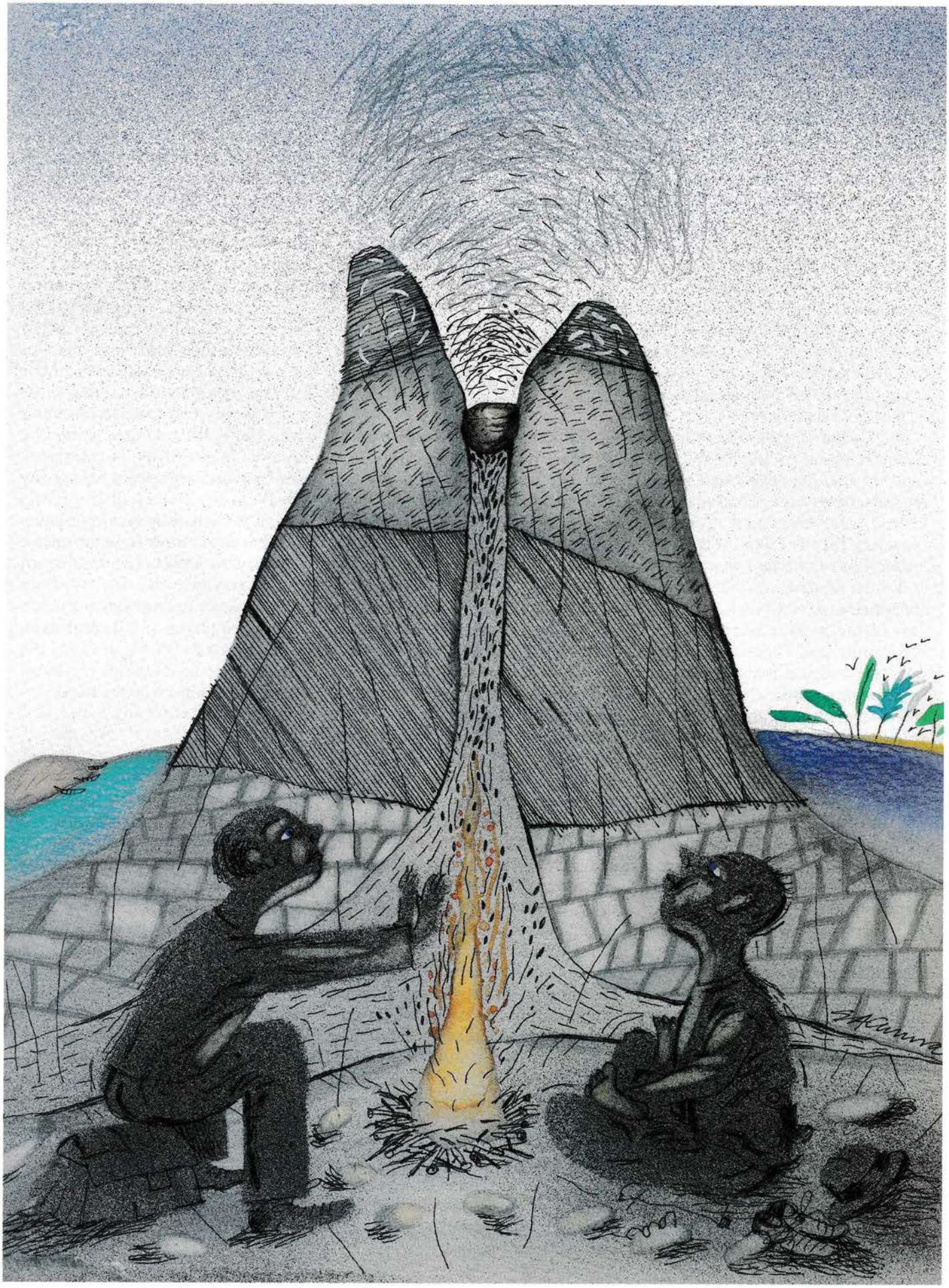
$$v_{esc} = \sqrt{2gR_E} \approx 11 \text{ km/s}, \quad (1)$$

where R_E is Earth's radius. Therefore, the projectile must move at supersonic speed. Indeed, approximating the speed of sound as $c = 300$ m/s, we get the following speed ratio (known as the Mach number):

$$\frac{v_0}{c} \geq \frac{v_{esc}}{c} = \frac{11 \cdot 10^3 \text{ m/s}}{3 \cdot 10^2 \text{ m/s}} \sim 40.$$

Here we're in the realm of *hypersonic* speeds, where no designer of flying objects would ever think of neglecting air resistance.

Art by Ekaterina Silina



What's the magnitude of the air resistance? Time and again we've deduced it using dimensional analysis. It depends on the air's density ρ, the cross-sectional area of the moving object S, and its speed v (check for yourself that the units are newtons on both sides of the equation):

$$F = C\rho S v^2. \qquad (2)$$

Alas, the dimensionless coefficient C can't be found by dimensional analysis (after all, it's dimensionless). But a pleasant surprise lies in store for us: Sir Isaac Newton took an interest in it—for hypersonic motion, his theoretical calculations give a value of $C = 1/2$.

Well, what happens after the projectile is "shot" from the volcano? As it moves along the y-axis its potential energy increases while its kinetic energy decreases. It's tempting to say that their sum is constant (according to energy conservation) and obtain equation (1). However, air resistance acts on the shell, so part of the kinetic energy is converted into heat. Therefore, we must take into account that the decrease in total mechanical energy in a small segment Δy of the trajectory is equal to the work performed by air resistance in this segment:

$$\Delta\left(\frac{mv^2}{2} + mgy\right) = -F\Delta y. \qquad (3)$$

It can be shown that the kinetic energy of the projectile outside the atmosphere (at an altitude of, say, 100 km, which is characteristic of artificial satellites orbiting our planet) is far greater than its potential energy. Indeed, assuming $v = v_{esc}$, $g = 10$ m/s^2, and $y = 10^5$ m, we find that $v^2/2$ is about twenty times greater than gy. If we take into account the fact that the initial speed v_0 must be greater than v_{esc} to "punch through" the layer of atmosphere, the ratio of kinetic to potential energy must be even greater than 20. Therefore, the second term in the parentheses in equation (3) can be neglected (yielding an error of no more than a few percent).

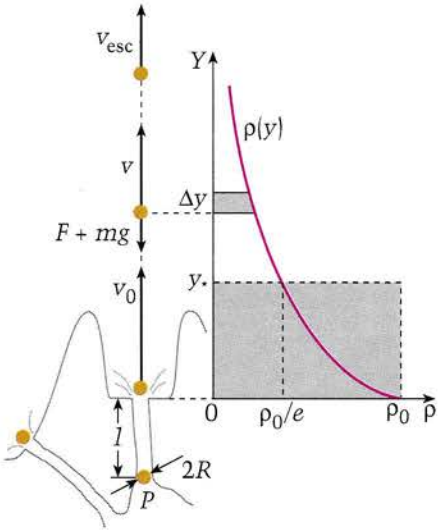


Figure 1

The force of air resistance given in equation (2) varies not only due to a change in speed but also due to a drop in air density, which is described by Boltzmann's barometric formula

$$\rho = \rho_0 e^{-y/y_*}, \qquad (4)$$

where ρ_0 is the atmospheric density at sea level ($y = 0$) and y_* is the characteristic thickness of the atmosphere (about the height of Mount Everest), where the air's density is less than ρ_0 by a factor of $e \cong 2.7$. This dependence is shown on the right side of figure 1. Clearly, the density of the Earth's atmosphere decreases very quickly (exponentially, as physicists say) with altitude.

Equation (4) is rather interesting: if we want to make the mass of an infinite vertical atmospheric column of variable density equal to the mass of a column of finite height and constant density ρ_0, we obtain the value y_* for this height. In other words, the area of the rectangle $\rho_0 y_*$, in figure 1, equals the area under the curve $\rho(y)$. We'll make use of this fact immediately.

Taking everything we've said into account, equation (3) can be written in the form

$$\frac{\Delta v^2}{v^2} = -\frac{S}{m}\left(\rho_0 e^{-\frac{y}{y_*}}\Delta y\right). \qquad (5)$$

A mathematician will recognize this expression as a simple differential equation with separated variables (the left side contains only v^2 and the right side only y). A businessman of the new stripe will also see something familiar—some sort of complex bank percentage (for kinetic energy instead of money).

Well, whatever it looks like, the time has come to solve it. Where to begin? First of all, we notice that the parentheses contain the elemental area $\rho(y)\Delta y$ shown in figure 1. This means that as the volcanic projectile rises along the y-axis, this area will sweep out the entire area under the curve $\rho(y)$, which is equal to $\rho_0 y_*$, as noted above. And somewhere not far from the Earth's surface (since the layer of atmosphere is relatively thin), air resistance stops having any effect on the energy of the rising object. At this (not very high) altitude the object must have the escape velocity to break the chains of Earth's gravity.

But what's happening on the left side of equation (5)? Integration along the y-axis (that is, piercing the atmospheric layer) yields the natural logarithm of v^2, so we have

$$\ln\frac{v_{esc}^2}{v_0^2} = -\rho_0 y_* \frac{S}{m}. \qquad (6)$$

On the other hand, the initial speed of the projectile emerging from the volcano's crater results from the effect of the pressure PS on the mass m. Assuming the pressure to be constant and neglecting friction arising from interaction with the inside wall of the volcano, we get a constant acceleration $a = PS/m$ imparted to the shell in the volcano's shaft. Therefore, according to the laws of uniformly accelerated motion, the object acquires the following specific kinetic energy (or energy per unit mass) while being accelerated along a shaft of length l:

$$\frac{v_0^2}{2} = al = \frac{PS}{m}l.$$

Plugging this equation into equation (6) and rearranging a few things, we get

$$\frac{v_0^2}{v_{esc}^2} = \frac{2Pl\frac{S}{m}}{v_{esc}^2} = \exp\left(\rho_0 y_* \frac{S}{m}\right). \quad (7)$$

There is something attractive about dimensionless parameters, so let's introduce two such parameters for the sake of beauty and simplicity:

$$x = \rho_0 y_* \frac{S}{m};$$

$$k = \frac{2Pl}{\rho_0 y_* v_{esc}^2}.$$

The first parameter depends on the characteristics of the projectile

$$\frac{S}{m} = \frac{\pi R^2}{4\pi R^3 \rho_S/3} = \frac{3}{4\rho_S R},$$

where ρ_S is its density. The second parameter depends on the volcanic "gun": on its length l and its pressure P. Now, using these new dimensionless parameters, we rewrite equation (7):

$$kx = e^x. \quad (8)$$

The left side describes a straight line, while the right side is our old friend, the exponential function. Both functions are plotted in figure 2. We can see that at small values of k (when, for example, the pressure of the accelerating gases P or the length of the shaft l are small), a solution to equation (8) doesn't exist: the dashed line doesn't cross the exponential function.

However, at some value k_0 there is a single tangent point, for which $x_0 = 1$ and $k_0 = e$ (check this by plugging these values into equation (8)). From this we obtain all the values we're interested in:
the radius of the projectile

$$R_0 = \frac{3}{4}\frac{\rho_0 y_*}{\rho_S},$$

the necessary length (depth) of the shaft

$$l_0 = e\frac{\rho_0 y_* v_{esc}^2}{2P},$$

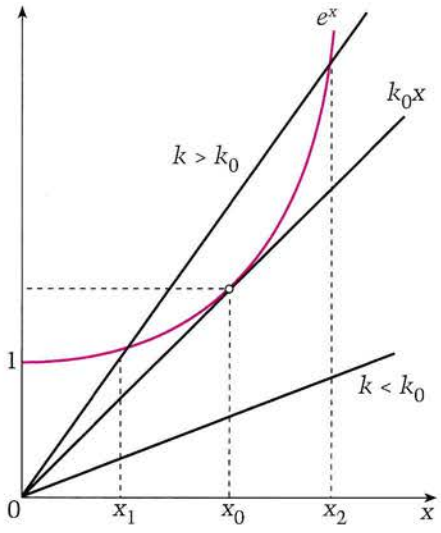


Figure 2

and the initial ejection speed

$$v_0 = \sqrt{e}v_{esc} \ (\approx 1{,}65v_{esc}),$$

needed to have escape speed after punching through the atmosphere.

(It should be noted that at such a speed it's not easy for the volcanic gas, which is providing the impetus, to keep up with the projectile. Even more problematic is preserving a constant pressure during the volcanic "shot put." In a more sophisticated theory we should consider the irreversible expansion of a rapidly heated gas, which is characterized by rapid changes in temperature and pressure exerted on the moving projectile. I hope our readers will eventually be able to handle such complications as they continue their studies.)

But what about the sphere and the shaft? Plugging the values $\rho_0 \sim 1$ kg/m^3, $y_* \sim 10$ km $= 10^4$ m, $\rho_S \sim 5 \cdot 10^3$ kg/m^3, and $p \sim 10^5$ atm $= 10^{10}$ N/m^2 into our answers, we get

$$R_0 \approx 1.5 \text{ m}, \ l_0 \approx 160 \text{ m}.$$

The mass of such a sphere is

$$m_0 = \frac{4}{3}\pi R_0^3 \rho_S \approx 70 \text{ t}.$$

Not bad for an artificial satellite!

However, our equation (8) also has other solutions. For example, if $k > k_0$, the corresponding straight line in figure 2 crosses the exponential curve in two points. The two roots x_1 and x_2 correspond to heavy and light spheres, because $x \sim 1/R$. The heavy sphere can be launched at a lower speed than v_0, while the launching velocity of the light sphere must be greater than this value. The reason is obvious: Air resistance is far less important for a stone than for a feather.

There's another side to the problem. We considered only a vertical "launch." Naturally, a volcano could launch a projectile along an inclined shaft (figure 1), thereby increasing the number of satellites orbiting the Earth. Less speed is needed to perform this task: It's equal to the orbital speed $v_{orb} \sim 8$ km/sec). The reader is invited to investigate this case independently.

What would a cautious professional physicist conclude after sorting through the numerous simplifications in our reasoning? "Well, if it's possible to provide volcanic gases at a constant pressure of about 100 atm along a shaft about 100 m long, then perhaps a volcano could eject from its crater an object with a mass of about 100 t and give it enough speed to launch this projectile to infinity. That's assuming, of course, you can find an object that can endure an acceleration of ten thousand g's."

Whether the volcanic catapult is just a mental toy or a real phenomenon is an open question, but the awesome spectacle of a volcanic eruption has certainly given us some interesting physics problems. ◙

Quantum on space travel:

Y. Osipov, "Catch as Catch Can," January/February 1992, pp. 38–43.

A. Stasenko, "From the Edge of the Universe to Tartarus," March/April 1996, pp. 4–8.

A. Byalko, "A Flight to the Sun," November/December 1996, pp. 16–20.

V. Surdin, "Swinging from Star to Star," March/April 1997, pp. 4–8.

V. Mozhaev, "In the Planetary Net," January/February 1998, p. 4–8.

I. Vorobyov, "High-Speed Hazards," May/June 2000, p. 24–26.

Electrical and mechanical oscillations

by A. Kikoyin

IN 1853 THE FAMOUS BRITISH physicist William Thomson (later Lord Kelvin) published a paper titled "On Nonstationary Electric Currents." In this paper he showed that a circuit composed of a capacitor with capacitance C and a coil with inductance L (the so-called

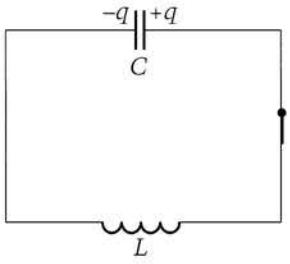


Figure 1

LC-circuit—figure 1) must produce electromagnetic oscillations with a period $T = 2\pi\sqrt{LC}$.

The word "oscillation" usually conjures up the oscillating mathematical pendulum (figure 2) or a mass attached to a spring (figure 3).

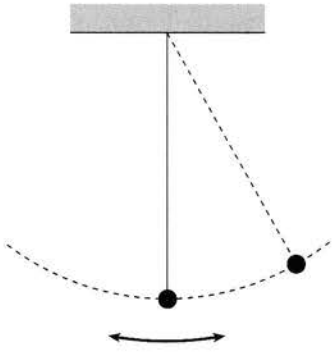


Figure 2

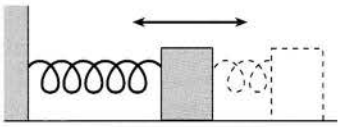


Figure 3

However, the term "oscillation" does not only refer to the mechanical motion of some material body. This term is used by physicists in a broad meaning encompassing any periodic variation of any parameter. Periodicity means repetition of the value of some parameter after a definite interval of time, called the "period" of the oscillations.

The mechanical oscillation of an object is the most obvious form of a periodic process. In this case, the periodic parameter is the coordinate x of the moving point.

In an LC-circuit, one of the periodic parameters is the electric charge q on the capacitor. It's assumed that the capacitor was initially charged by a battery and then connected to the coil. During each period, the charge on the capacitor's plates gradually vanishes, changes its sign, and then increases. Other periodic variables are the voltage $V = q/C$ across the capacitor and the current I in the circuit. The periodic change of the latter may be obvious for those who know a bit of differential calculus, because

$$V = L\frac{\Delta I}{\Delta t}. \qquad (1)$$

(If you haven't encountered such equations, just keep reading and compare equation (1) and equation (2) below.) The magnitude of the electric field E in the capacitor also oscillates (because $E = V/d$, where d is the distance between the plates), as does the magnetic field B within the coil (since it's proportional to the current I).

This is similar to what we know about mechanical oscillations, where not only the coordinate of an oscillating object varies periodically, but also its velocity, acceleration, kinetic energy, potential energy, and so on, change in a similar way. We might say that all these parameters "oscillate."

In a physics textbook you may come across an analogy between mechanical and electrical oscillations. This analogy can be made strict by noting that equation (1) has the same form as Newton's second law formulated for an object oscillating on a spring:

$$F = m\frac{\Delta v}{\Delta t}, \qquad (2)$$

where $\Delta v/\Delta t = a$ is the acceleration of the object. Therefore, equation (1) can be viewed in light of the well-known equation (2) according to the rule that "identical equations have identical solutions."

A comparison of equations (1) and (2) shows that the voltage V

corresponds to the elastic force F, the current I in the coil to the velocity v of the moving body, and the inductance L of the coil to the mass m of the body. Finally, the equations

$$V = q/C \text{ and } F = kx$$

show that the reciprocal of the capacitance $1/C$ corresponds to the spring constant k, and the charge q corresponds to the string's displacement x (remember that $I = \Delta q/\Delta t$, $v = \Delta x/\Delta t$). You might construct a table of the corresponding values within the framework of the electromechanical analogy and compare it with that given in physics textbooks.

Analogy is a very powerful instrument. The brilliant Irish mathematician Sir William Hamilton (1805–1865) used an analogy between optics and mechanics that helped him formulate classical mechanics in an elegant way. Almost a century later this analogy helped the outstanding Austrian physicist Erwin Schrödinger (1887–1961) formulate the basic equation of quantum mechanics (the famous Schrödinger equation).

To conclude, let's look at an interesting question involving the current in an oscillating circuit. Despite the fact that the circuit is "open" (there is no conducting material between the capacitor's plates), charge nevertheless flows in such an open circuit. Moreover, if the resistance of the coil and the connecting wires were zero (that is, cooled to the superconducting state), the current induced by discharging the capacitor would oscillate forever.

Strange as it may seem, the oscillating current in an LC-circuit can be considered "closed." To understand this paradoxical view, let's consider the phenomenon of electromagnetic induction. The British physicist Sir James Clerk Maxwell (1831–1879) was the first to show that the essence of this phenomenon is not generating *current*, but inducing a vortical electric *field* (that is, a field with closed lines of force) by a varying magnetic field. Therefore, magnetically induced current is a phenomenon that is secondary to the generation of the electric field that drives the charged particles in the conductors. Maxwell brilliantly flipped this formulation on its head: He proposed that the opposite is also true—that any varying electric field generates a magnetic field.

In our LC-circuit there's a place where there is nothing more than a varying electric field. This is the space between the plates of the capacitor. According to Maxwell, this capacitor must be surrounded by a magnetic field (figure 4). On the

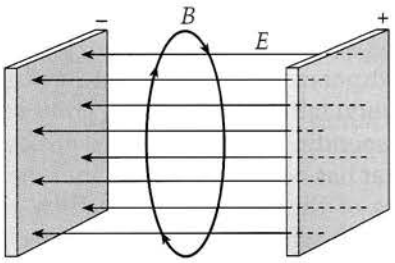


Figure 4

other hand, we know that a current generates a magnetic field around it. So a varying electric field is analogous to moving charged particles (that is, to current). Since Maxwell's time the rate of change of a variable electric field is called a current—specifically, *displacement current*.

So here's how we might describe what's happening: "ordinary" current in the conducting part of the oscillating circuit continues in the capacitor as chargeless current (displacement current). ◙

Quantum on the kinship of electric and magnetic fields:

S. M. Rytov, "From the Prehistory of Radio," May 1990, pp. 39–42.

A. Chernoutsan, "Michael, Meet Albert," September/October 1993, pp. 43–44.

A. Leonovich, "Surfing the Electromagnetic Spectrum," January/February 1995, pp. 32–33.

P. Bliokh, "The Advent of Radio," November/December 1996, pp. 4–9.

V. Dukov, "Convection and Displacement Currents," March/April 1999, pp. 4–8.

CONTINUED FROM PAGE 29

In the process, the corresponding number of zeros is added on the right. Then the results are added. The advantage of this method is that the most significant digits are obtained first, which is important for approximate calculations.

The second method is called *jealousy*. In this method, a grid is drawn in which the results of intermediate calculations are written (in fact, these results are extracts from the multiplication table). The grid is a rectangle divided into cells, each divided by a diagonal (figure 9). Such

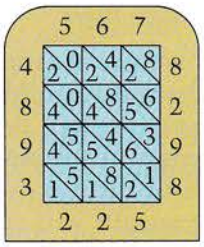


Figure 9

a grid looks like the latticed shutters that were attached to Venetian windows to prevent passers-by from seeing women sitting at the windows, according to Luca Pacioli.

Let's multiply 567 by 3,984 using this method. One factor is written at the top of the grid and the other on its left. Then we write the product of the factors' digits in the corresponding row and column in every cell. Tens are written in the left lower triangle and units in the right upper one. After the grid has been filled, the numbers along the diagonals are added. This method is very simple. Indeed, the cells are filled directly from the multiplication table, and it remains only to add them.

The six other methods described by Pacioli, as well as the first two, are based on the multiplication table. There are many other methods invented in different countries and at different times, but I don't know of any method, except for the Russian one, that doesn't use the multiplication table. ◙

—Anatoly Savin

ANSWERS, HINTS & SOLUTIONS

Physics

P316

First, note that $T > mg$; otherwise equilibrium would be impossible. At the equilibrium position the string is stretched by

$$\Delta l_0 = mg/k. \qquad (1)$$

When the weight is dropped from height x above the equilibrium position, it starts to fall downward. After the position corresponding to the length of the free (unloaded) string, the weight will stretch the string. Denote by Δl the maximum elongation of the string. The condition of breakage of the string is

$$k \cdot \Delta l = T \Rightarrow \Delta l = T/k. \qquad (2)$$

Depending on the relationship between T and m, two cases are possible: (1) $x < \Delta l_0$ and (2) $x > \Delta l_0$. Let's consider each case.

(1) $x < \Delta l_0$. We use energy conservation, taking the zero of potential energy at the position of the unloaded string:

$$-mg(\Delta l_0 - x) + \frac{k}{2}(\Delta l_0 - x)^2$$
$$= -mg \cdot \Delta l + \frac{k}{2}(\Delta l)^2.$$

Plugging in the values Δl_0 and Δl from equations (1) and (2), we get x:

$$x = \frac{T - mg}{k}.$$

Clearly the condition $x \le \Delta l_0$ is equivalent to $T \le 2mg$.

(2) $x > \Delta l_0$. The energy conservation yields

$$mg(x - \Delta l_0) = -mg \cdot \Delta l + \frac{k}{2}(\Delta l)^2.$$

Solving this equation together with (1) and (2) we get

$$x = \frac{mg}{2k}\left(\left(\frac{T}{mg} - 1\right)^2 + 1\right) \text{ at } T \ge 2mg.$$

P317

The melting point of ice does indeed drop under increased pressure. However, melting requires energy, so the temperature under the wire drops. This process goes on until the temperature of the pressurized ice region falls to the melting point corresponding to the increased pressure that has been exerted. Further melting of the ice is controlled by the rate at which heat is conducted to the low-temperature region.

In the case of the metal wire this heat will be conducted efficiently from the water freezing above the wire, so the cutting will proceed rapidly. In contrast, the nylon thread has a negligible thermal conductance, so the heat will be transferred mainly due to cooling of the entire block of ice. Therefore, the cutting will be very slow.

P318

In this bit of kitchen physics we'll neglect the heat capacity of the vessel, which is quite reasonable if it has thin walls and its mass is small compared to that of the liquids. In addition, the specific heats of metals are considerably smaller than that of water (but since this thermal parameter wasn't mentioned in the statement of the problem, and since we have no intention of calculating it, we'll just ignore it).

The heat transferred through a partition per unit time is known to be proportional to the contact area and the temperature drop across it. In our problem the contact area of each pair of liquids is identical, so the following quantities of heat transferred:

(1) from soup to stewed fruit

$$Q_1 = k(65° - 35°),$$

(2) from soup to kvass

$$Q_2 = k(65° - 20°),$$

and

(3) from stewed fruit to kvass

$$Q_3 = k(35° - 20°),$$

where the constant k is the coefficient of proportionality.

Thus the heat loss for the soup is

$$Q_1 + Q_2 = k \cdot 75°,$$

while the heat gain for the stewed fruit and kvass are

$$Q_1 - Q_3 = k \cdot 15°$$

and

$$Q_2 + Q_3 = k \cdot 60°,$$

respectively. Taking into consideration the double mass of the soup, and comparing the amounts of heat lost and gained, we find the temperature increase of the stewed fruit

$$\Delta t_2 = \frac{\Delta t_1 \cdot 15 \cdot 2}{75} = 0.4°\text{C}$$

and that of the kvass

$$\Delta t_3 = \frac{\Delta t_1 \cdot 60 \cdot 2}{75} = 0.6°\text{C},$$

where $\Delta t_1 = 1°\text{C}$ is the drop in temperature for the soup.

In principle, it's possible to make more precise calculations. The decreases in temperature across the partitions varied during the heat exchange processes, which means that the above formulas for Q_1, Q_2, and Q_3 are only approximations. However, the given temperature decrease for the soup (1°) is far less than all the temperature differences in this problem, so the corrections aren't that significant. In any case, they would affect the final result far

less than our neglecting of the heat capacity and heat transfer in the metal vessel.

P319

Electric current starts in the loop after the switch S is closed. However, this current cannot increase immediately due to the presence of the inductor, which, like any honest and reliable inductor, always dislikes variations in the current through it (although it's quite loyal to any constant current). Still, the current will increase steadily to some maximum value.

This current charges the capacitor. According to energy conservation, the following equation is valid at any given moment:

$$\frac{LI^2}{2}+\frac{CV^2}{2}=q\mathcal{E}=CV\mathcal{E}, \qquad (*)$$

where I is the electric current, L is the inductance, $LI^2/2$ is the magnetic field energy stored in the inductor, V is the voltage drop across the capacitor, and $CV^2/2$ electric field energy stored in the capacitor. In addition, $q = CV$ is the charge supplied by the battery, which performed the corresponding work $q\mathcal{E}$.

The voltage across the capacitor attains the value $\mathcal{E}$ at some value I_0 of electric current. At this instant the battery cannot drive current to the capacitor, so from now on the coil will take the leading role. As usual, it resists any changes in the current (and so it "tries" to maintain its present value). Therefore, the charging of the capacitor proceeds at the expense of the magnetic field of the coil. Since this source of energy is limited, the current will gradually fade.

Thus I_0 is the maximum current in the circuit. Equation (*) with $V = \mathcal{E}$ yields

$$I_0=\mathcal{E}\sqrt{C/L}.$$

Charging of the capacitor will continue until the current drops to zero. Thus the maximum voltage across the capacitor is given by (*) at $I = 0$:

$$V_{max} = 2\mathcal{E}.$$

P320

Assume that the layer in which total internal reflection occurs hovers at a distance R_0 from the Earth's center (point O in figure 1) and has a thickness ΔR. The refractive index of this layer is n_0. Since the layer is thin, we can consider that the refractive index n of the atmosphere doesn't vary within this layer (that is, over a distance ΔR).

Consider the beam AB directed to the external boundary of the sparkling layer, which is tangent to its internal boundary at point A. If this beam undergoes total internal reflection at point B on the external boundary, any other beam traveling within the sparkling layer will also be reflected at the external boundary.

The condition of total internal reflection for beam AB is

$$\frac{\sin\alpha}{1}=\frac{n_0+\Delta n}{n_0}=1+\frac{\Delta n}{n_0},$$

where $n_0 + \Delta n$ is the refractive index outside the upper boundary of the sparkling layer. Since this layer is thin, the geometry can be simplified to yield

$$\sin\alpha=\frac{R_0}{R_0+\Delta R}$$
$$=1-\frac{\Delta R}{R_0+\Delta R}\approx 1-\frac{\Delta R}{R_0}.$$

Therefore,

$$1+\frac{\Delta n}{n_0}=1-\frac{\Delta R}{R_0},$$

from which we get

$$\frac{\Delta n}{n_0}=-\frac{\Delta R}{R_0}$$

or

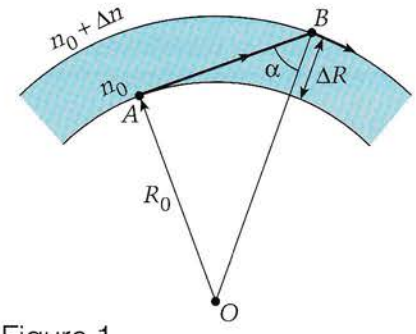


Figure 1

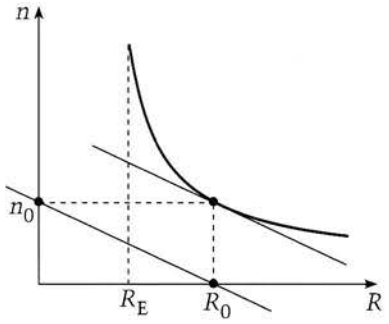


Figure 2

$$\frac{\Delta n}{\Delta R}=-\frac{n_0}{R_0}=n'(R_0).$$

Thus the tangent to the graph $n(R)$ at the point corresponding to the values n_0 and R_0 must be inclined at an angle ϕ, for which

$$\tan\varphi=-\frac{n_0}{R_0}.$$

In other words, this tangent line must be parallel to the line crossing the points with coordinates $(0, n_0)$ and $(R_0, 0)$ in figure 2.

Therefore, the altitude H_0 is

$$H_0 = R_0 - R_E.$$

Math

M316

Let $k > 8$, and let $m = a_{k-6}$. Consider the remainders upon division by m of a_i (where $i > k - 6$). We will use the (standard) notation $a \equiv b \pmod m$ to denote the fact that the numbers a and b have the same remainder upon division by m. We also use the fact that the remainder upon division (or multiplication) of a sum equals the remainder upon division (or multiplication) of the sum of the remainders of the summands (or factors).

Since

$$a_{k-5} = a_{k-6}\,a_{k-7} + 1 \equiv 1 \pmod m$$

and

$$a_{k-4} = a_{k-5}\,a_{k-6} + 1 \equiv 1 \pmod m,$$

then

$$a_{k-3} = a_{k-4}\,a_{k-5} + 1 \equiv 2 \pmod m,$$
$$a_{k-2} = a_{k-3}\,a_{k-4} + 1 \equiv 3 \pmod m,$$
$$a_{k-1} = a_{k-2}\,a_{k-3} + 1 \equiv 7 \pmod m,$$
$$a_k = a_{k-1}\,a_{k-2} + 1 \equiv 22 \pmod m.$$

Thus we see that $a_k - 22$ is divisible by a_{k-6}. We can similarly prove that, beginning with a certain k, the numbers $a_k - b$ are composite, where b is any number from the sequence 1, 1, 2, 3, 7, 22, ... that is defined by the same recurrent relation: Every next number is one greater than the product of the two preceding numbers.

M317

We will use the following lemma to solve this problem.

Lemma. *Let the points A, B, and C lie on the circle centered at M. Then triangle ABC is equilateral if and only if*

$$\vec{OA}+\vec{OB}+\vec{OC}=3\vec{OM}.$$

Proof. The given equality immediately implies that

$$\vec{MA}+\vec{MB}+\vec{MC}=\vec{0}.$$

This means that the point M coincides with the center of mass of triangle ABC—that is, with the point of intersection of its medians (prove this fact). Thus the lengths of all the medians of triangle ABC are equal, which implies that the triangle is equilateral. The converse statement is not difficult to prove.

We now pass to the solution of the problem. Let the coordinates of the points A, B, C, and M be (x_A, y_A), (x_B, y_B), (x_C, y_C), and (x_M, y_M), respectively. By the statement of the problem we have

$$\begin{cases} xy=1, \\ (x-x_0)^2+(y-y^0)=4(x_0^2+y_0^2). \end{cases}$$

Substitute $y = 1/x$ from the first equation of this system into the second one and perform simple manipulations to obtain the following equation in x:

$$x^4 - 2\,x_0x^3 + \ldots = 0.$$

We write out only the first two terms, since the other terms are irrelevant. The sum of all roots of this equation, including the root $-x_0$, is $2x_0$. Thus, $x_A + x_B + x_C = 3x_0$. Similarly, $y_A + y_B + y_C = 3y_0$. These equalities imply that

$$\vec{OA}+\vec{OB}+\vec{OC}=3\vec{OM},$$

where O is the origin of the coordinate system. It remains to use the lemma proved above.

M318

We use the following factorization:

$$x^4 + x^2 + 1 = (x^2 + 1)^2 - x^2 = (x^2 + 1 - x)(x^2 + 1 + x).$$

Setting

$$x = 2^{2^{n-1}},$$

we have

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

It's not difficult to prove that the numbers

$$2^{2^n} - 2^{2^{n-1}} + 1$$

and

$$2^{2^n} + 2^{2^{n-1}} + 1$$

are relatively prime for every natural n. Indeed, if they had a common (odd) divisor $q > 1$, their difference

$$2^{2^{n-1}+1} = 2^{2^n}$$

would have had the same divisor.

Now assume that

$$2^{2^n} + 2^{2^{n-1}} + 1$$

has not less than n different prime divisors. Then, by induction,

$$2^{2^{n+1}} + 2^{2^n} + 1$$

has not less than $(n + 1)$ different prime divisors.

Remark 1. For $n > 4$, the number in question has not less than $n + 1$ different prime divisors, since

$$2^{2^4} - 2^{2^3} + 1 = 97 \cdot 673.$$

Remark 2. The statement of the problem implies that the number of different primes is infinite.

M319

It is clear that a sector always exists that contains more than one stone.

We prove a stronger assertion: that after a certain number of steps no pairs of adjacent free sectors will remain (it immediately follows from this fact that not less than half of the sectors will be occupied by stones). First we note that if a pair of sectors is free, it was also free at the previous step; in other words, no new pairs can occur during the transformations.

Now we prove our stronger assertion. Assume the opposite. Let there exist at least one free pair of adjacent sectors after every nth step; thus one of the pairs exists for an infinitely long time.

Let's cut the circle along the radius that separates the sectors of this free pair and consider the problem for the line. In this case, the segments of length 1 play the role of sectors, and we assume that the stones are located at the center of the corresponding segment. Consider the sum of distances between all stones. After every step, this distance increases at least by 2. Indeed, let two stones A and B be moved, A to the left and B to the right. The distances from A to all the stones located to its left decrease, and the distances from B to the same stones increase by the same value. The distances from A and B to the stones that remained in the sector where A and B were before moving increase (or remain the same if this sector becomes empty). The distance between A and B increases by 2. Since we can make an infinite number of steps, the distance will increase infinitely. However, it cannot be greater than the length of the entire segment multiplied by the number of stones. The contradiction obtained proves that all free pairs of sectors will eventually disappear.

Brainteasers

B316

It's clear that four receptionists can do the job. Indeed, the following schedule is possible: Everyone works 24 hours and then rests for three full days (72 hours). We prove that the number of receptionists cannot be less. Indeed, if any of the receptionists works in a 24-hour period, then

at least three people are required to work while the first one is off (60 hours). If no one works more than 12 hours, then at least three people must work while the receptionist who worked the night shift is off.

B317

Notice that the number 101 is prime. Therefore, if $101 = a + b$, then a and b are prime relative to each other (indeed, if they had a common divisor d, it would also be a divisor of the sum). Thus Eric is always the winner, regardless of his strategy.

B318

The answer is six days. To prove the baron's statement can't be true for seven days, let's put a point corresponding to each day under consideration on a line, and connect with an arrow the days for which we know Münchhausen shot fewer ducks (the beginning of the arrow) than on the second day (the end of the arrow). We obtain figure 3. The

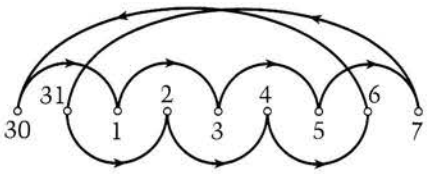


Figure 3

chain of arrows is closed, which means that the assumption that Münchhausen could make that statement for seven days was wrong. If we consider only six days, it's easy to construct an example satisfying the conditions of the problem. On July 31 he shot one duck, on August 2 two, on August 4 three, and so on. On August 5 he shot eight ducks, and every day before July 30 he shot, say, 2,000 ducks.

B319

The answer is eight players. Indeed, beginning with the sixth game, the pair of opponents is completely determined by the results of the previous games. In the sixth game, the loser of the first game and the winner of the third will play; in the seventh game, the loser of the second game and the winner of the fourth will play; and so on. This means that no new player can enter the tournament after the first five games. In the course of the first five games, eight players entered the tournament—two in the first game, two in the second, two in the third, one in the fourth, and one in the fifth.

B320

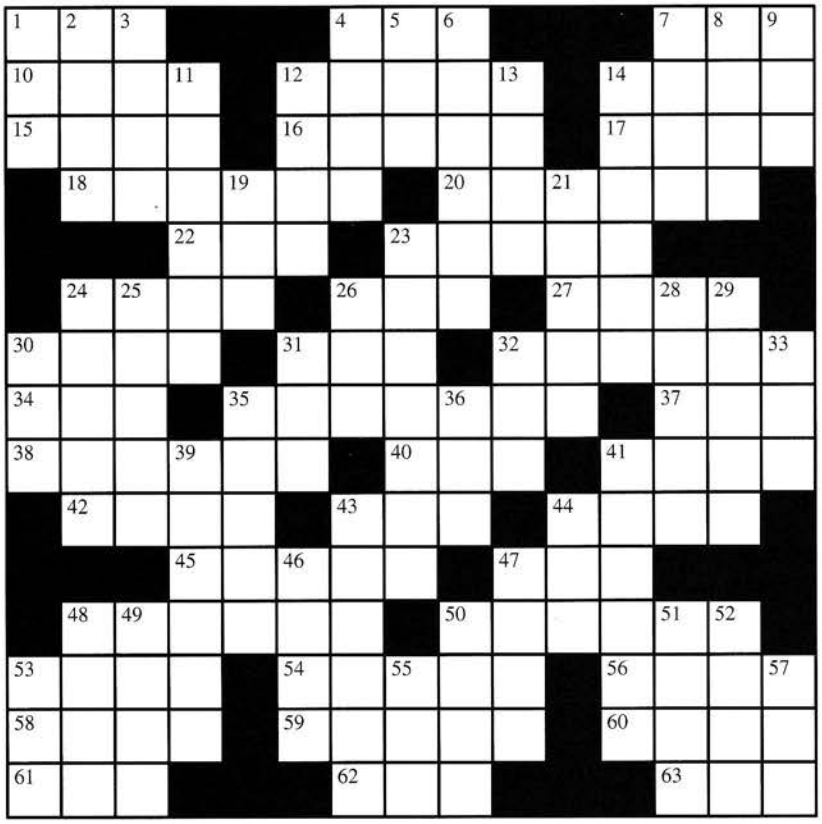
Near the shore the water is shallow. The water at the bottom of the wave is slowed by friction against the sand below, causing the upper layers of the wave to "outrun" the lower layers. As the uppermost layers run past the layers below, they are no longer supported by the water below and fall because of gravity. And so the wave curls.

Index of Advertisers

University of Toronto Press 53

Criss × cross science

by David R. Martin



ACROSS

1 Liquid meas.
4 Superconductivity theory: abbr.
7 ___ X-1 (X-ray pulsar)
10 Biblical Mount of curses
12 Sierra ___, Africa
14 60,074 (in base 16)
15 ___-Civita symbol
16 Writer Joyce Carol ___
17 Unit of length
18 Type of heat
20 Monochromatic light sources
22 Energy unit
23 ___ bob
24 10^{-24} cm^2
26 ___ Angeles
27 43,708 (in base 16)
30 Pressure units
31 Ten decibels
32 Matrices
34 100 m^2
35 Less dense than water
37 Digital display: abbr.
38 British astrophysicist ___ Arthur Milne (1896–1950)
40 Student's concern
41 Beautiful: comb. form
42 ___ lily
43 Trig. function
44 Fail
45 Noun-forming suffix
47 Kind of rally
48 Six feet
50 Drawings of functions
53 Hyperbolic function: abbr.
54 South American river
56 60,075 (in base 16)
58 Width times length
59 Vaporized water
60 Swiss mountain
61 Football QB Dawson
62 Unhappy
63 Curl

DOWN

1 Solidified colloidal solution
2 Norwegian mathematician Niels Henrik ___ (1802–1829)
3 Volcanic output
4 Interference sound
5 Trig. function
6 ___ law of refraction
7 French river
8 Fourth planet
9 Modern design tool
11 Units of volume
12 Lengthy
13 Birthright seller
14 Truss part
19 Sea eagle
21 Intelligent
23 Multisided figure
24 Ancient Celtic poets
25 "___ having fun yet?!"
26 Physicist ___Szilard
28 Rose-red spinel
29 Period
30 2990 (in base 16)
31 Flower
32 Collection of anecdotes
33 Star Wars program: abbr.
35 Watery soup
36 Physics assoc.
39 Novelist Christie
41 Element 29
43 Small planets
44 Meadow
46 Charged particles
47 Dance
48 Front
49 Asteroid
50 Common differential operator
51 1960s Broadway musical
52 Starch source
53 4.19 joules: abbr.
55 Place: comb. form
57 Binary digit

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

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

Breakfast of champions

by Don Piele

WHEN I WAS GROWING UP IN THE FIFTIES, collecting baseball and football cards was a serious extracurricular activity. Within my group of neighborhood friends, we bragged about our collections of 2 x 3" cards with the faces of famous players on the front side and short biographies on the flip side. Mickey Mantel was a highly prized baseball card and Doak Walker was the top football card. The primary sources for these coveted sports cards was Wheaties, "The Breakfast of Champions." Each cereal box promised to hold the card of a famous sport figure inside, but you would have to get Mom to buy the box before you would find out whose it was.

So Mom got the cereal as requested. It didn't matter to her as long as I was willing to eat the contents after shoving my hand down into the flakes to retrieve the card. If it turned out that I already had it in my collection, it could be traded among my friends. Of course, I hoped that this would not be necessary. At that time it never dawned on me that there was a mathematical way to calculate how many boxes, on average, Mom would have to buy before I got all the players being offered that season.

Recalling my card-collecting days suggests the following problem for investigation. Every season there were n cards being offered for collection by General Mills, the maker of Wheaties. If the goal was to collect them all, then how many boxes, on average, would Mom have to buy before I had them all? I'll make it simple and assume there is no trading with my friends.

Simulation

Armed with a computer loaded with *Mathematica*, let's take a look at how to easily answer this question via simulation. Suppose the number of cards offered for the season is $n = 12$. We first assign a number to each of the twelve cards.

```
n=12;

cards=Range[1,n]
{1,2,3,4,5,6,7,8,9,10,11,12}
```

Next we draw a random number from the set of cards

```
drawOne:=Random[Integer,{1,n}]

drawOne
4
```

Now we take this card away from the list of cards. This is done by taking the complement of the cards with the set of one element {4} consisting of the card drawn.

```
cards=Complement[cards,{4}]
{1,2,3,5,6,7,8,9,10,11,12}
```

Now all we must do is to repeat this process until all the cards have been drawn, to see how long it takes. I'll keep track of how many cards I have left in a list called cardsLeft.

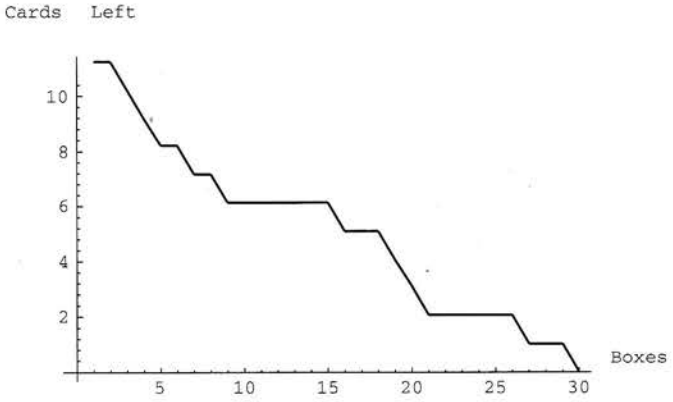
```
cards= Range[1,n];

cardsLeft={};

While[Length[cards] ≠ 0, cards=
Complement[cards,{drawOne}];
cardsLeft=Join[cardsLeft,{Length[cards]}]]

cardsLeft
{11,11,10,9,8,8,7,7,6,6,6,6,6,6,6,5,5,5,4,3,2,2,2,2,2,2,1,1,1,0}
```

The length of the cardsLeft list is how many boxes Mom had to buy before I have them all.

```
Length[cardsLeft]
30
```

When I plot this list I see the progress of my card collection until I have them all.

```
ListPlot[cardsLeft,PlotJoined → True,
AxesLabel → {"Boxes","Cards Left"}]
```



Now I'm ready to write my program to put this all together in a function that asks me to enter the number of cards to be collected and returns the number of boxes purchased in the simulation to get the whole collection. I call this function BOC for "Breakfast of Champions." Of course BOC is a random variable, since the outcome varies with each simulation.

```
BOC[n_] := Module[{cards,boxes = 0},
cards = Range[1,n];
While[Length[cards] ≠ 0,
cards = Complement[cards,{drawOne}];boxes++];
boxes]
```

Let's run BOC for each of my 10 buddies.

```
Table[BOC[12],{10}]
```

```
{32,28,41,45,24,30,35,31,40,55}
```

The luckiest one completed his collection after 28 boxes and the unluckiest one in 55 boxes. A big difference. To get a good estimate of the expected number of boxes needed, let's let each of my 10 buddies run the program 100 times and report their average.

```
Table[Apply[Plus,Table[BOC[12],{100}]]/
100//N,{10}]
```

```
{35.02,38.03,36.56,36.86,35.83,38.94,35.5,37.4,38.7,38.05}
```

That's more like it. Now I'll average these to come up with my best estimate for the number of boxes needed.

```
Apply[Plus,%]/10//N
```

```
37.089
```

On average, Mom is going to have to buy 37 boxes before my collection of 12 cards for the season is complete.

This simulation required the use of nine *Mathematica* commands: **Range**, **Random**, **Complement**, **Length**, **While**, **Module**, **Apply**, **Table**, **Apply**, **Plus**. All of these commands are commonly used by beginning users of *Mathematica*. What *Mathematica* does is harness the power of programming and make it available for everyone—even my neighborhood buddies.

Probability

The simulation method provided us with a simple and direct way to get a good estimate on the solution to this question. If you are advanced in your understanding of probability, it is possible to derive a probability density function for this problem which has an expected value for the number of boxes needed. However, because of the complexity of the distribution function, it would be nearly impossible to reach a solution without the use of the computer. Let's take a look.

The easiest case to compute is the probability that I will get the whole collection in the first 12 boxes. This means I get a new card on each and every draw for a series of 12 successes. The probability of a success on the first draw is 1 followed by a success on the second draw of $^{11}/_{12}$, followed by a success on the third draw of $^{10}/_{12}$, and so down to the last draw with a probability of success being $^{1}/_{12}$. Thus the probability of all events occurring in succession is $^{12}/_{12}$, $^{11}/_{12}$, $^{10}/_{12}$, ..., $^{11}/_{12}$.

The chance of that happening is very low.

$$\frac{12!}{12^{12}}$$

$$\frac{1925}{35831808}$$

```
%//N
```

```
0.0000537232
```

That's about once in every 20,000 times. This can be computed another way with a recursive function p that computes the probability of getting n straight successes (a new card) based on the knowledge of the probability of $n - 1$ successes. In other words, if you have 3 successes in a row then the probability you will get another success is $p(3)^{(12-3)}/_{12}$. This is expressed recursively in *Mathematica* as

```
p[1] = 1;
```

$$\texttt{p[n\_] := p[n - 1]} * \frac{\texttt{(12 - n + 1)}}{\texttt{12}}$$

We want to compute the chances of 12 successes in a row.

```
p[12]//N
```

```
0.0000537232
```

That one was easy. But we need to compute many more. In reality, we have a series of successes and failures until we have a total of 12 successes in all. So we need to compute a probability density function $p[j, k]$ which represents the probability of j successes and k failures. We will generate this distribution recursively with the following observations:

The probability of one success and no failures is 1. That means the first box you buy will have a new card in it.

```
Clear[p]

p[1,0]=1;
```

If you have one success followed by $k-1$ failures, then the probability you will have one more failure is the current probability times $^1/_{12}$. You have to pick the one card you already have to have another failure (no new card).

```
p[1, k_] := p[1, k] = p[1, k - 1] 1/12
```

Also, if you have a string of $j-1$ successes and no failures, then, as above, the probability you will add another success is the current probability times $^{(12-(j-1))}/_{12}$ or $^{(12-j+1)}/_{12}$. There are $12-j+1$ cards left that you don't have.

```
p[j_, 0] := p[j, 0]
= p[j - 1, 0] (12 - j + 1)/12
```

The key recursion is knowing how you arrive at j successes and k failures. This can be arrived at in only two ways: 1) you have j successes and $k-1$ failures and record another failure with a probability of $^j/_{12}$, or 2) you have $j-1$ successes and k failures and record another success with a probability of $^{(12-j+1)}/_{12}$. This is represented by the recursion:

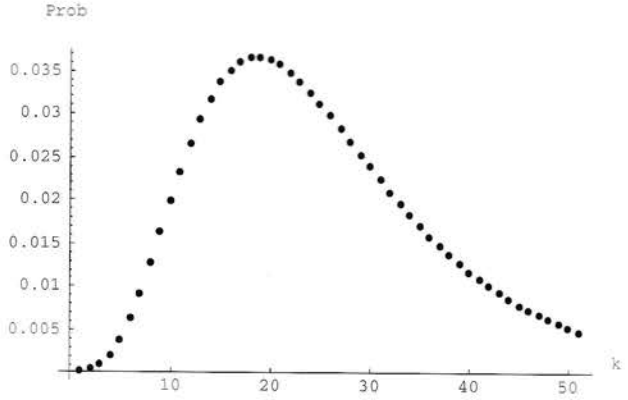
```
p[j_, k_] := p[j, k]
= p[j, k - 1] j/12 + p[j - 1, k] (12 - j + 1)/12
```

Finally, you can arrive at 12 successes and k failures from only one direction. From 11 successes and k failures you get the final card. It is not possible to arrive at 12 successes and k failures from 12 successes and $k-1$ failures, since once you reach 12 successes the game is over. The final recursion is expressed as

```
p[12, k_] := p[12, k] = p[11, k] 1/12
```

Here is a plot of the probability distribution $p[12, k]$ as the number of failures runs from 0 to 50.

```
pd=Table[p[12,k],{k,0,50}];

ListPlot[pd,AxesLabel → {"k","Prob"},
PlotStyle → PointSize[.02]]
```



The total probability is

```
Apply[Plus,pd]

0.946323
```

So, 94.6% of the time the collection is complete within 12 + 50 or 62 boxes. The expected number of boxes is the product of the probability $p[12, k]$ times the number of boxes $(12 + k)$ summed over all values of k. Since we need to stop somewhere, let's assume that the probability of more than 200 failures is insignificant.

```
Sum[p[12, k] (12 + k), {k, 0, 200}] // N

37.2385
```

Well, what do you know? We get nearly the same result we expected from our simple simulation! Does anyone believe they could get this anwer without an informatics tool?

Distribution image

To view a picture of how $p[j, k]$ changes as j varies from 1 to 12 and k varies from 0 to 70, we simply color a rectangle the value of $p[j, k]$. Each row corresponds to j, an increasing number of successes (new cards), and each column corresponds to k, an increasing number of failures. The large area of red corresponds to small probabilities. The green, blue, and purple correspond to increasingly larger probabilities. The bottom row is an image of the final distribution $p[12, k]$. Actually, I stopped at $p[11, k]$ since the colors show up better and $p[12, k]$ is a constant ($^1/_{12}$) times $p[11, k]$.

```
Show[Graphics[Table[{Hue[(p[j, k]) * 2.1],
Rectangle[{k, -j},{1 + k,1 - j}]},{k, 0, 70},
{j, 1, 11}]],
  AspectRatio → 1/4]
```

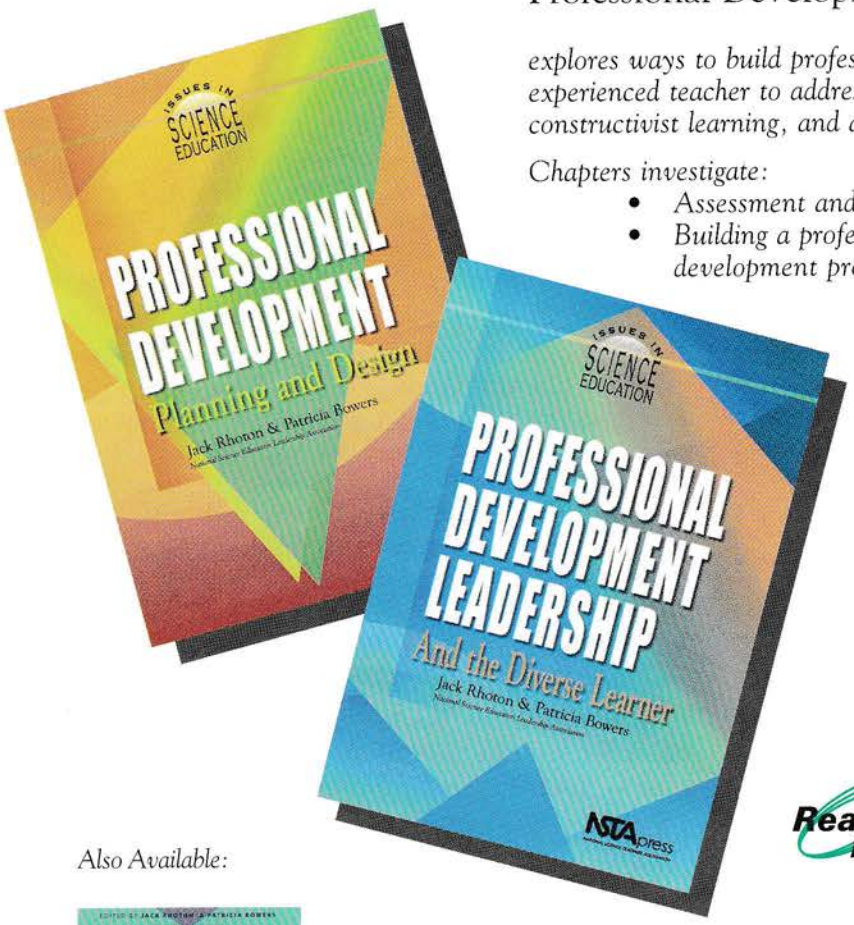
Final thoughts

The level of mathematical sophistication needed to understand the probability argument above is significantly greater than the level needed to understand the simulation argument. This is often the case with many problems in probability. Without the use of tools such as *Mathematica*, a whole line of reasoning is closed to the student. What this problem illustrates to me is, if you want to be a well rounded problem solving champion today, you had better learn how to use a computer algebra system such as *Mathematica*. And of course, you had better eat your Wheaties, "The Breakfast of Champions." ◙

ISSUES IN SCIENCE EDUCATION

Brought to you by ...

The National Science Teachers Association and
The National Science Education Leadership Association



Professional Development Planning and Design

explores ways to build professional development for the new and experienced teacher to address national standards, reform efforts, constructivist learning, and assessment strategies.

Chapters investigate:

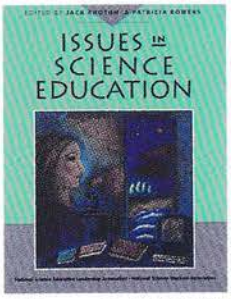
- *Assessment and evaluation*
- *Building a professional development program*
- *Curriculum development*
- *Education reform*
- *Online learning*

208 pp., © 2001 NSTA
#PB127X2, ISBN 0-87355-185-0
Members: $22.46
Non-Members: $24.95

To order, call
1-800-277-5300
or visit
www.nsta.org/store/

Read it! Read these entire books online before you order—for free!—http://www.nsta.org/store/

Also Available:



176 pp., © 1996 NSTA
#PB127X
Members: $23.36
Non-Members: $25.95

Professional Development Leadership and the Diverse Learner

discusses specific ways that professional development methods—classes, workshops, collaborative work—can give teachers the skills, resources, and knowledge necessary to help learners from different backgrounds and with different learning styles achieve scientific literacy.

Chapters investigate:

- *Community collaboration*
- *Equity/diversity*
- *Leadership development*
- *Motivation*
- *Teaching techniques*

176 pp., © 2001 NSTA
#PB127X3, ISBN 0-87355-186-9
Members: $22.46
Non-Members: $24.95

NSTApress™
NATIONAL SCIENCE TEACHERS ASSOCIATION