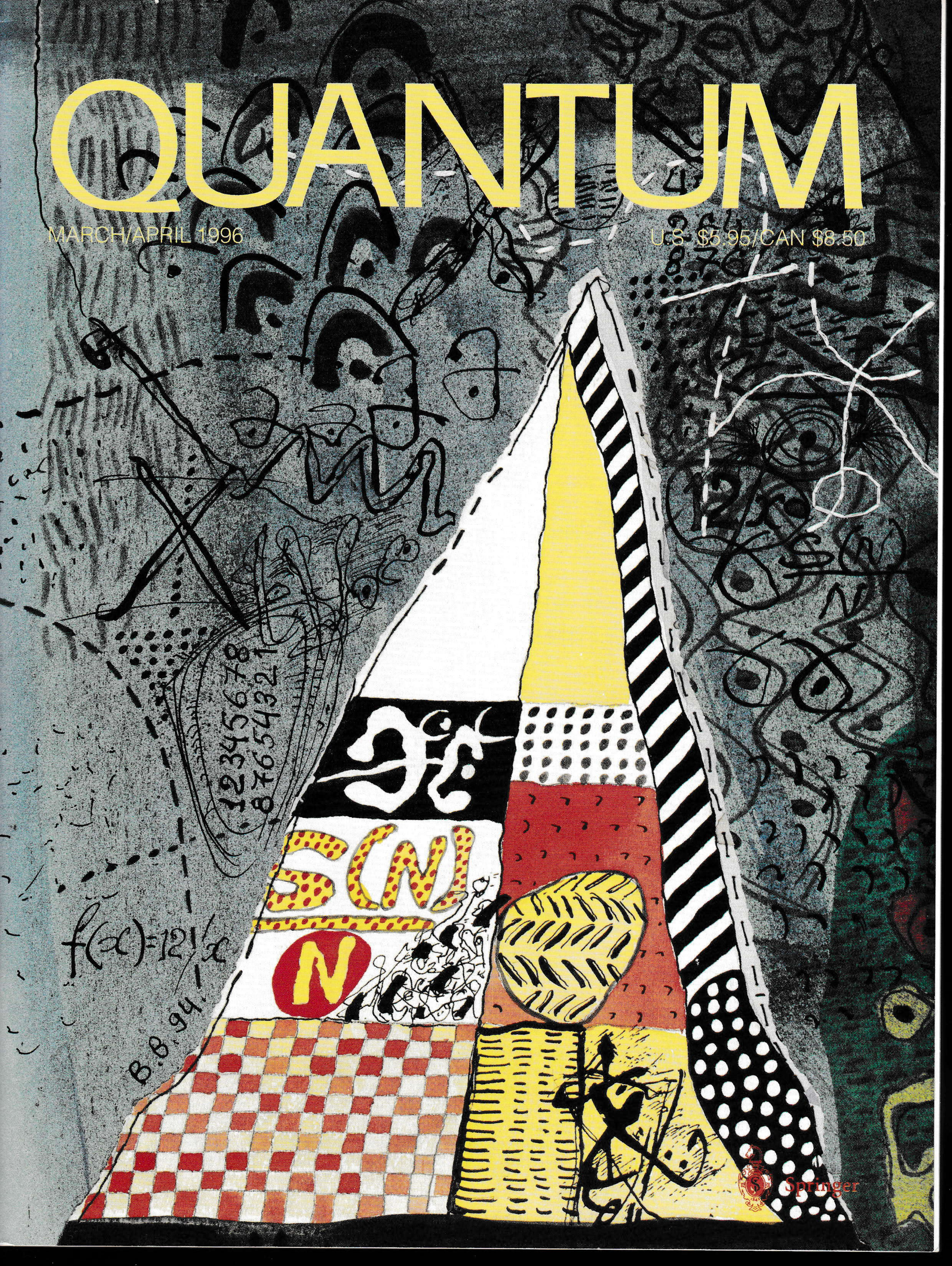
QUANTUM

MARCH/APRIL 1996

U.S. $5.95/CAN $8.50



Springer

GALLERY Q

Gift of Lila Acheson Wallace © 1996 Board of Trustees, National Gallery of Art, Washington

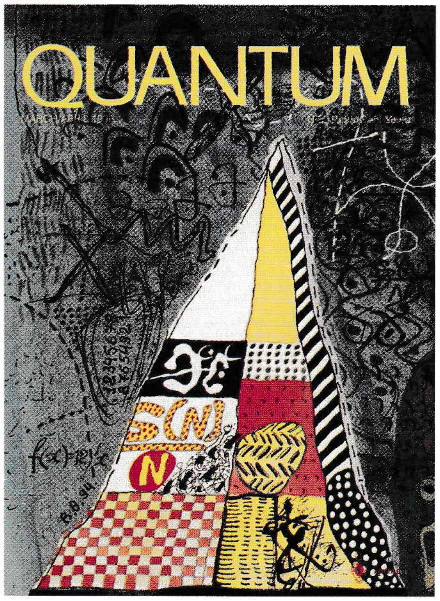
Spinner (1985) by Nancy Graves

THIS FANTASTIC CREATURE, COBBLED TOGETHER from castings of food and tools, is a kind of homage to the seemingly infinite creativity of the natural world. It's also a celebration of the physical. The sardines at the left are so precisely cast, one can see the fine detail of their scales. The salt on the pretzels is meticulously preserved in bronze. The sculpture as a whole is adorned in a riot of colors. We take a primitive, one might say "animal" pleasure in feeling these things with our eyes—objects that we normally ingest or handle without a second's thought.

The sculpture is called "Spinner" because its head and neck rotate. It recalls the mobiles of Alexander Calder, which are also kinetic balancing acts. It may also cause the scientifically minded viewer to think of problems the sculptor works through intuitively—questions of center of mass and friction. This issue of *Quantum* contains several articles devoted to these topics—see "A Gripping Story," "So What's the Point?" and "Up the Down Incline."

QUANTUM

MARCH/APRIL 1996 VOLUME 6, NUMBER 4



Cover art by Vasily Vlasov

The stylized mountain jutting up brightly through the blue gloom, populated by a phantasmagoria of numbers and shapes, contains the seeds of its own leveling. The enigmatic "$S(N)/N$," the checkerboard grid, the dots arrayed in neat rows—they are all steps on the way to a formula that raises the valleys and lowers the peaks of a certain jagged curve.

You can find that curve, and witness the drama foretold on our cover, by turning to page 24, where the author asks: "How Many Divisors Does a Number Have?"

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE*, and *General Science Source*

FEATURES

DEPARTMENTS

But what does it mean?

Putting the math into words is good practice

LEARNING MATHEMATICS is, for many of us, an awesome experience. We can use symbols to represent various kinds of numbers, and then we can define operations to perform on these numbers using logic. We can also represent geometrical objects mathematically. What is most impressive is the fact that we can use postulates or primitives, which are very simple, and from just a few arrive at a vast array of generalizations—for instance, the Peano Axioms.

Mathematics is interesting in its own right, and one should not depend on the relevance of that math to practical problems to be motivated to learn it. I remember asking my math professor at the University of Kansas, Dr. Wealthy Babcock, what possible use non-Euclidean geometry could ever have. She replied, "None, I hope!" Such a reply is essential to progress in mathematics. Young bright minds must not be constrained by the correspondence of mathematics to reality. We need explorations into areas that appear not to have application. What is surprising is that in almost every case, applications are ultimately found, even in the non-Euclidean geometries.

Now, some of us are not really mathematicians. We use math as a language to understand nature. The postulates of math can be arbitrary, and the consequent derivations can lead anywhere. In physics, however, our postulates must derive from observations of nature. We call such observations and their mathematical summaries *empirical science*. Boyle's law is an example of empirical science. It is a mathematical statement of what nature does, but it's not an explanation. Scientists use mathematics to create and test those explanations. For example, Boyle's law states that the product PV is a constant for a gas. Another empirical law, Charles's law, states that V/T is a constant. The act of combining these two empirical laws in one law—PV/T = constant—is often done by a wave of the hand, but in fact, it can't be done without solving a differential equation.

When the empirical law $PV = KT$ is combined with theory, we get something new. In this case, if we create a kinetic theory that states that the gas is represented by point particles moving and elastically colliding with the walls of the container, but never with each other, we get the equation $PV = \text{constant} \cdot E_k$, where E_k is the average kinetic energy of the particles. But from mathematics we know that if $a = b$ and $a = c$, then $b = c$. In this case, the temperature must be proportional to the average kinetic energy of the particles. This is the fundamental idea of kinetic theory and gives meaning to temperature.

My point is that the symbols used and manipulated in physics have something behind them—some meaning. They represent something. Deep understanding requires that the person manipulating these equations have the deeply felt intuitive sense of what these symbols and quantities represent. As an example, one of Maxwell's equations is $\nabla \cdot \mathbf{B} = 0$. This is a representation of a sum of partial derivatives of a magnetic field **B** with respect to direction. But what does it mean? A careful examination of the derivatives and their meaning as rates of change shows that this expression means that a magnetic field has no sources or sinks. The field must consist of closed lines. So far, no one has found a monopole for magnetism. (See "Magnetic Monopoly" by John Wylie in the May/June 1995 issue of *Quantum*.) In the case of the electric field, the equation $\nabla \cdot \mathbf{E} = \rho$ means that electric fields originate and terminate on electric charges. Great scientists—people like Feynman and Fermi—have a remarkable intuitive sense of the meaning of the mathematical expressions they use.

The problem is, too many people learn such equations without having the slightest idea what they represent. When you use such mathematics, you should ask yourself: "What does this symbol mean?" And then state it in words. For example, $\nabla\mathbf{A}$—a vector called the gradient—gives the value of the maximum rate at which **A** is changing and its direction. When a scalar product of this vector and a unit vector in a particular direction is found, the result is the rate at which **A** is changing in that particular chosen direction.

These common words and thoughts give meaning to a very complex mathematical expression. Pure mathematicians may be permitted to revel in their purity, but as scientists we should make it a practice to articulate similar "translations" of the mathematics we use.

—Bill G. Aldridge

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

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

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 6 (6 issues) will be published in 1995–1996. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau of the Russian Academy of Sciences, as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000, telephone (703) 243-7100. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) Brian Skepton (212) 460-1575; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone (0) 30-82 07-1, telex 185 411.

Second class postage paid at New York, NY, and additional mailing ofices. **Postmaster:** send address changes to: *Quantum*, Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485.

Subscription Information:

North America: Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $39; Single Issue Price: $5.95. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus NJ 07096-2485; or call 1-800-SPRINGER (777-4643) (in New York, call [201] 348-4033).

All Countries Outside North America: Personal rate is US$28 including surface delivery charges; institutional rate is US$39; airmail delivery is US$13 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.

From the edge of the universe to Tartarus

"Hesiod, the teacher of nearly everyone, considered knowledgeable about well-nigh everything . . ."
—Heraclitus of Ephesus

by Albert Stasenko

THIS HIGH PRAISE FROM AN ancient philosopher for an even more ancient poet[1] suggests an idea: can we also learn something from that person who knows "well-nigh everything"? Let's see. Here is what my reference book says about this teacher's cosmogony: "Hesiod measured the length of the universe by how long it takes an anvil to fall from the heavens to the Earth (nine days) and then from the Earth's surface to the bottom of Tartarus (also nine days). Below that is Chaos, where all downward motion ceases."

A question arises: what numerical estimates of the universe's size would the ancient poet obtain if he applied modern physics to his time scheme? Like Hesiod, we'll divide the investigation into two stages: first, the fall from the sky to the ground, which has a duration t_1 = 9 days; and second, from the ground to Tartarus (t_2 = 9 days as well). Naturally, we'll consider the day to be equal to 24 hours, because we can hardly expect the anvil to move only during the day and to rest at night. Also, we must keep in mind the main actors: the forces acting on the anvil. While the dominant force in space is that of gravity, we should take into account air resistance as the anvil drops through the Earth's atmosphere. Since nobody knows where Tartarus is, let's suppose that it lies at the center of the Earth (after all, one can't descend any further than that!). Clearly there must be a way to get to Tartarus—for instance, through a shaft dug strictly along a radius.

And now, let's begin.

Stage 1: From the sky to the ground

The force of gravity acting on a body of mass m_a located outside the Earth at a distance r from its center is known to be

$$F = -G\frac{M_\oplus m_a}{r^2} = -m_a g_\oplus \frac{R_\oplus^2}{r^2},$$

where $R_\oplus$ is the Earth's radius and $g_\oplus = GM_\oplus/R_\oplus^2$ is the acceleration due to gravity at the planet's surface. The work performed by moving a body over a small distance $dr > 0$ is

$$dW = \frac{GM_\oplus m_a}{r^2}dr.$$

So, to lift a body from the Earth's surface ($r = R_\oplus$) to the boundary of "Hesiod's universe" ($r = R_H$), we must perform work equal to

$$W = \int_{r=R_\oplus}^{R_H} \frac{GM_\oplus m_a}{r^2}dr = GM_\oplus m_a\left(\frac{1}{R_\oplus} - \frac{1}{R_H}\right).$$

If the lifted body is now set free, it will fall from a distance $r = R_H$ to the Earth's surface, and all the potential energy we gave the body will be transformed into kinetic energy:

$$\frac{m_a v_\oplus^2}{2} - 0 = GM_\oplus m_a\left(\frac{1}{R_\oplus} - \frac{1}{R_H}\right),$$

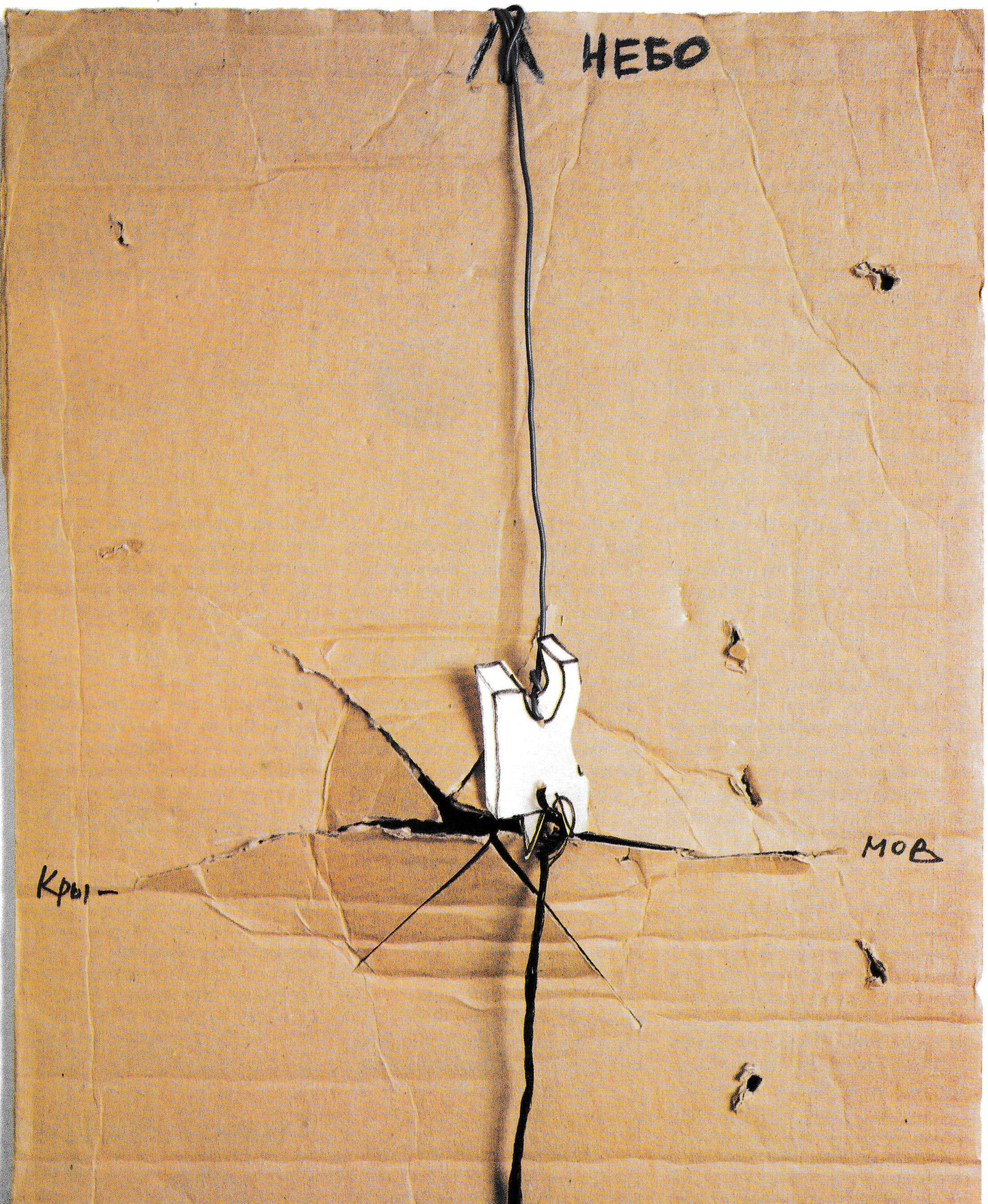
where $v_\oplus$ is the body's velocity near the Earth's surface and the zero on the left side means that the initial velocity of the body is zero.

Similarly, at any distance $r < R_H$ the law of conservation of energy can be written as follows:

$$\frac{v^2(r)}{2} = GM_\oplus\left(\frac{1}{r} - \frac{1}{R_H}\right) = g_\oplus R_\oplus^2\left(\frac{1}{r} - \frac{1}{R_H}\right).$$

[1]Hesiod (c. 700 B.C.) is the first Western poet whose name has come down to us from antiquity.

Art by Dmitry Krymov


НЕБО
Кры-
мов
ТАРТАР

Here we have canceled m_a out of the equation. We can generate another useful form of this equation,

$$\frac{v^2(r)}{2} + \phi(r) = 0 + \phi(R_H),$$

in terms of the "gravitational potential"

$$\phi(r) = -g_{\oplus} R_{\oplus}^2 \frac{1}{r}$$

and taking into account the null value of the anvil's initial velocity $v_H = 0$. This form of the equation clearly reflects the conservation of energy, written as a sum of potential and kinetic energies per unit mass.

The functions for the gravitational potential $\phi(r)$ and the acceleration due to gravity $g(r)$ can be shown both for $r < R_{\oplus}$ and $r > R_{\oplus}$ (fig. 1). Similar graphs have been drawn many a time.[2]

Thus, the anvil falling freely from the "altitude" of Hesiod's radius R_H will have a free-fall velocity

$$v(r) = v_{esc}\sqrt{\frac{1}{r/R_{\oplus}} - \frac{1}{R_H/R_{\oplus}}} = -\frac{dr}{dt}$$

at an arbitrary point outside the Earth. Here $v_{esc} = \sqrt{2g_{\oplus}R_{\oplus}} = \sqrt{-2\phi_{\oplus}}$ is the escape velocity.[3]

To find the desired distance R_H we should integrate the equation displayed above:

$$\frac{1}{v_{esc}} \int_{r=R_H}^{R_{\oplus}} \frac{dr}{\sqrt{\frac{1}{r/R_{\oplus}} - \frac{1}{R_H/R_{\oplus}}}} = -t_1. \qquad (1)$$

However, it's boring to integrate—we can come up with a passable estimate some other way. For example, as it "falls" to the Earth, the Moon is known to make a complete revolution around the planet in approximately 28 days. (This is where Newton saw an analogy with a falling apple.) Thus, the Moon travels from point P to point Q (fig. 2) in 28/4 = 7 days. Perhaps it isn't by chance that this value is very close to Hesiod's estimate of "nine days"? If there is something to this conjecture, the size of Hesiod's universe should be about the radius of the Moon's orbit—that is, about 380,000 km $\cong 60R_{\oplus}$.

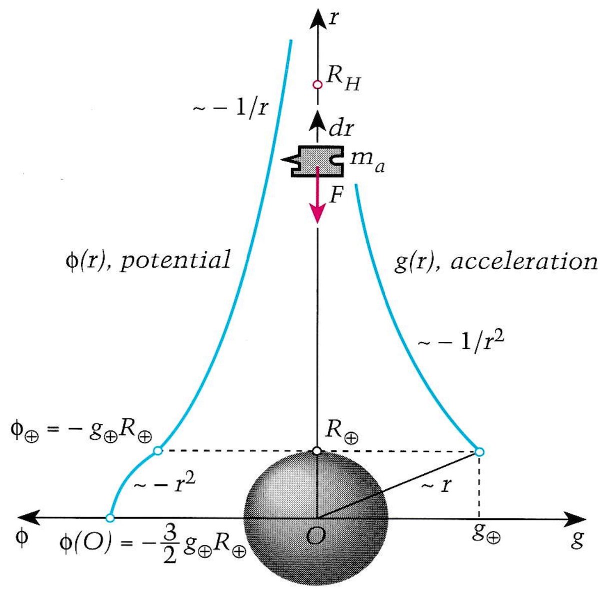


Figure 1

The integral (1) can be written in dimensionless form by expressing all the linear sizes in units of the radius R_H we seek:

$$\left(\frac{R_H}{R_{\oplus}}\right)^{3/2} \int_{r=R_{\oplus}}^{R_H} \frac{\sqrt{r/R_H}}{\sqrt{1 - r/R_H}} d\left(\frac{r}{R_H}\right) = \frac{v_{esc}t_1}{R_{\oplus}}. \qquad (2)$$

Denoting the integral by a certain dimensionless constant C we get

$$\left(\frac{R_H}{R_{\oplus}}\right)^{3/2} = \frac{v_{esc}t_1}{CR_{\oplus}}. \qquad (3)$$

Since $v_{esc}/CR_{\oplus}$ is also constant, this equation, rewritten in the form $R_H^3 \sim t_1^2$, looks very similar to Kepler's third law, which states that the cube of the semimajor axis of a planet's elliptical orbit is proportional to the square of its period of revolution around the Sun. It looks as if we're revolving in the same circle of ideas. Inserting the numerical data $v_{esc} = 11.2 \cdot 10^3$ m/s, $t_1 = 9 \cdot 24 \cdot 3{,}600$ s $= 7.8 \cdot 10^5$ s, and $R_{\oplus} = 6.4 \cdot 10^6$ m into equation (3), we get

$$\frac{R_H}{R_{\oplus}} = \frac{123}{C^{2/3}}.$$

If we suppose that our dimensionless integral C (equation (2)) is equal to about 1 (which is common practice in dimensional analysis), we can say the radius we seek is two orders larger than that of the Earth ($R_H = 10^2R_{\oplus}$), and this value is similar to that obtained previously in the "lunar" approach.

Those who still want to calculate the size of Hesiod's universe more exactly must overcome the difficulties of calculating the integral in equation (2). Let's begin by substituting variables: $r/R_H = \sin^2\theta$, so

$$\sqrt{1 - \frac{r}{R_H}} = \sqrt{1 - \sin^2\theta} = \cos\theta,$$

$$\frac{dr}{R_H} = 2\sin\theta\cos\theta\, d\theta,$$

and the integral will be

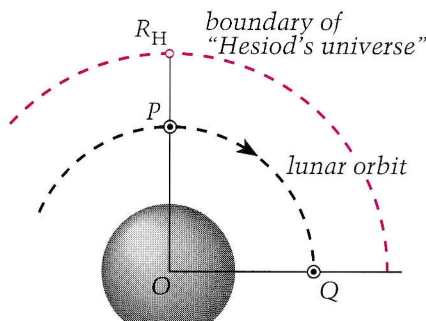


Figure 2

[2]See "Late Light from Mercury" in the November/December 1993 issue of *Quantum*.

[3]We must keep in mind that v_{esc} is the minimum velocity that must be given to a body at the Earth's surface in order for it to escape the planet's gravitational field ($\phi_{\oplus} = -g_{\oplus}R_{\oplus}$ is the potential of this field at the Earth's surface).

$$C = \int\limits_{r=R_\oplus}^{R_H} \frac{\sqrt{r/R_H}}{\sqrt{1-r/R_H}} d\left(\frac{r}{R_H}\right)$$

$$= \int\limits_{\theta=\arcsin\sqrt{\frac{R_\oplus}{R_H}}}^{\arcsin 1} 2\sin^2\theta d\theta$$

$$= \int\limits_{\theta=\arcsin\sqrt{\frac{R_\oplus}{R_H}}}^{\pi/2} (1-\cos 2\theta) d\theta$$

$$= \left(\theta - \frac{\sin 2\theta}{2}\right)\Bigg|_{\arcsin\sqrt{\frac{R_\oplus}{R_H}}}^{\pi/2}.$$

Our estimate above showed that Hesiod's radius is larger than that of the Earth by two orders of magnitude—that is, $R_\oplus/R_H \ll 1$. We can therefore set the lower limit of the integral equal to zero: $\arcsin(R_\oplus/R_H) \cong 0$. Then $C = \pi/2$ (as we expected, this value is about 1). Thus,

$$\frac{R_H}{R_\oplus} = \frac{123}{(\pi/2)^{2/3}} \cong 90,$$

which works out to

$$R_H \cong 6 \cdot 10^8 \text{ m}.$$

So Hesiod's Universe includes the entire lunar orbit and more—it stretches one and a half times farther.

Of course, this is a far cry from modern estimates of the Universe's size, but still it's not too bad if we remember that Hesiod lived 27 centuries ago and was not a physicist but a country poet. Scholars who lived well after his time considered the Earth to be flat and to rest on the backs of three whales . . .

However, the anvil did not stop yet—it continues to fly on.

Stage 2: From the Earth's surface to Tartarus

Everything is much more complicated here. How, in Hesiod's view, did the anvil fall further? Did he know about air resistance? Let's try to enlist Hesiod himself in searching for the answer. We consider two cases.

1. *There is no air in the shaft leading to Tartarus, and according to our assumption Tartarus can be no further from the surface than the Earth's center.*

So let's say that the anvil passes the Earth's surface at time $t = 0$ and enters a vertical shaft dug down to the very center of the planet. To write the equation of a body's motion under the influence of gravity only, we recall that inside a homogeneous planet this force is proportional to the distance from the center—that is, $F = -m_a g_\oplus r/R_\oplus$:

$$m_a \frac{d^2r}{dt^2} = -m_a g_\oplus \frac{r}{R_\oplus},$$

or

$$\frac{d^2r}{dt^2} + \frac{g_\oplus}{R_\oplus} r = 0. \qquad (4)$$

This is the equation for simple harmonic motion with frequency $\omega = \sqrt{g_\oplus/R_\oplus}$, which has the solution

$$r(t) = A \cos \omega t + B \sin \omega t.$$

Taking into account the initial conditions of the problem, at $t = 0$ we have

$$r = R_\oplus,$$

$$v(R_\oplus) = \frac{dr}{dt}\bigg|_{t=0} = -v_H,$$

and so

$$r(t) = R_\oplus \cos \omega t - \frac{v_H}{\omega} \sin \omega t.$$

The anvil will arrive at the Earth's center ($r = 0$) in the time $t = t_2$ (the second time interval of the anvil's fall). From here it follows that

$$0 = R_\oplus \cos \omega t_2 - \frac{v_H}{\omega} \sin \omega t_2,$$

$$\tan \omega t_2 = \frac{R_\oplus}{v_H}\omega = \frac{1}{\sqrt{2}},$$

$$t_2 = \frac{1}{\omega}\arctan\frac{1}{\sqrt{2}} = \sqrt{\frac{R_\oplus}{g_\oplus}} \arctan\frac{1}{\sqrt{2}}.$$

Since

$$g_\oplus = G\frac{M_\oplus}{R_\oplus^2} = G\frac{\frac{4}{3}\pi R_\oplus^3 \langle\rho\rangle}{R_\oplus^2}$$

$$= \frac{4}{3}\pi G R_\oplus \langle\rho\rangle,$$

where $\langle\rho\rangle$ is the average density of the Earth, we get a formula for evaluating t_2:

$$t_2 = \frac{1}{\sqrt{\langle\rho\rangle}}\sqrt{\frac{3}{4\pi G}} \arctan\frac{1}{\sqrt{2}}.$$

Note that the duration of the fall from the surface to the center depends only on the average density of the planet $\langle\rho\rangle$. On this point Hesiod went astray. One must either agree that, in accordance with the value of $t_2 = 9$ days $= 7.8 \cdot 10^5$ s, the planet's density is

$$\langle\rho\rangle = \frac{3}{4\pi G}\left(\frac{\arctan\frac{1}{\sqrt{2}}}{t_2}\right)^2$$

$$\cong 2 \cdot 10^{-3} \text{ kg/m}^3$$

—that is, three orders of magnitude less than air (patent nonsense)—or admit that the anvil reaches the center far more quickly:

$$t_2 = \frac{1}{\sqrt{5.5 \cdot 10^3}}\sqrt{\frac{3}{4 \cdot 3.14 \cdot 6.67 \cdot 10^{-11}}} 0.616 \text{ s}$$

$$\cong 500 \text{ s}.$$

As a result of this computation, Hesiod would have to reject the assumption that there is no braking force. So now we consider the second assumption.

2. *Air exists after all.*

In this case, the motion equation (4) must be written with a right-hand term instead of zero—that is, the air resistance (divided by the anvil's mass), which is proportional to the density of the air; the square of the anvil's velocity; the square of its cross-sectional area $S_\perp$; and the buoyancy force, which increases, one would think, due to the inevitable increase in the air's density ρ with depth. Thus we get

$$\frac{dv}{dt}+\frac{g_\oplus}{R_\oplus}r=\frac{\rho(r)v^2(r)S_\perp}{m_a}C_m+\frac{\rho(r)}{\rho^0}\frac{g_\oplus}{R_\oplus}r, \tag{5}$$

where ρ^0 is the density of the anvil: $m_a = \rho^0 V_a$ (V_a is its volume). A certain dimensionless drag coefficient C_m is introduced into equation (5) that depends on many parameters, but is nevertheless equal to about 1. However, even if this coefficient is assumed to be constant, an important question remains: how does the air's density ρ depend on the shaft's depth $h = R_\oplus - r$ (or, in other words, on the distance r from the Earth's center)?

It's known that the dependence of the air's density on the altitude $h = R_\oplus - r$ above the Earth's surface is described by Boltzmann's barometric height formula $\rho(h) = \rho_\oplus e^{-mgh/kT}$, where T is the temperature of the atmosphere (assumed to be constant), m is the molecular mass, k is Boltzmann's constant, and $\rho_\oplus \sim 1$ kg/m³ is the value of ρ at the Earth's surface (that is, at sea level—see figure 3). What is the exponent in the exponential function? It's the ratio of two forms of energy: the gravitational potential energy mgh (which is zero at sea level) and the average molecular kinetic energy kT. Suppose we make a shaft whose walls are at a constant temperature ($T \sim 300$ K) and assume that the acceleration due to gravity is equal to $g_\oplus$ (strictly speaking it decreases with depth—see figure 1—but we don't plan to go too deep at this point). Under these conditions equation (5) is valid and yields a depth $h_* < 0$ where the air is compressed so much (to the density ρ_*) that its

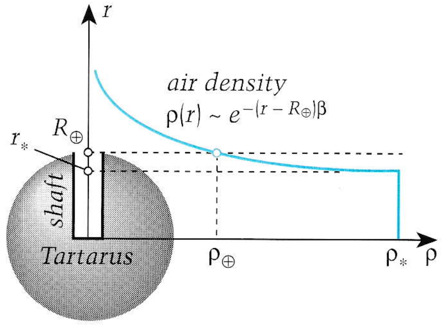


Figure 3

molecules touch one another. Whether this air could be considered a liquid is a question belonging to the thermodynamics of the phases of matter. Just to be on the safe side (and for good reason!) we'll put the phrase "liquid air" in quotation marks. Let's assume that ρ_* is of the order of the density of water (~10³ km/m³). Then equation (5) leads to

$$\begin{aligned} h_* &= \frac{kT}{mg_\oplus}\ln\frac{\rho_\oplus}{\rho_*}=\frac{RT}{Mg_\oplus}\ln\frac{\rho_\oplus}{\rho_*} \\ &= \frac{8.31\cdot 300}{29\cdot 10^{-3}\cdot 10}\ln 10^{-3}\ \text{m} \\ &\cong -60\cdot 10^3\ \text{m}. \end{aligned}$$

Clearly, at this depth—about 1% of the Earth's radius—the acceleration due to gravity does not vary appreciably.

So what will happen to the anvil? Having acquired almost escape velocity during its free fall from the boundary of Hesiod's universe, it will crash into the Earth's atmosphere and begin to decelerate while being heated due to friction with the air. If in the process it doesn't melt, burn, and disintegrate (after all, it's the handiwork of the immortal blacksmith Hephaestus!), it will fall deep into the shaft and meet denser and denser air, and then beginning at the depth h_* (or $r_* = R_\oplus - h_*$) it will move in almost "liquid" air. It's clear that the anvil will pass through air with a characteristic thickness of 10 km in about a second. For some dozens (or hundreds) of seconds it will fall through ever thicker air in the shaft until almost all the anvil's kinetic energy is expended in working against the air resistance. This spectacular disregard for the time it takes for the anvil to "get used to" its new falling conditions (this period is often referred to in technical slang as the "relaxation time") merely reflects our hope that the relaxation time is small compared to t_1 and t_2, and also our reluctance to spend the time, energy, and paper needed to prove it (it isn't too difficult, by the way).

The further motion of the anvil in the "liquid air" of density ρ_* will be characterized by the balance of all the forces involved: gravity, buoyancy, and air resistance. So equation (5) leads to

$$\frac{\rho_* v^2(r)S_\perp C_m}{m_a}=\frac{g_\oplus}{R_\oplus}r\left(1-\frac{\rho_*}{\rho^0}\right),$$

from which we get

$$v(r)=\sqrt{r}\,\frac{v_{esc}}{R_\oplus\sqrt{2}}\sqrt{\frac{\rho^0}{\rho_*}-1}\sqrt{\frac{m_a}{\rho^0 S_\perp C_m}}=-\frac{dr}{dt}.$$

Therefore,

$$\int_0^{t_2}\frac{v_{esc}}{R_\oplus\sqrt{2}}\sqrt{\frac{\rho^0}{\rho_*}-1}\sqrt{\frac{m_a}{\rho^0 S_\perp C_m}}\,dt=-\int_{R_\oplus}^{0}\frac{dr}{\sqrt{r}}.$$

This last equation yields the value we seek for the time t_2 required for the anvil to drop to the center of the Earth:

$$t_2=\frac{R_\oplus^{3/2}2\sqrt{2}}{v_{esc}\sqrt{\frac{\rho^0}{\rho_*}-1}\sqrt{\frac{V_a}{S_\perp C_m}}}.$$

We can estimate the order of magnitude of t_2 by inserting the numerical data $\rho^0/\rho_* - 1 \sim 10$, $V_a/S_\perp \sim 1$ m, and $C_m \sim 1$. This gives us $t_2 \sim 10^6$ s, which is close to Hesiod's "nine days." Those who like to play with computers may obtain a more precise solution to equation (5). One might also ponder how the ancient poets can stimulate us to investigate certain physical phenomena. Are poetry and physics really that far apart?

Just for the fun of it!

B166

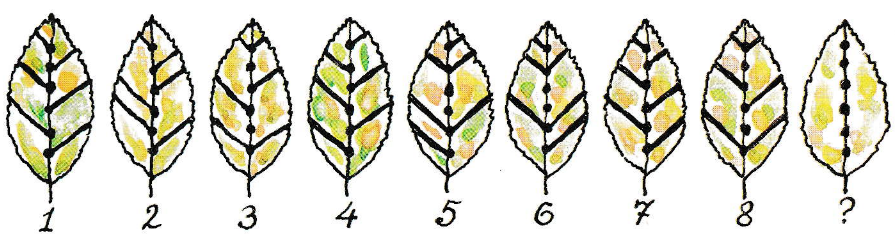
Pinocchio's prehistory. Master Ciliegia received an order to make a certain number of stools. "If I make three stools a day, starting from today," the carpenter thought aloud, "I'll just finish on Sunday. If I make five stools a day, I'll be done on Friday."—"And what day is it today?" asked a curious talking block of wood. Indeed, what day is it? (A. Shevkin)

B167

Ordering by triples. The seven volumes of an encyclopedia stand on a shelf in the order 1, 5, 6, 2, 4, 3, 7. Put them in increasing order using a series of the following operations: any three consecutive volumes are moved to the left or right end of the shelf or inserted between any two of the other volumes in the same order. (A. Savin)

B168

Lost cargo. When Pitzius was a little boy, he loaded a toy boat with some metal pieces from his construction set and set it afloat in his bath tub. Suddenly the ship began to list, and the metal pieces sank to the bottom of the tub. Did the water level change?

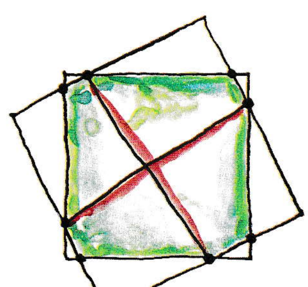


B169

Botanical logic. The pattern of veins on the first eight leaves in the figure above is determined by a certain law. Find the law and draw the veins on the ninth leaf. (Z. Chromý [Czech Republic]—2nd World Puzzle Championship)

B170

Intersecting squares. The intersection of two squares (not necessarily of equal size) is an octagon—see the figure at right. It is divided into four quadrilaterals by two diagonals (joining opposite vertices). Prove that these diagonals are perpendicular to each other. (V. Proizvolov)



ANSWERS, HINTS & SOLUTIONS ON PAGE 58

Art by Pavel Chernusky

Construction program

Regular polygons, Euler's function, and Fermat numbers

by Alexander Kirillov

Art by Sergey Ivanov

EDITOR'S NOTE: WE PUBLISH this article in connection with the bicentenary last year of the first great achievement of Carl Friedrich Gauss: his straightedge-and-compass construction of the regular 17-gon. Young Gauss was so impressed by the discovery he chose mathematics as his profession. This construction thus became a crucial point in the history of mathematics as well as in his life. We even know the exact date—March 31, 1795 (Gauss started his diary on this day). Later he developed his method into an important and beautiful theory and proved the constructibility of regular n-gons for all numbers n of a certain form, described in terms of Fermat primes. This article approaches the problem from the other direction: it explains why regular polygons are constructible only for these values of n.

Prologue

Geometric constructions are one of the most popular kinds of problem in school mathematics. And by no means is this a matter of chance. The history of geometric constructions covers several millennia, and even as early as in ancient Greece this mathematical art reached an extraordinarily high level. Suffice it to mention the famous problem of Apollonius: *construct a circle touching three given circles.*

I think many of our readers must have heard about the three famous problems of antiquity, which turned out to be unsolvable: squaring the circle, trisecting the angle, and doubling the cube. But perhaps the most beautiful is the problem of constructing regular polygons. In fact, this isn't *one* problem, it's an entire series of problems: *for each integer $n \geq 3$, a regular n-gon must be constructed using only a straightedge and a compass.*

For some values of n this is a very simple problem (for example, for $n = 3, 4, 6, 8, 12$); for some other values it's somewhat more difficult ($n = 5, 10, 15$—I'll explain later how a regular decagon and pentagon can be constructed); then there are values of n for which the problem is extremely hard ($n = 17$ or 257). And finally, values of n exist such that the problem can't be solved at all (for instance, $n = 7, 9, 11$).

Let's write out a number of integers starting with $n = 3$ and highlight in red the values for which the regular n-gon can be constructed with straightedge and compass:

3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, ...

Is there any law as to how the red and black numbers are distributed? Yes, there is. But it's fairly difficult to find. This law is arithmetic in nature—to describe it we'll have to leave geometry for the time being and take up some aspects of number theory, the highest branch of arithmetic.

Euler's function

Leonhard Euler, the renowned 18th-century mathematician, was one of the first to notice that the number of positive integers less than a given n and relatively prime with n is a useful and important arithmetic characteristic of n. Euler

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$\phi(n)$	1	1	2	2	4	2	6	4	6	4	10	4	12	6	8	8	16	6	18	8	12
n	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
$\phi(n)$	10	22	8	20	12	24	12	28	8	30	16	20	16	24	12	36	18	24	16	40	12

introduced the notation $\phi(n)$ for this number, and since then the function $n \to \phi(n)$ has been referred to as *Euler's function*. For example, there are four numbers less than and relatively prime with $n = 10$: 1, 3, 7, and 9. So $\phi(10) = 4$.

The function ϕ has many interesting properties. Euler himself discovered one of them: *for any two relatively prime numbers m and n,*

$$\phi(mn) = \phi(m)\phi(n). \tag{1}$$

It can also be seen that, *for any prime p*, $\phi(p) = p - 1$, $\phi(p^2) = p^2 - p$, and in general,

$$\phi(p^k) = p^{k-1}(p-1). \tag{2}$$

These properties are sufficient to compute Euler's function. This is quite easy for small values of n—for instance,

$$\phi(10) = \phi(2)\phi(5) = 1 \cdot 4 = 4,$$
$$\phi(100) = \phi(4)\phi(25) = 2 \cdot 20 = 40.$$

The first 42 values of $\phi(n)$ are shown in the table above. Compare these data with the sequence of red and black numbers above. The connection between the "color" of a number n and the value of $\phi(n)$ is now almost obvious, isn't it? We see that if a regular n-gon can be constructed with straightedge and compass, then $\phi(n)$ is a power of two. This observation turns out to be a necessary and sufficient condition for the constructibility of a regular n-gon.

I won't give a rigorous proof of this fact here. But I'll present quite simple and convincing considerations in its favor. Similar reasoning can be applied to many other problems in geometric construction—for example, to trisecting the angle.

What does it mean to "construct"?

Before we begin to study the problem of what is constructible, we ought to explain what "constructible" means. That is, it would be good to give an exact formulation of the rules for straightedge-and-compass constructions. For instance, the straightedge can be used only for drawing lines through a pair of points—it has only one straight, unmarked edge. Similarly, a compass is used only for drawing circles with a given radius and center. However, these matters have repeatedly been discussed in the literature, so I'll confine myself to this brief reminder, relying on the reader's intuition.

What's more important to me here is that the net result of solving a construction problem is (at least, in principle) a sequence of elementary operations resembling a computer program.

For example, the midpoint of a segment AB is constructed by the following "program" (fig. 1):

1. With the compass, draw a circle ω_1 with center A and radius AB.
2. With the compass, draw a circle ω_2 with center B and radius BA.
3. Mark the intersection points M_1 and M_2 of circles ω_1 and ω_2.
4. Use the straightedge to draw the straight line M_1M_2.
5. Mark the intersection point X of M_1M_2 and AB.

Here's another example: a construction of the bisector of a given angle AOB (fig. 2 on the next page). The corresponding command system can take the following form:

1. Use the compass to draw a circle

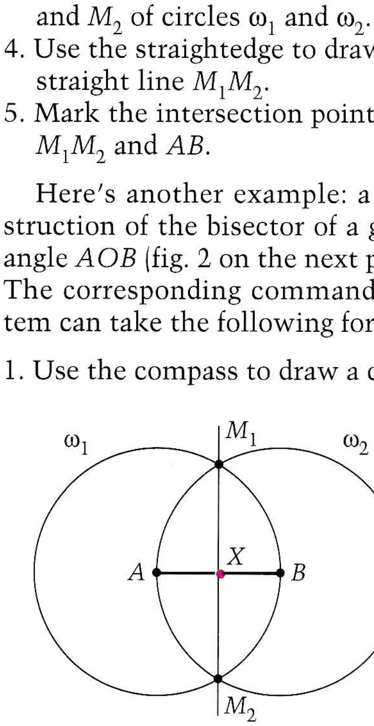


Figure 1

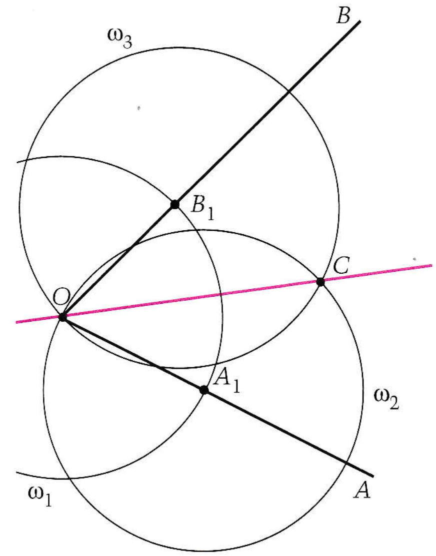


Figure 2

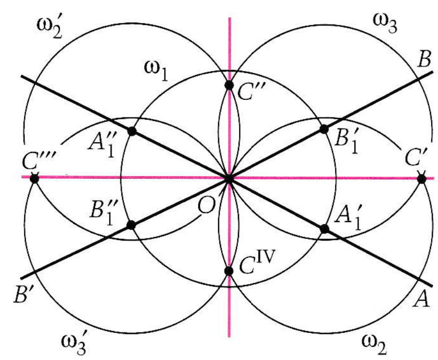


Figure 3

ω_1 with center O and an arbitrary radius R.

2, 3. Mark the intersection points A_1, B_1 of this circle with the lines OA and OB, respectively.

4, 5. Use the compass to construct circles ω_2, ω_3 with centers A_1, B_1, and radius R.

6. Mark as C the other intersection point of ω_2 and ω_3 (that is, the one that isn't O).

7. Use the straightedge to draw the line OC.

However, in this case, steps 2 and 3 of the program aren't formulated precisely. The problem is that the circle ω_1 meets the lines OA and OB at two points each, and it's unclear which of these points is to be labeled A_1 and which is to be B_1. You may protest that it should go about the *half-lines OA* and *OB*, which meet the circle at a single point each. But the notion of a "half-line" falls outside the scope of our "construction machine's" understanding. It can only handle the notion of "line."

Let's see what happens if the term "intersection point" is understood as "any intersection point." Our program will then produce figure 3: in place of points A_1 and B_1 we mark two new points for each old one: A_1', A_1'', and B_1', B_1''. And so what was originally a single point C turns into four different points C', C'', C''', and C^{IV}. This, however, leads to two rather than four different answers: the lines OC' and OC''' coincide, as do OC'' and OC^{IV}.

So how many different answers can be produced by the same program solving a given construction problem? Any program of this kind consists of elementary operations. There are only five of them: drawing a line through a given pair of points; drawing a circle with a given center and radius; marking the intersection points of two lines; marking the intersection points of a line and a circle; and marking the intersection points of two circles. The first three operations take a single value; the last two contain a two-valued uncertainty.[1]

If a program consists only of one-valued operations, we get only one answer. If there is one two-valued operation, it leads to two realizations of the program (as in the example above). And in general, if a program contains k two-valued operations, it can be realized in 2^k ways.

We have seen that some ambiguities can eventually "cancel out" without affecting the final answer. But it turns out that these cancellations always take place in such a way that the eventual indefiniteness is always 2^l-fold ($l \le k$). This coordination is algebraic rather than geometric in nature (the corresponding branch of algebra is called Galois theory) and could be strictly proved, but that isn't the aim of this article.

Let's return to the construction of the angle bisector. Along with the bisector of angle AOB, our program yields the bisector of the "external" (that is, adjacent) angle (fig. 3). It should not be regarded as an extraneous solution. From the viewpoint of our straightedge and compass, which "understand" angles only as pairs of lines, the external angle is as good as the original angle AOB. If we try to define the notion of bisector in terms "understandable" to the straightedge and compass, we'll see that the external bisector will satisfy this definition as well as that of the "normal," internal kind.

This phenomenon is general in nature. All the 2^l solutions produced by a program with ambiguities are "genuine" rather than extraneous solutions, as long as the problem is given the proper wording.

For example, the problem "*inscribe a circle in a triangle*" is solved by a program with 16-fold uncertainty (we construct the bisectors of two angles), which leads to four different answers (one inscribed and three escribed circles). All of them become equally legal if the problem is formulated as "construct a circle touching three given lines." The difference between inscribed and escribed circles is based on the notion of "between" (or "interior") and is beyond our "computer's" comprehension.

The examples discussed above also show that if a construction problem has several solutions, the construction program yields them all. This statement is true in the general case as well.

An instructive example: the geometric construction of one of the roots of a quadratic equation automatically gives the second root.

Thus we arrive at the following principle: *any solvable problem in straightedge-and-compass constructions has 2^l solutions for some integer l.*

A rigorous proof of this assertion is given by Galois theory and can't be presented in this article. But the

[1]Of course, two circles, or a circle and a line, can be disjoint or touching each other (at a single point). It's possible to include these cases also in the general scheme, but I'd better ignore them here.

assertion itself looks very simple and could perfectly well have been discovered by the mathematicians of antiquity. The question arises why this discovery was made only in the last century, although many corroborative examples have been known for thousands of years. (For instance, the problem of Apollonius mentioned above has, in general, eight solutions.)

One possible explanation is that the modern, "computer" setting of the problem never occurred to the geometers of the past.[2] Another reason is that they considered separately each single problem instead of an entire series of similar problems (such as the construction of the regular n-gons for all n).

Perhaps this question will draw the attention of historians of mathematics and they will give us a more complete explanation why this opportunity was missed.

Regular polygons

Let's get back to our main problem. We want to know when a regular n-gon can be constructed with straightedge and compass. Our previous reasoning suggests investigating the possible number of solutions to this problem. To get a sensible answer, we must refine its formulation. Let's fix the *size* and *position* of the desired n-gon (otherwise, there will certainly be infinitely many solutions, provided there's at least one). To this end, let's fix the circumcircle ω of our n-gon and the location of one of its vertices A_0 on this circle. Then we have to find the positions of the other $n-1$ vertices $A_1, A_2, \ldots, A_{n-1}$. Obviously it will suffice to find the position of A_1: marking off successive arcs equal to A_0A_1, we'll plot the points A_2, A_3, A_4, ... on the circle.

The easiest case is when $n = 6$. The side length a regular hexagon inscribed in a circle equals the radius of the circle. So the required

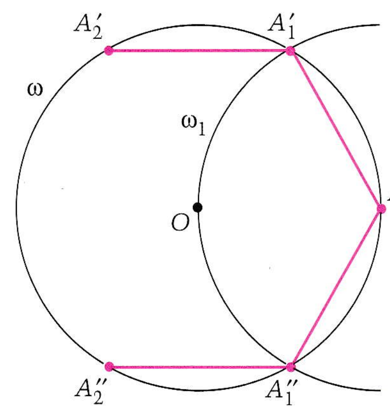


Figure 4

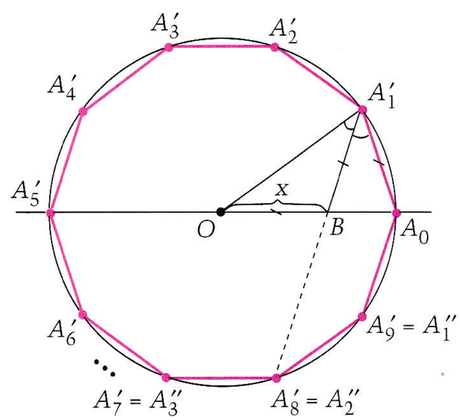


Figure 5

"program" boils down basically to two steps (fig. 4—hereinafter O is the center of circle ω):

1. Use the compass to draw a circle ω_1 with center A_0 and radius OA_0.
2. Mark an intersection point A_1 of the circles ω and ω_1.

We see that this program yields two points (A_1' and A_1'' in figure 4), but the corresponding hexagons $A_0A_1'A_2'A_3'A_4'A_5'$ and $A_0A_1''A_2''A_3''A_4''A_5''$ differ only in the order of numeration of their vertices.

The same thing occurs for $n = 3$ and $n = 4$. The cases $n = 5$ and $n = 10$ are more interesting. Let's examine $n = 10$.

Suppose we start with any decagon $A_0A_1A_2A_3A_4A_5A_6A_7A_8A_9$, inscribed in a circle. Draw the bisector A_1B of the angle OA_1A_0 in triangle OA_0A_1 (see figure 5, where the label A_1 is supplied with a prime for future use). It's easy to see (for example, by direct calculation of angles) that OA_1B and BA_1A_0 are isosceles triangles (so that $OB = BA_1 = A_1A_0$) and that the triangles OA_1A_0 and A_1A_0B are similar to each other. Think of the line OA_0 as a number axis with its origin at O and point 1 at A_0. Let point B correspond to the number x. Then from the similarity of the triangles mentioned above, we get

$$\frac{x}{1-x} = \frac{1}{x},$$

or

$$x^2 + x - 1 = 0.$$

Solving this equation, we get a number whose length, it turns out, we can construct. Then we can find point B, and the required point A_1 can be constructed as an intersection point of the given circle ω and the circle with center A_0 and radius x. There are two such points—that is, two solutions, A_1' and A_1'' (fig. 5).

But our equation has two roots: $x_1 = (-1 + \sqrt{5})/2$ and $x_2 = -(1 + \sqrt{5})/2$. The second root is negative and for this reason should be ignored, or so it seems. However, rather than rush to discard this root, let's try to understand its geometric sense.

Let us redraw figure 5, assuming that point B lies to the left rather than to the right of O, at a distance of $|x_2|$. We'll get figure 6, which has two new possible positions A_1''' and A_1^{IV} for the point A_1.

All in all, we found four different possibilities for A_1, which result in two different decagons—a convex one and a star-shaped one. The vertices of each can be numbered in two

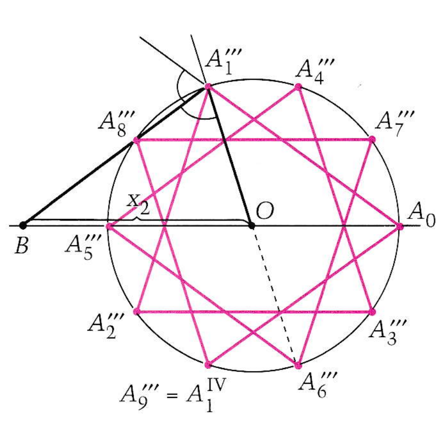


Figure 6

---

[2]By the way, constructions have in fact been given a computer setting by such recent software as "Geometer's Sketchpad" and "Cabri Geometry."—*Ed.*

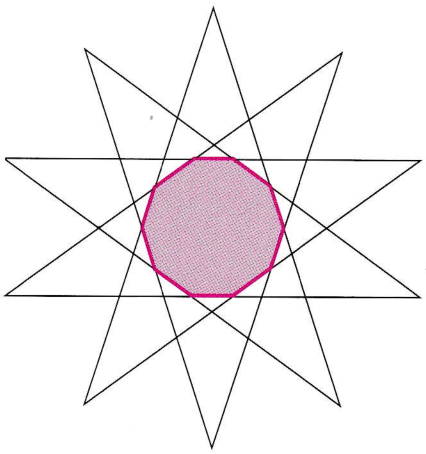


Figure 7

different ways (see figures 5 and 6).

Notice that from the "point of view" of the straightedge and compass the star-shaped decagon is as legitimate as the convex one.

You may argue that nonadjacent sides of the convex decagon are disjoint, whereas in the star-shaped decagon they intersect. But this objection falls away if a "side" of a polygon is understood as the entire line joining two vertices, not as the segment between them (we don't deal with "betweenness"!). Then the correct drawing of a "convex" decagon will differ from the star-shaped decagon's depiction only in size (fig. 7).

A similar picture emerges in the case of pentagons. Here we also obtain four solutions yielding two different pentagons (fig. 8) with two different numerations on each.

Now, without actually constructing an arbitrary regular n-gon we can try to establish the number of solutions to this problem for a given n. (Recall that the circumcircle ω and vertex A_0 on it are considered fixed.) Denote by x the length of the arc A_0A_1. Point A_1 is a solution to our problem (from the compass's "viewpoint") if, consecutively marking off arcs of length x, starting from A_0 and doing it n times, we arrive back at point A_0, whereas by doing this fewer than n times we can't come back to A_0. (The last condition is essential—otherwise, in the case of, say, $n = 6$, we would have to count as a "regular inscribed hexagon" such objects as a triangle traced twice, a diameter traced three times, or even one point A_0 repeated six times.)

In arithmetic terms, assuming that the circumference is of unit length, the condition for x can be formulated as follows: *nx is an integer and the numbers x, $2x$, $3x$, ..., $(n - 1)x$ are not integers.*

For instance, if $n = 10$, then x can be taken to be equal to $1/10$. But this isn't the only choice. Although the values $2/10 = 1/5$, $4/10 = 2/5$, $5/10$, $6/10$, and $8/10$ don't satisfy our condition for x, we can take $x = 3/10$, $7/10$, or $9/10$. These values correspond to the four solutions found above geometrically. Notice that the value $x = 11/10$ (as well as $13/10$, $17/10$, ...) doesn't yield any new geometric solutions—the location of the point corresponding to $x = k/n$ on the circle depends on the remainder of k when divided by n rather than the number k itself.

It's clear that the number mx $(0 < m \le n)$ is an integer (that is, hits the initial point on the circle) only for $m = n$ if and only if x is an irreducible fraction k/n $(k < n)$. This means that each number less than and relatively prime with n gives a solution to the problem of the regular n-gon. So the number of solutions to this problem is the Euler function $\phi(n)$!

In particular, $\phi(3) = \phi(4) = \phi(6) = 2$, $\phi(5) = \phi(10) = 4$ in accordance with the results obtained above geometrically. Now we recall that any solvable straightedge-and-compass construction problem must have 2^l solutions. This leads to a convenient condition for the constructibility of a regular n-gon:

A regular n-gon is constructible with straightedge and compass if and only if $\phi(n) = 2^l$ for a certain integer l.

(For instance, it's impossible to construct a regular heptagon, because $\phi(7) = 2^l$ is not a power of two.)

I've tried to explain why this condition is necessary. The fact that it's also sufficient is a separate result,[3] and I'm not going to deal with it here.

Fermat numbers

However, the problem with which we started is not completely exhausted yet. The question "What are the numbers n satisfying $\phi(n) = 2^l$?" remains open.

Of course, any specific number can fairly quickly be attributed to either the "red" or "black" kind (recall our list on page 11): all we have to do is compute $\phi(n)$. But this doesn't give any general description of the entire collection of "red" numbers. In searching for such a description we encounter a difficult, thus far unsolved problem in number theory. I'll briefly explain the gist of it.

Factor the number n:

$$n = p_1^{m_1} \cdot p_2^{m_2} \cdot \ldots \cdot p_k^{m_k},$$

where p_1, p_2, ..., p_k are different primes, and compute $\phi(n)$. From properties 1 and 2 of Euler's function,

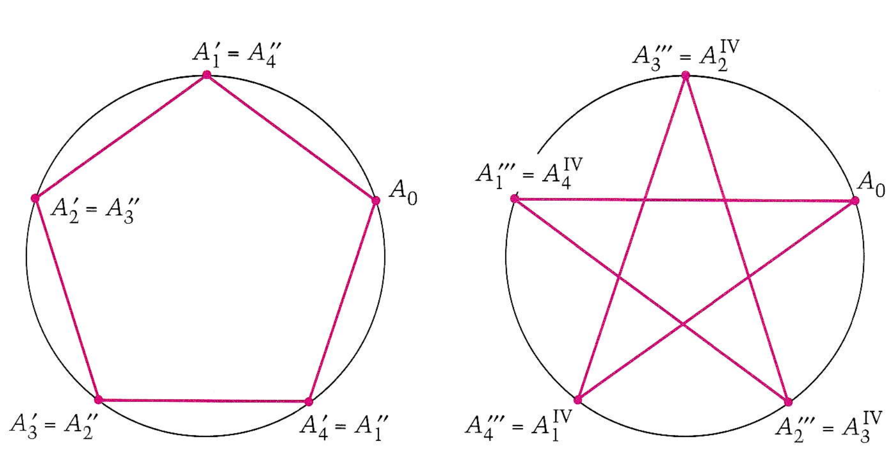


Figure 8

[3]The one proved by Gauss.—*Ed.*

we obtain

$$\begin{aligned}\phi(n) &= \phi\left(p_1^{m_1}\right)\cdot\phi\left(p_2^{m_2}\right)\cdot\dots\cdot\phi\left(p_k^{m_k}\right)\\ &= p_1^{m_1-1}\cdot p_2^{m_2-1}\cdot\dots\\ &\quad\cdot p_k^{m_k-1}(p_1-1)(p_2-1)\cdots(p_k-1).\end{aligned}$$

The last expression is a power of two if each odd prime factor p_i enters the factorization with the exponent $m_i = 1$ and is of the form $p_i = 2^l + 1$. On the other hand, a number of the form $2^l + 1$ can be a prime only if l is a power of two. (If l is divisible by an odd number $m > 1$, then $2^l + 1$ is divisible by $2^{l/m} + 1$, because $a^m + 1 = (a + 1)(a^{m-1} - a^{m-2} + a^{m-3} - \dots + 1)$ for any a—in particular, for $a = 2^{l/m}$.) Thus, each odd factor $p_i = 2^{2^k} + 1$.

Numbers of the form $2^{2^k} + 1$ are called *Fermat numbers*. The first five of them (for k = 0, 1, 2, 3, 4) are 3, 5, 17, 257, and 65,537. They are indeed primes. Euler discovered that the sixth Fermat number, $2^{2^5} + 1$, is divisible by 641.

Since Euler's time, Fermat numbers have been a matter of interest to many mathematicians all over the world. One of the sessions at the St. Petersburg Academy of Sciences in 1878 was dedicated to a report made by I. F. Zolotaryov about a work submitted to the Academy by the priest Ioann Pervushin. This work established the divisibility of $2^{2^{23}} + 1$ by $167{,}722{,}161 = 5 \cdot 2^{25} + 1$.

Nowadays, numbers are investigated by computer. Many Fermat numbers have been examined recently, but no primes were discovered among them, so it remains unknown whether there are any Fermat primes other than the first five. So at this point I can only formulate the answer to our problem in the following, not yet final form:

A regular n-gon can be constructed with straightedge and compass if and only if $n = 2^s \cdot p_1 \cdot p_2 \cdot \dots \cdot p_k$, where p_i are pairwise distinct Fermat primes.

Perhaps one of you reading this article will make your own contribution to a complete solution of this very interesting and difficult problem.

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M166

Tangential intersections. Two circles intersect at points A and B. The tangents to them drawn through A meet them again at M and N. The lines BM and BN meet the circles for the second time at points P and Q, respectively (fig. 1).

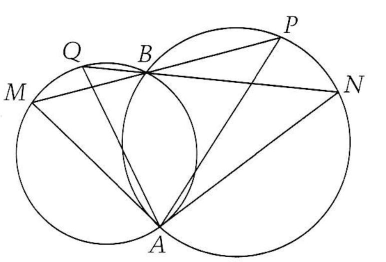


Figure 1

Prove that $MP = NQ$. (I. Nagel)

M167

Erase by raising. Does there exist a polynomial with at least one negative coefficient such that, when raised to any power n, $n > 1$, it will have only positive coefficients? (O. Kryzhanovsky)

M168

Arithmetic of a triangle. The side lengths x, y, z of a triangle are integers, and one of its altitudes is equal to the sum of the other two. Prove that $x^2 + y^2 + z^2$ is the square of an integer. (D. Fomin)

M169

Battleships. In the classic game of battleships each player starts with a 10×10 square grid ("ocean") on which a fleet of ten ships must be hidden: one battleship measuring 1×4, two cruisers measuring 1×3, three destroyers measuring 1×2, and four submarines measuring 1×1. No two ships can have points in common (even corners), but they can border on the sides of the ocean. Prove that (a) if the vessels are drawn in the order given above (starting with the battleship), it will always be possible to fit them all into the grid, even if at any step we take care only of the next ship rather than of the subsequent ones; (b) if the ships are arranged in the opposite order (starting with the submarines), then it may happen that there's no room for the next ship (give an example). (K. Ignatyev)

M170

Ph.D.'s in sepulation. In a paper on *sepulkas* Prof. Tarantoga gave n definitions of sepulation. His graduate students proved step by step that all these definitions were equivalent to one another. Each of them defended a doctoral dissertation proving that sepulation in the sense of the ith definition implies sepulation in the sense of the jth definition. What is the greatest number of graduate students that the professor could have if the dissertations were defended one by one and their main results could never have been derived directly from those defended earlier? (K. Mishachyov)

Physics

P166

Snake in a tube. A snake crawled halfway into a stationary narrow tube lying on a horizontal plane. The other half of the snake's body is coiled arbitrarily on this plane. Taking the snake to be a thin homogeneous cord of length l, determine which region of the plane could contain the snake's center of mass. (V. Gorbunova)

P167

Bar on rollers. A homogeneous bar of mass M and length l begins to move downward along an inclined plane that makes an angle α with the horizontal. The initial portion of the inclined plane of length l is occupied by closely packed rollers made of tubes of mass m and radius $r \ll l$ (fig. 2), which rotate on ball bearings without friction. The rest of the inclined plane

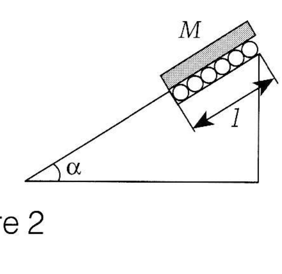


Figure 2

is frictionless. Find the dependence of the bar's acceleration on the position of the moving bar. (A. Stasenko)

P168

Vapor absorption. After the air has been pumped out of a 1-liter vessel to a very low pressure, 1 gram of water is still present. In order to remove it, an absorbent substance is used to take up the free water molecules. The total surface of the absorbent S = 100 m^2, and the area from which the water evaporates s = 0.001 m^2. The temperature of the vessel T = +5°C, and the saturated vapor pressure for water at this temperature P = 870 Pa. How much time is needed for all the vapor to be absorbed? Assume that the absorbent takes up all the molecules that come into contact with it. How much time would it take for all the water to evaporate if there were no absorbent? (D. Makarov)

P169

Capacitor and spring. One plate of a parallel plate capacitor of area S is suspended from a spring, while the other is firmly fixed. Initially the distance between the plates is l_0. The capacitor is briefly connected to a battery, which charges it to a voltage V. What must the spring constant be in order to prevent the plates from touching? Disregard any displacement of the upper plate during the charging. (1975 All-Union Physics Olympiad)

P170

Light in low orbit. The atmospheric refractive index of a planet X of radius R decreases with the altitude h above its surface according to the relation $n = n_0 - \alpha h$. Find the altitude h_0 of the optical channel where light rays will circle the planet at a constant height. (N. Sedov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

What's happening?

Summer study ... competitions ... new books ... ongoing activities ... clubs and associations ... free samples ... contests ... whatever it is, if you think it's of interest to *Quantum* readers, let us know about it! Help us fill Happenings and the Bulletin Board with short news items, firsthand reports, and announcements of upcoming events.

What's on your mind?

Write to us! We want to know what you think of *Quantum*. What do you like the most? What would you like to see more of? And, yes—what *don't* you like about *Quantum*? We want to make it even better, but we need your help.

What's our address?

Quantum
National Science Teachers Association
1840 Wilson Boulevard
Arlington VA 22201-3000

Be a factor in the QUANTUM equation!

From mouse to elephant

Cell size and other zoological constants

by Anatoly Mineyev

THE FAUNA ON OUR PLANET Earth is incredibly diverse. Nevertheless, among the different parameters characterizing living things there are those that vary little compared to the broad range of masses assigned to different animals. Let's refer to them as "zoological constants." A short list of these constants (from the mouse to the elephant) is shown in table 1.

First we should explain what we mean by the word "constant" in the context of the zoological world. There are, of course, much longer cells in the bodies of animals— for instance, nerve fibers. However, in the midst of all the other cells they are negligibly few and far between. Similarly, the temperature of diseased animals may increase sharply, or the proportion of muscular mass may differ in animals who engage in different physical activities. So the data in table 1 relate to the average and most numerous representatives of each species of animal. In other words, the distribution of the probability of a certain value for a zoological parameter is generally a curve with a clear maximum, which is the characteristic value of the zoological variable. Thus, we use the term "zoological constant" in a somewhat different sense than say, "physical constant" (like the speed of light, the mass of an electron, and so on, which are quite precise values).

This article will mainly consider the nature of cell size in animals. Other questions will be touched only slightly, among them the most enigmatic constant—the heart resource (one billion strokes in a mammal's life).

"One needn't think about pocket money," said Ostap Bender in Ilf and Petrov's *The Twelve Chairs.* "There's always some lying about, and we'll just pick some up along the way as need be." We'll take this advice and also try to "pick up" data about cells when needed, working our way to our goal step by step.

Table 1

Cell diameter	$d_c \sim 10$–$20\ \mu m$
Ratio of longevity to heart cycle	$t_l/t_h \sim 10^9$
Ratio of respiratory cycle to heart cycle	$t_r/t_h \cong 4$
Body temperature	$T_{bd} \cong 37$–$38°C$
Ratio of organ mass to body mass (m_{bd}): heart lungs blood skeleton muscles	 $m_h/m_{bd} \cong 0.6\%$ $m_l/m_{bd} \cong 1\%$ $m_{bl}/m_{bd} \cong 5\%$ $m_s/m_{bd} \cong 6\%$ $m_m/m_{bd} \cong 40\%$

And now, let's try to answer two simple questions:

Why is the average diameter of a mammalian cell d_c 10–20 μm and not, say, 1 or 100 μm?

Why is d_c approximately the same for all mammals, while their masses differ tremendously? For example, the mass of a shrew is 3 g, while the mass of an elephant is 3 tons—that is, the range of mammalian mass is as much as six orders of magnitude.

We'll try several approaches in our attempt to estimate the characteristic cell size. Let's start with one that doesn't use any information about the structure and functions of the cell.

Taking a stab at it

A cell must be much larger than an atom ($\sim 10^{-10}$ m) and much smaller than "human size" (~ 1 m). The first postulate allows us to ignore quantum effects; the second allows us to use cells like bricks to

At right: "I love you, people."

Art by Leonid Tishkov

Я ЛЮБЛЮ ВАС, ЛЮДИ
Тишков

build the complex structure of a living organism with diverse functions.

The geometric mean of the "atomic" and "human" sizes satisfies the demands of the postulates and results in the correct order of magnitude—

$$(10^{-10}\ \text{m} \times 1\ \text{m})^{1/2} = 10^{-5}\ \text{m}$$

—but it doesn't clarify the problem. Moreover, if we insert the characteristic sizes of the shrew and the elephant into this formulation, we get 2 μm and 20 μm, which amounts to a fundamental difference in cell size. So we can't get by without some information about cell structure, however primitive.

Let's go at it from the opposite direction and take a "culinary" approach.

Baking a cell from scratch

Let's construct a mammalian cell from the simplest component parts and estimate its size. (Just like in cooking: for one serving you need one egg, a tablespoon of sugar, a cup of milk . . .)

Now, what kind of object are we dealing with? The cell is the elementary unit of life. Basic information about the cell itself as well as the organism as a whole is written in the DNA molecule and stored in the cell's nucleus. This information is processed by means of RNA molecules, which leads to the synthesis of proteins and other necessary substances. The energy for this synthesis is accumulated in the mitochondria. Water is the physical medium where all cellular processes occur. Membranes keep both the cells and their inner organelles separate from one another.

Table 2

Diameter of double helix	$2 \cdot 10^{-9}$ m
Distance between pairs of bases	$3.4 \cdot 10^{-10}$ m
Number of nucleotide pairs in a mammalian cell	$(1 \text{ to } 5) \cdot 10^{9}$
Mass of nucleotide pair in atomic units ($1.67 \cdot 10^{-27}$ kg)	~500

DNA will be our starting point. Some of its parameters are given in table 2. According to the table, the length of DNA molecules in different mammals is within the range

$$(1 \text{ to } 5) \cdot 10^9 \times 3.4 \cdot 10^{-10} \cong 0.3 \text{ to } 2\ \text{m},$$

and its mass is about

$$1.67 \cdot 10^{-27} \times 500 \times (1 \text{ to } 5) \cdot 10^9 \cong (0.8 \text{ to } 4) \cdot 10^{-15}\ \text{kg}.$$

What volume can such a molecule occupy? If we wound the DNA into a very tight ball with a gap between individual layers of the order of the distance between bases (that is, $4 \cdot 10^{-10}$ m), we would get a volume of

$$V \sim 2 \cdot 10^{-9} \times (1 \text{ to } 5) \cdot 10^9 \times 3.4 \cdot 10^{-10} \cong (0.3 \text{ to } 1.4) \cdot 10^{-18}\ \text{m}^3$$

and a characteristic size of

$$d \sim V^{1/3} \cong 0.7\text{–}1.1\ \mu\text{m}.$$

The density of this ball of DNA is $\sim 3 \cdot 10^3$ kg/m^3.

Upon reflection, these figures seem questionable:

- Since the density of this nucleus far exceeds that of water, it might "sink to the bottom" of the host cell if it's not held somehow near the cell's center;
- DNA isn't a flat structure—it's a helix, and so it can't be wound as tightly as we supposed above;
- It would be impossible to quickly unwind such a ball.

These considerations suggest that the DNA must be wound more loosely.

It's reasonable to suppose that the average density of a ball of DNA in the volume enclosed by the membrane of the nucleus must approach that of water—that is, the nucleus floats more or less freely in the cell. In this case, the volume of the nucleus of a mammalian cell is $\sim(0.8 \text{ to } 4) \cdot 10^{-18}$ m^3 and the diameter of the corresponding sphere is ~1.2–2 μm. The membrane that confines the nucleus is rather thin ($\sim 10^{-8}$ m) and contributes a negligible amount to the volume of the nucleus.

Now let's mix in the rest of our ingredients. We need several times more RNA than DNA. And the mass of other substances (proteins and other parts of the cell synthesized by means of DNA and RNA) should exceed by far the mass of the RNA. For our estimate we'll assume that the ratio of the masses of the proteins and the DNA is about 10 : 1. All this content held by the membranes must float in water. We'll suppose that there is four times more water by volume than the other ("dry") ingredients of the cell.

All told, the minimum volume of mammalian cells must be

$$(V_c)_{\min} \sim (0.3 \text{ to } 1.6) \cdot 10^{-16}\ \text{m}^3,$$

and the corresponding diameter of the sphere is

$$(d_c)_{\min} \sim \left(\frac{6V_c}{\pi}\right)^{1/3} \cong 4\text{–}7\ \mu\text{m}.$$

That is to say, mammalian cells with diameters less than 4 μm cannot exist. This result is a decided improvement over our earlier "stab" at an estimate.

In addition, our culinary approach provides partial answers to our main questions posed at the outset:

- The range of minimum cell diameters isn't large—it's about a factor of two for all mammals (in contrast to DNA, whose length varies by a factor of five);
- The diameter of the mammalian cell isn't 1 μm—it's much larger.

So our culinary approach has resulted in a recipe for baking a mammalian cell. To obtain 100 parts of a living cell, you need (by weight)

84 parts water
7 parts protein
4 parts each lipids
4 parts carbohydrates
a pinch of RNA (0.7 part)
a pinch of DNA (0.3 part).

The volume of such a cell is $4 \cdot 10^{-15}$ m^3, its diameter is 20 μm, and its mass is $3.5 \cdot 10^{-12}$ kg.

Now let's turn our attention to one of the most important functions of a cell: the exchange of substances with the outside world. For a cell to stay alive, it needs to obtain oxygen and fuel and it needs to get rid of carbon dioxide, the products it has synthesized, and garbage.

We should note that not only is the characteristic size of mammalian cells the same, the diameters of erythrocytes and capillary vessels are roughly the same as well. Mammalian erythrocytes, which take part in cellular gas exchange, vary in size from 5 μm to 10 μm. The diameter of capillaries is 3–30 μm. The closeness of these values has a profound basis: at a distance of about 10 μm, a drastic change occurs in the character of substance transport in an organism. So, we'll name our last approach to estimating the size of cells, erythrocytes, and capillaries, rather enigmatically, the "convection and diffusion approach."

Watching the cell at work

First let's clarify what we mean by "convection" and "diffusion." How does one ensure the ready supply of substances to a great number of cells? One way is to organize the directed motion of substances along with the flow of blood—that is, by convection. This is what happens when blood flows through the blood vessels of animals. Convection is quite effective when the blood flows in large vessels. However, when it moves away from the heart and approaches the "consumers" (cells), the network of blood vessels branches more and more, and the diameter of the vessels becomes smaller and smaller. As this happens, the blood must necessarily flow more slowly. This is because the pressure of a viscous fluid decreases as it flows in a vessel. The drop in pressure is proportional to the fluid's speed and the length of the vessel and inversely proportional to its cross-sectional area.[1]

To prevent the pressure losses from becoming too large (in which case the blood wouldn't flow at all through such a channel), the blood flow must decrease with the diameter of the vessels. The same is true for the vessel's length. Indeed, these tendencies are realized in the vascular system of animals. As an example, figure 1 shows the dependence of average convective speed on the vessel's diameter in humans. The relation is approximately linear:

$$v_c \sim (20 \text{ to } 100)d.$$

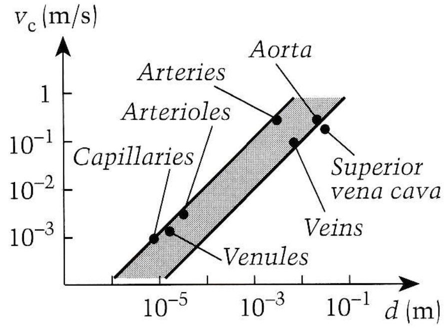


Figure 1
Dependence of the rate of convective blood flow on the diameter of blood vessels in humans.

Thus, blood flow indeed decreases with vascular diameter. In a certain sense, convection loses its effectiveness in smaller vessels.

Now, what about diffusion? Diffusion is the transfer of substances due to the chaotic wandering of molecules. During random motion, a molecule now moves away from its starting point, now moves toward it. As this happens, the mean distance from the molecule to its starting point is proportional not to t (as with linear uniform motion) but to $\sqrt{t}$:

$$d = \sqrt{2Dt},$$

where D is the diffusion coefficient.[2] Using t from this equation, we get the average speed of diffusion for a distance d:

$$v_d \sim \frac{d}{t} \sim \frac{D}{d}.$$

Thus diffusive transfer slows at large distances, but it's very effective at short range. Comparing the formulas for the speeds of convection and diffusion (v_c and v_d, respectively) yields a certain characteristic distance d_0 (fig. 2)—

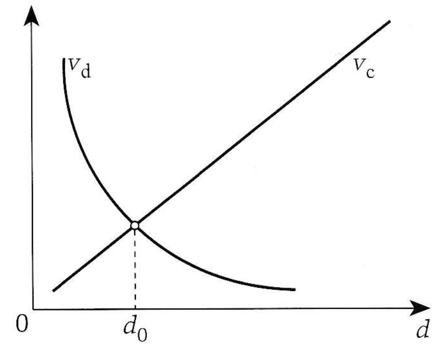


Figure 2
Rates of convection (v_c) and diffusion (v_d) versus characteristic distance d.

$$d_0 \sim \left(\frac{D}{20 \text{ to } 100}\right)^{1/2}$$

—at which both speeds are equal.

The diffusion coefficient for substances in water is of the order $D \sim 10^{-9}$ m^2/s, so $d_0 \sim 3$–7 μm. And this is indeed the scale of capillaries, cells, and erythrocytes. In other words, convective transport in living organisms dominates down to a distance d_0, after which the major role is played by diffusion.

Our new estimate is consistent with our culinary estimate. This is indirect evidence that the cell has limited "choices" when it comes to its size. The difference between 3–7 μm (our estimate) and 10—20 μm (its actual size) can be explained by simplifications we've used in our approach. It's the price we sometimes pay for clarity.

I'll conclude with some comments about table 1. The underlying nature of some of the items in it isn't entirely clear at present.

1. The heart resource is about one billion strokes in a lifetime. This figure corresponds to the world of mammals. However, it seems to be too small for human beings—a billion strokes at the rate of one per second (the human heart rate) would translate to a life span of only 30 years. It is currently two to three times that. At the end of the last century, our life expectancy was

[1]You can read about this in more detail in "Trees Worthy of Paul Bunyan" by the same author in the January/February 1994 issue of *Quantum.—Ed.*

[2]The details of diffusion are described in "Airplanes in Ozone" by Albert Stasenko in the May/June 1995 issue of *Quantum.—Ed.*

pretty much "according to spec" (that is, table 1). In the intervening years we have made much progress in eliminating certain diseases, reducing infant mortality, and improving our work conditions and life styles. Among mammals, only human beings have managed to improve their statistical standing.

The heart rate decreases as the body mass increases. For example, the mouse has a heart rate of 600 beats per minute and lives for three years. The corresponding figures for an elephant are 30 beats per minute and 60 years.

2. The heart resource seems to be extremely large. Indeed, among moving mechanisms in inanimate nature only clocks might compete with it, but unlike the heart, they need periodic cleaning and repair.

3. The last three lines in table 1 reflect certain optimal relationships worked out during the long course of evolution.

In the case of temperature, its increase results in a rapid increase in the activity of enzymes catalyzing the metabolic processes. On the other hand, proteins begin to disintegrate at temperatures of 40–45°C. Thus, the temperature range 37–38°C is the best one for enzymatic processes in land mammals.

The relation between the respiratory and heart cycles ($t_r/t_h \cong 4$ for all mammals) must follow from the fact that the erythrocytes in the blood carry oxygen from the lungs and bring carbon dioxide back to them. Thus the respiration and heart rates must be closely linked. However, this relationship isn't direct.

Scientists still don't know why the relative masses of the heart, lungs, muscles, and certain other organs is constant. For some of them (muscles and bones) it could be explained in principle by the need to move to find food. At the same time the relative mass of some other important organs (kidneys, liver, and brain, in particular) increases as the body size decreases.

Many interesting questions remain open for the next generation of researchers. ◘

How many divisors does a number have?

"The newly harrowed vast expanses
So evenly are spread about,
As though the valley had been spring-cleaned
Or else the mountains flattened out."
—Boris Pasternak[1]

by Boris Kotlyar

HAVE YOU EVER WONDERED why the number one is usually not considered a prime? There's something unclear about it: this number is divisible only by itself and by one, so it satisfies the definition of a prime, doesn't it?

There were reasons why it came to be accepted that one isn't a prime—the same sort of reasons that make us consider any straight line parallel to itself.[2] If coincident lines aren't assumed to be parallel, we have to consider the case of the coincidence of lines separately in our phrasing of many geometric facts. But if we agree that a line is parallel to itself, we can state these results without exclusions.

The number one had been listed among the primes for a long time, but it was deprived of this title for quite practical reasons. It's very convenient to have a *unique* factorization of any positive integer into primes. But if one is counted a prime, this statement becomes false.

For example, let's factor some number—say, 84—into primes:

$$84 = 2 \cdot 2 \cdot 3 \cdot 7 = 2^2 \cdot 3 \cdot 7.$$

Can it be factored differently? Of course. We can rearrange the factors, but these factorizations are naturally considered the same. The fact that we can't obtain anything essentially different follows from the so-called Fundamental Theorem of Arithmetic, according to which *any natural number (positive integer) can be factored into primes, and this factorization is unique* (*up to the order of the factors*). That is, a natural number N is uniquely representable in the form

$$N = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k},$$

where $p_1, \dots, p_k$ are primes and $\alpha_1, \dots, \alpha_k$ are natural numbers. This is called the *canonical factorization* of the number N.

Let's get back to our example. The number 2 enters the factorization of 84 in the second power, 3 and 7 in the first power. And in what power does the prime factor 5 enter this factorization? In "no power"—that is, in the power zero.

So we can assume that all prime numbers enter any factorization, but some of them in the zeroth power. Of course, such "ghost factors" aren't written out (as a rule).

Now we see why it's not convenient to consider 1 a prime: this number can be included in any factorization in any power:

$$\begin{aligned} 84 &= 1^5 \cdot 2^2 \cdot 3 \cdot 7 \\ &= 1^{100} \cdot 2^2 \cdot 3 \cdot 7 \end{aligned}$$

and so on, which violates the uniqueness.

There are a number of other arguments—simple as well as rather sophisticated—in favor of excluding one from the primes. For instance, let's write out the few first natural numbers in a row and the number of divisors for each one (counting only different divisors of each number). The number of divisors of a number n is denoted $\tau(N)$. For $N = 1, 2, \dots, 12,$

N	1	2	3	4	5	6	7	8	9	10	11	12
$\tau(N)$	1	2	2	3	2	4	2	4	3	4	2	6

[1]Translated by Lydia Pasternak Slater.

[2]Perhaps this is a statement you aren't likely to read in a school textbook, but in more "serious" geometry texts. As a rule, parallelism is understood in this wider sense.—*Ed.*

the numbers $\tau(N)$ are given in the table below.

We see that the number 1 has only one divisor, while all the other numbers have more (primes have exactly two divisors). So it's reasonable to isolate the number 1 in a special class of natural numbers—neither prime nor composite.

Divisors of positive integers

Can the function $\tau(N)$ be expressed analytically? Yes, it can, and the expression is simple enough. Let's derive this formula. Write out the canonical factorization of a number $N > 1$:

$$N = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$$

($p_1, p_2, \dots, p_k$ are different primes, $\alpha_1, \alpha_2, \dots, \alpha_k$ are their respective exponents).

Theorem. *If $N = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ is the canonical factorization of a number N, then*

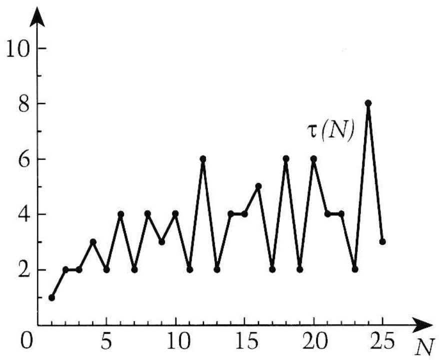
$$\tau(N) = (\alpha_1 + 1)(\alpha_2 + 1)\dots(\alpha_k + 1).$$

Proof. Any natural number that divides N has the form $p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k}$, where $0 \le \beta_1 \le \alpha_1$, $0 \le \beta_2 \le \alpha_2$, …, $0 \le \beta_k \le \alpha_k$. For instance, if $\beta_i = 0$ for all i, the corresponding divisor is 1, and if $\beta_i = \alpha_i$ for all i, the divisor is N itself. So how many such products can be formed? The exponent β_1 takes exactly $\alpha_1 + 1$ values: 0, 1, 2, …, α_1; β_2 takes $\alpha_2 + 1$ values; and so on. So there are $\alpha_1 + 1$ divisors of the form $p_1^{\beta_1}$, $\alpha_2 + 1$ divisors of the form $p_2^{\beta_2}$, and consequently $(\alpha_1 + 1)(\alpha_2 + 1)$ divisors of the form $p_1^{\beta_1} p_2^{\beta_2}$. Proceeding in the same way, we arrive at the required result.

Using this formula, we can find the number of divisors of any natural number, but only after factoring it into primes in order to find the exponents $\alpha_1, \alpha_2, \dots, \alpha_k$. But this may not be very easy—if a number is big, it's hard even to understand whether it's prime or composite, to say nothing of finding its canonical factorization.

This isn't the only disadvantage of our formula. The behavior of the function $\tau(N)$ is chaotic. On the one hand, $\tau(p) = 2$ for any prime p, and since there are infinitely many primes, we'll encounter twos in the second row of our table arbitrarily far. On the other hand, the number of divisors can become arbitrarily large—it suffices only to take a number n with large enough exponents $\alpha_1, \alpha_2, \dots, \alpha_k$ (even with only one sufficiently large exponent). The graph of $\tau(N)$ is plotted in figure 1 (it actually consists of isolated points, but we joined them to make its haphazard nature easier to grasp). Now you see these "mountains and valleys"!

So the exact formula is of little use—our function is too much irregular. Can we find a more graphic, even if approximate, formula that would directly show what's to be expected from $\tau(N)$?



Let's see how $\tau(N)$ behaves "on average." Consider the arithmetic mean $\bar{\tau}(N)$ of the numbers of divisors of the first N natural numbers:

$$\bar{\tau}(N) = \frac{1}{N}\big(\tau(1) + \tau(2) + \dots + \tau(N)\big).$$

This function happens to have a nice formula. It's not absolutely precise, but it expresses the "average number of divisors" in terms of a well-known function:

$$\bar{\tau}(N) \cong \ln N.$$

Why logarithms?

At first glance the appearance here of a logarithm is a bit odd. But in fact it should come as no surprise. For example, for $N = 2^n$ we have

$$\begin{aligned}\tau(N) = \tau(2^n) &= \log_2 N + 1\\ &= \log_2 N + \log_2 2\\ &= \log_2 (2N).\end{aligned}$$

This particular example may not seem too convincing. After all, numbers of this form (powers of primes) are rather rare, and besides, the base of the logarithm here is 2 rather than e. Nevertheless, we'll manage to prove our formula later. First I want to refine it. What does approximate equality mean here? There exists a constant number μ ($\mu \cong 0.154$) such that

$$\bar{\tau}(N) = \ln N + \mu + a_N,$$

where a_N is an infinitesimal sequence—that is, its limit is 0 as N approaches infinity. When N is large, the "correction term" a_N becomes arbitrarily small and negligible compared to the constant μ, and even more so compared to the infinitely increasing $\ln N$. This is what the approximate equation $\bar{\tau}(N) \cong \ln N$ means.

Average number of divisors

The divisors of a number can be visualized by means of coordinates. Take, for instance, the number 12. Write out all its divisors:

$$1, 2, 3, 4, 6, 12.$$

Consider the function $f(x) = 12/x$. You certainly know that its graph is a hyperbola. We'll need only one of its two branches—the one in the first quadrant. To sketch the graph, we can compute the values of the function for a number of the values of x: for $x = 1, 2, 3, \ldots$ we get $y = 12, 6, 4, \ldots$. It's convenient to compute y when x is a divisor of 12. Plot the points with integer coordinates (x, y) thus obtained and draw a curve through them (fig. 2).

Now let's count the points with integer coordinates in the vicinity of the origin and on the branch of the hyperbola we've constructed. (All these points are shown in figure 3.) There are exactly 6 of them—as many as the divisors of 12, because any natural divisor x is coupled with the natural y such that $xy = 12$.

Figure 4 illustrates both a branch of the hyperbola $y = 12/x$ and a branch of $y = 6/x$, which carries as many integer points as there are divisors of 6. But any integer point (x_0, y_0) in the first quadrant below the hyperbola $y = 12/x$ (except the points on the axes) lies on exactly one hyperbola $y = n/x$, where $n = x_0y_0 < 12$. For instance, point (1, 11) belongs to the hyperbola $y = 11/x$ and (2, 2) is on the

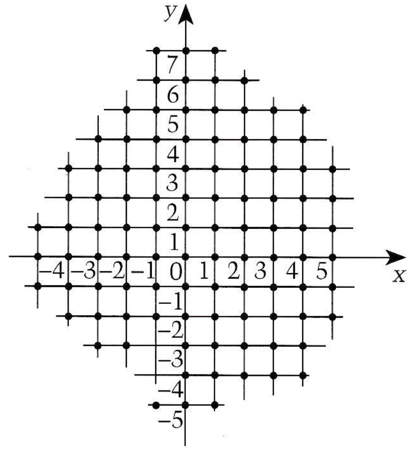


Figure 3

graph of $y = 4/x$. It follows that the total number of divisors of all natural numbers not exceeding 12 is equal to the number $S(12)$ of integer points in the first quadrant below and on the hyperbola $y = 12/x$ (excluding the points on the axes). This number can be written as

$$S(12) = \tau(1) + \tau(2) + \ldots + \tau(12).$$

Similarly, for any positive integer N,

$$S(N) = \tau(1) + \tau(2) + \ldots + \tau(N)$$

—the argument above remains valid without any change. Therefore,

$$\bar{\tau}(N) = \frac{1}{N}\big(\tau(1) + \cdots + \tau(N)\big) = \frac{S(N)}{N}.$$

Thus, the arithmetic problem of finding the average number of divisors of a number is replaced by the geometric problem of counting integer points in the first quadrant

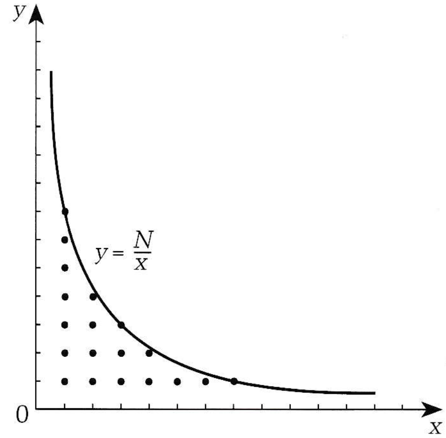


Figure 5

under (and on) the hyperbola $y = N/x$ (fig. 5).

It's difficult to solve this problem exactly, but an approximate solution is simple enough—and it will involve a little bit of calculus.

Consider the curvilinear trapezoid T bounded by the vertical lines $x = 1$ and $x = N$ on the left and right, by the x-axis $y = 0$ on the bottom, and by our graph $y = N/x$ on the top (fig. 6). The points we want to count are all the integer points in T except the N points $(1, 0), (2, 0), \ldots, (N, 0)$ on the x-axis. So there are $S(N) + N$ integer points in T altogether. Each of these points is the *left bottom* vertex of a certain unit square with integer vertices—all these squares are shaded in figure 7 (for the case $N = 6$).

You see that entire shaded area approximates the trapezoid T with a bit left over. We denote the area of T by A and will get an upper and lower bound for the difference $A - S(N)$.

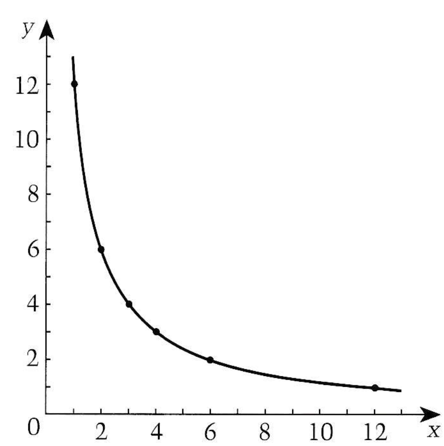


Figure 2

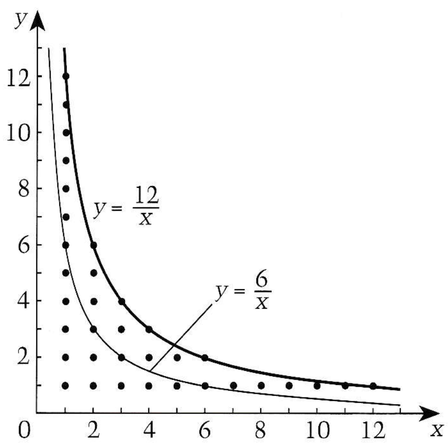


Figure 4

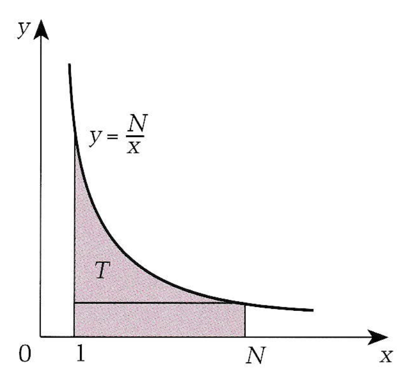


Figure 6

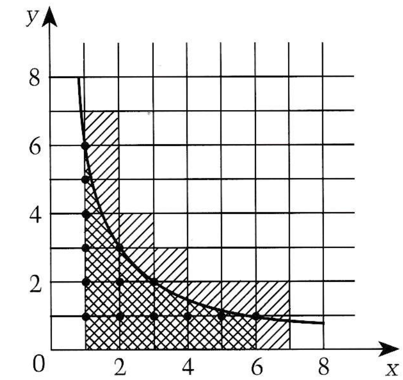


Figure 7

To get an upper bound, we note that any point (x, y) is covered by a square of the integer grid shown in figure 3. If there are several such squares, we choose the one that doesn't have the point on its top or right side. Let (n, m) be the left bottom vertex of this square. Then $nm \le xy \le N$, and our rule for choosing the square ensures that (n, m) is also in T, so (x, y) is covered by a shaded square. Noting that the total area of the shaded squares is equal to their number, we get $A \le S(N) + N$, or

$$A - S(N) \le N.$$

(As is clear from the figure, T can always be covered by a slightly smaller set of squares—we can harmlessly remove, say, the two rightmost shaded squares, and perhaps some others. This allows us to improve the estimate, but such an improvement isn't really significant, as we'll soon see.)

To estimate the difference $A - S(N)$ from below, consider the $S(N)$ integer grid squares whose *right top* vertices lie in the trapezoid T but not on the x-axis (the shaded squares in figure 8). Obviously, all these squares (except the column of N squares adjoining the y-axis) lie in T. (We can prove this using the fact that the function $y = N/x$ decreases for $x > 0$.) It follows that $A \ge S(N) - N$, or

$$A - S(N) \ge -N.$$

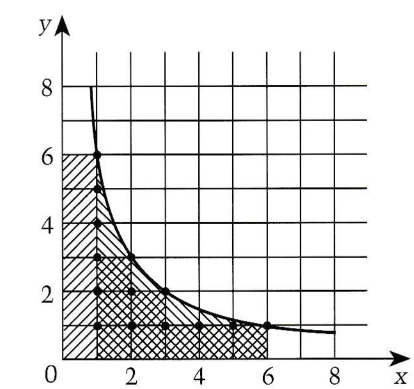
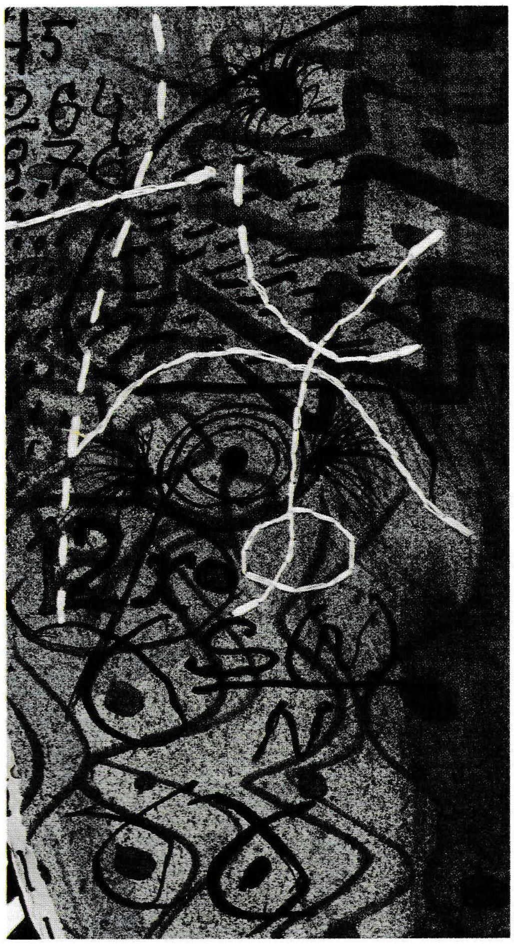


Figure 8



Combine the two estimates:

$$-N \le A - S(N) \le N,$$

or

$$|S(N) - A| \le N.$$

For readers familiar with calculus, computing the area A is an easy exercise:

$$A = \int_1^N \frac{N}{x}\,dx = N \ln x\Big|_1^N = N \ln N.$$

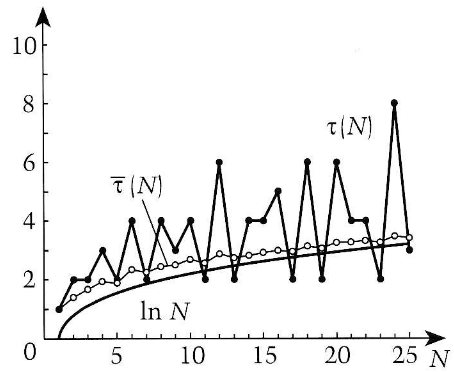


Figure 9

Those who haven't studied integration yet will have to take this formula as given.

At any rate, we have arrived at the relation

$$|S(N) - N \ln N| \le N,$$

or, after dividing it by N,

$$|\bar{\tau}(N) - \ln N| \le 1.$$

The function $\ln N$ grows indefinitely with N. Therefore, the same is true for $\bar{\tau}(N)$. But the difference between them remains bounded—it never exceeds 1. So the relative error of the approximation $\bar{\tau}(N) \cong \ln N$ becomes arbitrarily small with the growth of N.

Calculate the values of the functions $\tau(N)$, $\bar{\tau}(N)$, and $\ln N$ for small values of N and sketch their graphs in the same coordinate system (fig. 9). We see that the functions $\bar{\tau}(N)$ and $\ln N$ allow us to "spring-clean the valley and flatten out the mountains," to paraphrase Pasternak.

Dirichlet and the Divisor Problem

Earlier I mentioned a refinement of our approximate equality. It was obtained by the outstanding German mathematician Peter Gustav Lejeune Dirichlet (1805–1859), who invented the geometric approach we used. Dirichlet proved that

$$\bar{\tau}(N) = \ln N + (2C - 1) + a_N,$$

where C is the so-called Euler constant, defined as the limit

$$C = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln n\right).$$

(The proof that this limit exists is not too difficult.) The constant C is approximately equal to 0.577; the term $2C - 1 \cong 0.154$ in the expression for $\overline{\tau}(N)$ was denoted by μ above.

Dirichlet showed that the sequence a_N decreases rather rapidly to zero—there exists a constant a such that

$$|a_N| \le a \cdot N^{-1/2}.$$

In his proof, Dirichlet cleverly used the symmetry of the hyperbola with respect to the bisector of the first quadrant.

The prominent Russian mathematician Georgy Fedoseyevich Voronoy (1868–1908) proved that a_N decreases even faster: for any number $\varepsilon > 0$, however small it is,

$$|a_N| \le k \cdot N^{-2/3+\varepsilon}$$

for a certain constant k. Is it possible to replace the function on the right with a function that decreases still faster? What is the ultimate exponent? These questions constitute a famous, still unresolved problem: the Divisor Problem. As of today, theorems are known with an exponent of N that is smaller than in Voronoy's estimate, but the final result has not yet been achieved. However, the function on the right can't be reduced too much—the English mathematician Godfrey Harold Hardy (1877–1947) proved that the inequality is already violated for the exponent $-3/4$. There is a conjecture, neither proved nor refuted, that even the slightest increase of Hardy's prohibitive exponent yields a function that majorizes the sequence—that is,

$$|a_N| \le k \cdot a^{-3/4+\varepsilon}$$

for any $\varepsilon > 0$ and some constant k.

The Divisor Problem is one of the most interesting problems in number theory. Let's wonder again at the marvelous interweaving of mathematical notions that arose in our investigation: the number of divisors of a natural number—which is natural too, of course—proved to be connected with the hyperbola, integer points on the plane, areas, integration, and the natural logarithm.

Exercises

1. Prove that the number of divisors of N is odd if and only if N is a perfect square.

2. A function f defined on integers is called *multiplicative* if $f(ab) = f(a)f(b)$ whenever a and b are relatively prime. Prove that $\tau(N)$ is multiplicative.

3. Denote by $\tau_m(N)$ (for $m \ge 1$) the number of solutions to the equation $x_1x_2\dots x_m = N$ in natural numbers $x_1, x_2, \dots, x_m$; in particular, $\tau_1(N) = 1$, $\tau_2(N) = \tau(N)$. Try to prove the following statements, first, for $m = 3$, then for an arbitrary m:

(a) $\tau_m(N)$ is a multiplicative function;

(b) $\tau_m(p^\alpha) = \dfrac{(\alpha+1)(\alpha+2)\cdots(\alpha+m-1)}{1\cdot 2\cdot \dots \cdot(m-1)}$;

(c) the number of solutions to the inequality $x_1x_2\dots x_m \le N$ in positive integers $x_1, x_2, \dots, x_m$ equals $\sum_{m\le N}\tau_m(N)$.

A reader writes . . .

Hello, *Quantum*:

I look forward to every issue of your magazine. You always make me think, and as a result I learn. I am a physicist, head of the Teacher Institute at the Exploratorium in San Francisco, and I enjoy playing music by twirling corrugated tubes. So I must comment on your answer to question 7 in "Fluids on the Move" in the January/February issue.

Frank Crawford did an investigation of twirling corrugated tubes and reported the results in the *American Journal of Physics* (1974, Vol. 42, p. 278). His analysis agrees with my simple experiments. The motion of air through the tube is caused by both the difference in pressure between the ends of the tube and also by forces on the air in the noninertial frame of the rotating tube.

An easy experiment will allow you to separate these two effects and find their relative importance. In brief, the pressure difference does not produce enough air motion through the tube to make the tube sing. Instead, the rotation of the tube forces air through the tube and makes it sing.

Simply blow over one end of the tube, using your mouth or any blower you wish (we sometimes stick one end out the window of the car). When the air flows across the end of the tube (perpendicular to the end of the tube), the tube does not sing. The pressure decrease as the air speeds up over the end of the tube is not enough to move air through the tube to make it sing. Then blow directly into the tube. It will sing, even if you blow with just your mouth. (Hold your mouth an inch or so from the tube when you blow.)

To get a feel for how the rotation of the tube drives the air through the tube, picture the tube full of marbles, then picture the tube being twirled—the marbles will flow through and out of the tube.

Frank Crawford, in his fine article, does the calculation and shows how inertial forces in the rotating frame cause the air motion.

—Paul Doherty, Ph.D.

The golden ratio in baseball

Fibonacci strikes yet again!

by Dave Trautman

THE FAMOUS FIBONACCI SEquence 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ... is generated by setting $f_0 = 1$, $f_1 = 1$, and for $n \geq 0$, $f_{n+2} = f_{n+1} + f_n$. This sequence arises in a wide variety of branches of theoretical mathematics and in many areas of the natural sciences, as discussed by Elliott Ostler and Neal Grandgenett in their article "Fibonacci Strikes Again!" (*Quantum*, July/August 1992). Ostler and Grandgenett also show that the ratios of successive elements of the Fibonacci sequence tend to the number known as the "golden ratio," which is approximately 1.618:

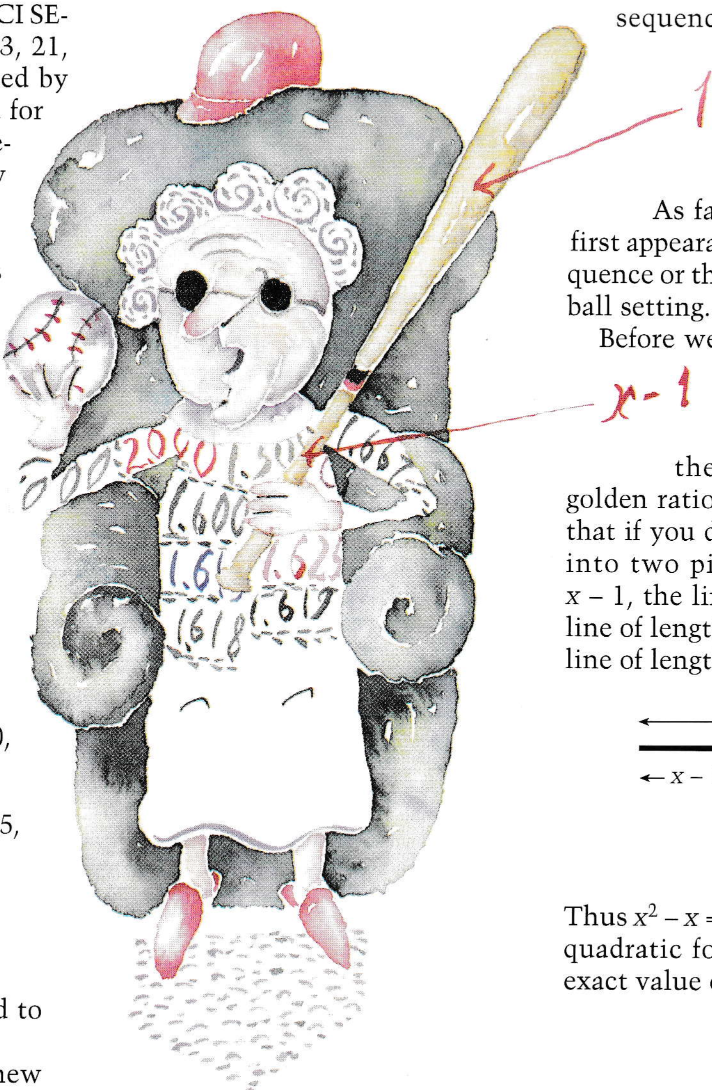
$$\frac{1}{1} = 1.000,\ \frac{2}{1} = 2.000,\ \frac{3}{2} = 1.500,$$

$$\frac{5}{3} \cong 1.667,\ \frac{8}{5} = 1.600,\ \frac{13}{8} = 1.625,$$

$$\frac{21}{13} \cong 1.615,\ \frac{34}{21} \cong 1.619, \ldots .$$

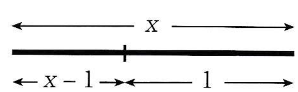
From then on all ratios round to 1.618.

Recently I came across a new appearance of the Fibonacci sequence and the golden ratio while reading *The Politics of Glory: How Baseball's Hall of Fame Really Works* by Bill James.



As far as I know, this is the first appearance of the Fibonacci sequence or the golden ratio in a baseball setting.

Before we see how these appearances arise, let me give the traditional description and exact value of the golden ratio. The golden ratio is the value of x such that if you divide a line of length x into two pieces of lengths 1 and $x - 1$, the line of length x is to the line of length 1 as this line is to the line of length $x - 1$:



$$\frac{x}{1} = \frac{1}{x-1}.$$

Thus $x^2 - x = 1$, or $x^2 - x - 1 = 0$. The quadratic formula then yields the exact value of the golden ratio:

$$x = \frac{\sqrt{5}+1}{2}$$

Art by Sergey Ivanov

(we reject, for now, the negative root of the equation).

Problem 1. Assuming that the limit of f_{n+1}/f_n as n goes to infinity exists, show that this limit must be the golden ratio.

Now to the world of baseball. In his book James discusses the problem of comparing the career won–lost records of pitchers. Of particular concern is the problem of comparing the record of a pitcher who had a fairly short but spectacular career and the record of a pitcher who had a longer but less spectacular career. For example, how does one compare Sandy Koufax's career record of 165 wins and 87 losses to the career record of 224 wins and 184 losses of Jim Bunning? On the one hand Bunning had 59 more wins than Koufax—a significant advantage. On the other hand, Bunning lost 97 more games than Koufax, and this seems to be a rather high price to pay for 59 more wins. Since the three most important statistics for pitchers are wins, winning percentage (wins divided by wins plus losses), and games over .500 (wins minus losses), James devised a formula that uses all three statistics.

If a pitcher has W wins and L losses, then credit him with $W \cdot W/(W + L) + W - L$ Fibonacci Win Points. Thus Koufax has $165 \cdot 165/252 + 165 - 87 = 186$ Fibonacci Win Points, while Bunning has $224 \cdot 224/408 + 224 - 184 = 163$ Fibonacci Win Points. (For convenience we will always round the calculation to the nearest integer.) So James concludes that Koufax was a better pitcher than Bunning. The conclusion itself is unimportant, as virtually every baseball analyst would concur with James. What's important is that there is now a numerical method for comparing records.

But why does James call this number Fibonacci Win Points? He noticed that some pitchers, such as Koufax, have more Fibonacci Win Points than wins, while other pitchers, such as Bunning, have fewer Fibonacci Win Points than wins. The division between the two sets of pitchers occurs when wins equals Fibonacci Win Points. This means

$$W \cdot \frac{W}{W+L} + W - L = W,$$

or

$$\frac{W^2}{W+L} - L = 0,$$

which we can further reduce to

$$W^2 - LW - L^2 = 0.$$

Using the quadratic formula to solve for W (and again we reject the negative root) gives

$$W = \frac{L + \sqrt{L^2 + 4L^2}}{2} = \frac{\sqrt{5}+1}{2} \cdot L.$$

So we see the golden ratio appear. James defines the golden ratio via the ratios of successive elements of the Fibonacci sequence, and thus he uses the term Fibonacci Win Points.

Baseball statistics never give wins as a multiple of losses. Instead they list the winning percentage for each pitcher, so we calculate the winning percentage that corresponds to $W = (\sqrt{5}\,) + 1)/2 \cdot L$:

$$\frac{W}{W+L} = \frac{\frac{\sqrt{5}+1}{2}}{\frac{\sqrt{5}+1}{2}+1} = \frac{\sqrt{5}+1}{\sqrt{5}+3}$$
$$= \frac{\left(3-\sqrt{5}\right)\left(\sqrt{5}+1\right)}{4} = \frac{\sqrt{5}-1}{2}.$$

A quick calculation reveals that, computed to three decimal places, $(\sqrt{5} - 1)/2 = 0.618$. This value is simply $x - 1$, where x is the golden ratio. It is also $1/x$:

$$\frac{1}{x} = \frac{2}{\sqrt{5}+1} = \frac{2\left(\sqrt{5}-1\right)}{5-1} = \frac{\sqrt{5}-1}{2}.$$

James calls $1/x$ Fibonacci's Number; a few mathematicians call $1/x$ the golden ratio.

It's quite possible for a pitcher to have a negative number of Fibonacci Win Points. For example, Milt Gaston had 97 wins and 164 losses for –31 Fibonacci Win Points. (This is probably the lowest score ever achieved by a Major League pitcher.) James guessed the zero point (the winning percentage corresponding to zero Fibonacci Win Points) would have something to do with Fibonacci's Number $1/x$. Inspired by the relationship $1 - 1/x = (1/x)^2$, he guessed that the zero point was this common value, which is 0.382, computed to three decimal places. But he soon discovered that the zero point is 0.414, computed to three decimal places.

Problem 2. Find the exact value of the zero point.

Problem 3. Suppose a pitcher has W wins and L losses for a winning percentage of $P = W/(W + L)$. Show that his Fibonacci Win Points are given by $(P^2 + 2P - 1)(W + L)$.

Although this isn't a practical way to calculate Fibonacci Win Points, it can be used for many theoretical calculations. For example, having Fibonacci Win Points equal to wins requires that

$$(P^2 + P - 1)(W + L) = W,$$

or

$$P^2 + 2P - 1 = P.$$

From this we see that $P^2 + P - 1 = 0$, and thus

$$P = \frac{\sqrt{5}-1}{2}.$$

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Dave Trautman *is a member of the Department of Mathematics and Computer Science at The Citadel in Charleston, South Carolina.*

Looking for answers?

The solutions to problems in Martin Gardner's article "The Magic of 3×3," which appeared in the January/February issue, can be found on page 60. There you will also find an update from the author, summarizing recent developments related to 3 × 3 magic squares.

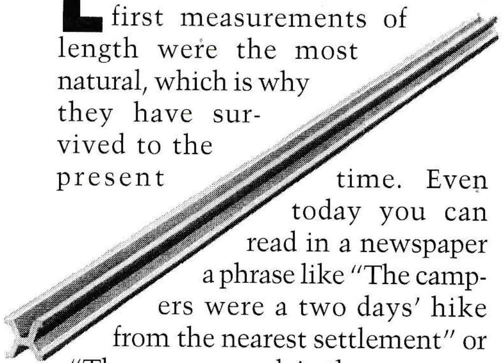
KALEIDOS

The long and

What do you really kno

by Anatoly S

LENGTH IS ONE OF THE first notions in geometry that mankind had to deal with. The first measurements of length were the most natural, which is why they have survived to the present time. Even today you can read in a newspaper a phrase like "The campers were a two days' hike from the nearest settlement" or "There was a crack in the pavement as wide as my hand."



But for all their convenience—we always had them "on us"—the primordial measures of length such as span (the distance between the tips of the thumb and little finger stretched out), cubit or ell (the distance from the fingertips to the elbow), and fathom (the length of two outstretched arms) were inexact: different persons have different units. So nation-states had to introduce standards of length—model units of measurement. Naturally, these units were different in different countries. For instance, three "Russian cubits" were equal to two "Persian cubits"; the latter unit received the name of *arshin* in Russia (*arsh* means "elbow" in the Turkic languages).

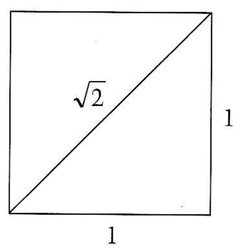
Even in the same country, relations between different units of length were sometimes quite intricate. For instance, Peter the Great issued an edict in the 17th century that was meant to put the Russian system of measurement in order. It introduced rather complicated relationships between the units used at that time:

1 mile = 7 versts = 3,500 sazhens (Russian fathoms) = 10,500 arshins = 168,000 vershoks (originally, a vershok was the width of the palm at the base of the fingers) = 294,000 inches = 2,940,000 lines = 29,400,000 points.[1]

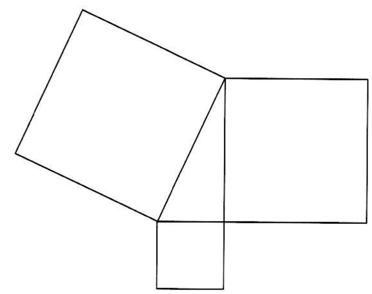
It's worth noting that in the last two relations the idea of the metric system can be glimpsed. But customary measures are so hard to root out that a revolution is needed to replace them—sometime literally. The French Revolution introduced the meter, kilometer, centimeter, and so on, to the French populace, and the October Revolution in Russia introduced these units there. The United States and Great Britain, on the other hand, still cling to their medieval systems of measurement, despite the efforts of some to convert them to the metric system.

Measuring length played a vitally important part in the history of mathematics. Indeed, what do we actually mean by the length of a segment? It is the number that shows how many times the chosen unit of length fits in the segment. If it doesn't fit an integer number of times, we must introduce a fractional length.

Already in ancient Greece people noticed that the diagonal of a square is incommensurable with its side—that is, can't be expressed as a rational multiple of the side. This observation resulted in the discovery of irrational numbers. Thus the notion of length was a bridge between geometry and algebra.



These two areas of mathematics were bound even more closely in the cornerstone philosophical treatise *Discourse on the Method of Rightly Conducting the Reason and Seeking for Truth in the Sciences* by the great French philosopher and mathematician René Descartes, whose 400th birthday is nigh upon us (March 31). In this work—or rather, in one of the three appendices to it—Descartes introduced coordinates, later called Cartesian coordinates, and thus laid the basis for analytic geometry. This made it possible to translate any geometric statement into algebraic language. For instance, the famous Pythagorean theorem—"the area of the square constructed on the hypotenuse of a right triangle equals the sum of the areas of the squares on its legs"—can be interpreted as the formula for the distance from a point in the coordinate plane to the origin: the square of this distance equals the sum of the squares of the point's coordinates. This theorem has many extensions in geometry, algebra, and number theory (Pythagorean triples, or the Great Fermat Theorem).



The attempt to measure the length of curves led to a variety of discoveries. The circumference of a circle was measured in ancient times (by polygonal approximation), although the question of the nature of the number π tortured mathematicians for hundreds of years and was answered only in the last century. The approach to defining the length of a curve was the same as for the circle. However, one has to be careful here, because sloppiness with limits may lead to absurdity. Draw a diagonal in a square and approximate it with "staircase" paths as shown in the figure at right. All these paths are the same length—twice the side length of the square. But on the other hand, geometrically, they approach the diagonal,

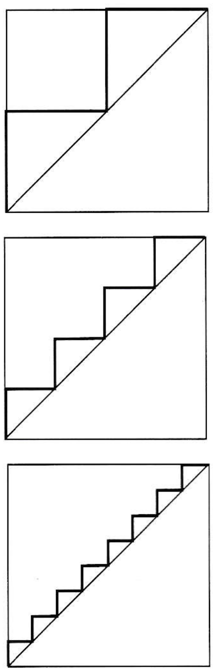
[1]To compare these measures to those you know, start with inches, which are the same magnitude everywhere.

OSCOPE

d short of it

now about "length"?

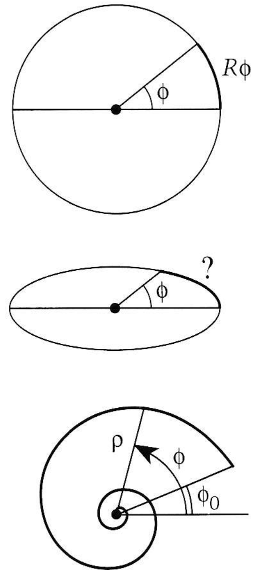
oly Savin



whose length is $\sqrt{2}$ times the side length. It looks like we've proved an absurd relation $2 = \sqrt{2}$! Haven't we? Problems of length measurement stimulated the development of the theory of limits.

It's curious that, whereas the length of a circular arc is proportional to the corresponding central angle ϕ, an arc of an ellipse (a "squeezed" or "flattened" circle) can't be expressed in terms of the corresponding angle even using the entire range of functions studied in high school (the "elementary functions"). Special *elliptic* functions had to be introduced to handle this problem. They proved useful for many other problems as well.

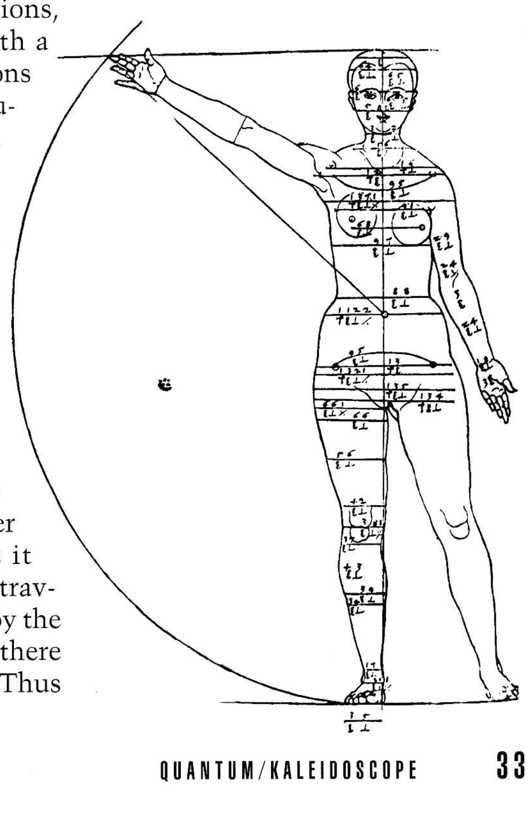
Another interesting class of curves is the spiral.[2] The so-called hyperbolic spiral, given by the equation $\rho = a/\phi$ (see the figure), winds about its "center" infinitely many times as the angle ϕ varies on the interval (ϕ_0, ∞), $\phi_0 > 0$. This spiral is infinitely long. The logarithmic spiral, given by $\rho = e^{-a\phi}$, also makes an infinite number of windings for $\phi > \phi_0$, but its total length is finite!



Speaking of length, I can't help mentioning the simplest and most important tool for measuring it: the ruler. It has markings in inches or centimeters. Two similar rulers can be used to add numbers. Put one ruler along the other with the zero mark on the first ruler against the 8-cm mark on the second. Find the 6-cm mark on the first ruler and read the number against it on the second ruler. It is 14, so $8 + 6 = 14$. The same idea was adjusted for multiplying numbers: we merely replace the uniform scale by the *logarithmic* scale. For convenience, one of two such rulers was made with a groove in which the other ruler could slide back and forth. Both pieces had many scales that allowed one to perform various operations, and a special glass runner with a cross hair helped match divisions on different scales. This calculating tool is called a slide rule. Not so long ago, before electronic calculators chased it from the field, it was an indispensable tool for every engineer.

When we speak of the length of the path from one place to another rather than the distance between them, kilometers may not be the best unit of measurement. It's more important for a passenger or a hiker to know the time it takes rather than the distance traveled, so we measure distances by the number of hours it takes to get there by plane, train, or automobile. Thus we introduce a completely different "metric" for points on Earth. In this metric, the distance, say, from Moscow to Uglich[3] may turn out to be greater than from Moscow to Marseilles. And a driver may have her own metric. What do they have in common? First, the distance from point A to point B is the same as from B to A; second, any distance is nonnegative; and third, the sum of the distances from A to B and from B to C is never smaller than the distance AC (the Triangle Inequality). They share one more property: if two points are zero distance apart, then they are the same point.

This brings us to one of the fundamental notions in modern mathematics—the notion of *metric space*. This notion includes not only ordinary space or the globe, but such uncommon "spaces" as, say, the set of all continuous functions on a segment, where the length seems to have completely lost any genetic connection with elbows, feet, fingers, and the like. This in itself is a kind of measure of how far we have come in our mathematical journey through the ages.



[2]See also the Kaleidoscope of the March/April 1995 issue of *Quantum*, devoted to these curves.

[3]A small town on Volga, several hundred kilometers away from Moscow, which played an important part in Russian history many centuries ago.—*Ed.*

PHYSICS CONTEST

Sea sounds

*"Come here Uncle John's band by the riverside,
Got some things to talk about, here beside the rising tide."
—Garcia–Hunter*

by Arthur Eisenkraft and Larry D. Kirkpatrick

REFRACTION IS RESPONSIBLE for distorting and correcting our vision. The shimmering of light as it passes over a hot barbecue requires a knowledge of refraction to be understood. Refraction of light at the surfaces of lenses benefits those of us who wear glasses or contact lenses.

Refraction was known as far back as the time of Ptolemy in the 2nd century A.D. It wasn't until 1621, however, that the correct relationship between the incident and refracted angles was discovered. Snell determined that

$$n_1 \sin \theta_1 = n_2 \sin \theta_2,$$

where θ_1 and θ_2 are the angles of the light measured relative to the normal to the surface in materials with indices of refraction n_1 and n_2, respectively. When Foucault measured the speed of light in water and other transparent materials in the 1850s, it was shown that the index of refraction is inversely proportional to the speed of light in the material—that is, $n \propto 1/v$.

Modern manufacturing techniques have allowed the development of a new type of lens that can be made out of a flat piece of glass. The graded index lens has an index of refraction that varies from the center to the edge. This variation causes the light to change direction as it would in a regular lens.

This technique also finds applications in fiber optics. The graded index fiber has a core with an index of refraction that decreases with distance from the center. This helps keep the rays traveling along the axis of the fiber.

Of course, Nature has exhibited the effects of a variable index of refraction for a very long time. The index of refraction of air varies with its density. Therefore, the index of refraction of the Earth's atmosphere decreases with altitude, and light rays bend toward the vertical as they approach the ground. As a consequence, the Moon appears about one diameter higher than it actually is when it is near the horizon. Temperature also affects the density of air, and this dependence creates the mirages we see on road surfaces as we drive down highways on hot days.

Identical effects occur with other types of wave—for instance, sound waves. This formed the basis for a theoretical problem given at the XXVI International Physics Olympiad held in Canberra, Australia, in July 1995.

The speed of propagation of sound in the ocean varies with depth, temperature, and salinity. Let's assume that the speed has a minimum v_0 midway between the surface and the ocean floor. For convenience, we choose the origin $z = 0$ at the level of the minimum, $z = z_s$ at the surface, and $z = -z_f$ at the floor. Let's also assume that the speed of sound increases linearly above and below $z = 0$ according to

$$v = v_0 + b|z|,$$

where b is a positive constant.

Let x be a horizontal direction, and let's place a source of sound S at the position $x = 0$, $z = 0$. A ray of sound is emitted from S at an angle $\theta_0 < \pi/2$ measured relative to the positive z-axis—that is, vertically upward (fig. 1).

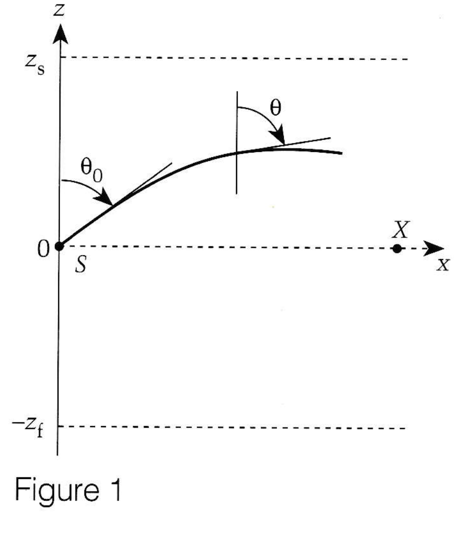


Figure 1

Art by Tomas Bunk


FISH SCHOOL PS1
REFRACTED LIGHT RAYS
TODAY: FISH & CHIPS
TUNA
TONIGHT:
the GRAND FISH BALL
REFRACTED SOUND RAY SPECIAL
TOMAS BUNK

A. Show that the trajectory of the ray (constrained to the xz-plane) is a circle with radius

$$R = \frac{v_0}{b \sin \theta_0}.$$

This can be derived by using calculus or demonstrated by means of a spreadsheet. In using the spreadsheet, assume that the ocean is divided up into a large number of horizontal sheets, each with a speed of sound equal to that in the mid-depth of the sheet. You can then apply Snell's law at the interfaces between sheets to obtain the trajectory of the ray. (It's sufficient to show that the path follows a circular arc at the beginning of the trajectory. The spreadsheet will have difficulties when the ray is horizontal. See physics challenge P170 on page 17.)

B. Derive an expression for the smallest angle of $\theta_0(z_s, b, v_0)$ that can be transmitted without the sound ray hitting the surface.

C. Assume that you have a microphone at a position $x = X$, $z = 0$. Find the series of values for $\theta_0(X, b, v_0)$ required for the sound ray emerging from S to reach the microphone. Assume that z_s and z_f are sufficiently large to remove the possibility of reflection from the ocean surface or floor.

D. Calculate the smallest four values of θ_0 for these rays given that X = 10,000 m, v_0 = 1,500 m/s, and $b = 0.02000\ \text{s}^{-1}$.

E. Let's now compare the times required for sound to travel along two different paths. The first path is the direct horizontal, or axial, path. The second path is the one corresponding to the smallest angle for θ_0. The time required for the second path can be obtained by integrating ds/v along the path—that is, by adding up the times that it takes to move small distances along the path. (The integral $\int dx/\sin x = \ln \tan(x/2)$ should prove useful.) This result can also be obtained numerically with a spreadsheet or a computer program. Which path takes the shorter time?

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

Split image

The split-lens problem that appeared in the September/October 1995 issue provided readers with a broken lens and asked for a description of the resulting light pattern. Because the problem noted the appearance of interference fringes, the first step in finding the solution is to determine how the multiple sources arise.

The bisection of the lens produces two point images of the point source of light, each displaced from the principal axis, which act as sources to produce the interference pattern. In figure 2, these sources are labeled S_1 and S_2. Our path to a solution is now revealed to us. We can use the lens equation to find the positions of the two new sources. A comparison of similar triangles can provide us with the distance between these sources. This will then allow us to find the spacing between the interference fringes following the analysis of a typical Young's double-slit experiment. A second set of similar triangles will provide us with the size of the overlap region from which we can find the number of fringes. Let's now journey down this solution trail.

We first use the lens equation to find the location of the images:

$$\frac{1}{f} = \frac{1}{d_0} + \frac{1}{d_i},$$

$$d_i = \frac{d_0 f}{d_0 - f}.$$

A comparison of similar triangles yields the distance between S_1 and S_2:

$$\frac{d_s}{\delta} = \frac{d_0 + d_i}{d_0},$$

$$d_s = \delta\left(\frac{d_0 + d_i}{d_0}\right) = \delta\left(\frac{d_0}{d_0 - f}\right).$$

The distance to the screen l can be found in terms of the given dimensions of the problem:

$$l = L - d_i = L - \frac{d_0 f}{d_0 - f}$$

$$= \frac{L(d_0 - f) - d_0 f}{d_0 - f}.$$

The distance x between interference fringes produced by two point sources of wavelength λ separated by a distance d_s on a screen l meters away is derived in most physics textbooks:

$$x = \frac{\lambda l}{d_s} = \frac{\lambda}{\delta d_0}\left[L(d_0 - f) - d_0 f\right].$$

CONTINUED ON PAGE 42

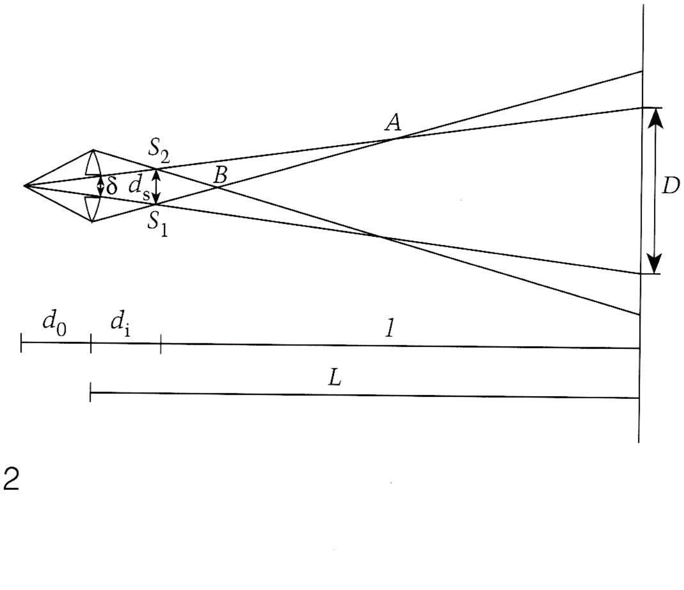


Figure 2

STICKING POINTS

Surprises of conversion

Flipping theorems on their heads

by I. Kushnir

IF YOU EXCHANGE THE HYpothesis and conclusion of a theorem, you come up with another statement that is the converse of the given one. For instance, the statement "if the square of one side of a triangle equals the sum of the squares of its two other sides, then the angle opposite the first side is a right angle" is the converse of the Pythagorean theorem. In this case, the converse theorem can be derived from the direct theorem. Indeed, if we have a triangle satisfying the Pythagorean relationship (but which may or may not have a right angle), a right triangle whose legs are equal to the shorter sides of the given triangle must be congruent to the original triangle. Thus we must have started with a right triangle initially.

But often the converse requires an independent proof, and sometimes the converse of a perfectly true theorem is in fact false. Anyway, it should be made very clear that two such theorems are different statements, each requiring a separate proof. Elementary geometry provides a host of examples that illustrate this simple truth. Below you'll find a number of geometric facts that you may find almost obvious. You'll surely have no problem proving them, but your assignment will be different: *formulate and prove their converse statements*. You'll see that all the converse statements are much harder to prove than their direct versions.

Direct theorems

1. The segment joining the midpoints of the bases of a trapezoid divides it into two figures of equal area.

2. In a parallelogram, the sum of the distances between the midpoints of the opposite sides is equal to the semiperimeter.

3. If H is the orthocenter (intersection point of the altitudes) of an acute triangle ABC, then $\angle HBA = \angle HCA$ and $\angle HAB = \angle HCB$.

4. If O is the circumcenter of triangle ABC, then $\angle OAB - \angle OBA = \angle OBC - \angle OCB = \angle OCA - \angle OAC$.

5. If a triangle is isosceles, then its incenter belongs to its Euler line. (The Euler line of a triangle passes through its circumcenter and the intersection points of its altitudes and medians. It can be shown that these three points are always collinear.)

6. The bisectors of the base angles of an isosceles triangle are the same length.

Hints for the converse theorems

1. Here the converse theorem states that if the line joining the midpoints M and N of the sides AB and CD in a quadrilateral $ABCD$ bisects its area, then these two sides are parallel.[1] To prove this, show that triangles AND and BNC (fig. 1) have equal areas. Since $DN = NC$, this will mean that $AB \parallel CD$.

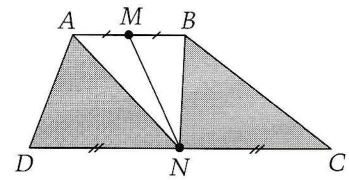


Figure 1

2. Let K, L, M, N be the midpoints of the sides AB, BC, CD, DA of a quadrilateral $ABCD$ (fig. 2). Given $KM + LN = (AB + BC + CD + DA)/2$, we are to prove that $ABCD$ is a parallelogram. Label E the midpoint of AC. Since KE and EM are midlines

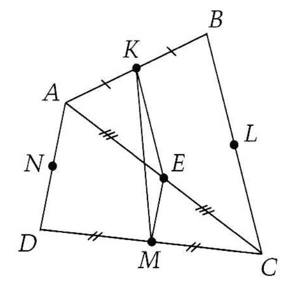


Figure 2

[1]So it's not strictly the converse of the original statement: we have to include the case of a parallelogram along with a trapezoid.

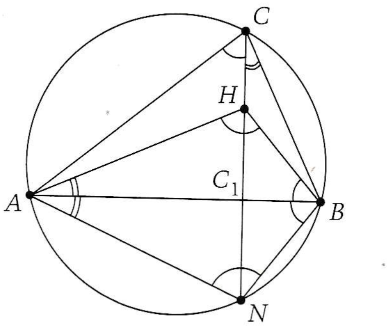


Figure 3

in triangles ABC and CDA, using the Triangle Inequality we get

$$KM \le KE + EM = \frac{BC + AD}{2}.$$

Similarly,

$$LN \le LE + EN = \frac{AB + CD}{2}.$$

Adding these inequalities and comparing the result with the given equation, we find that $KM = KE + EM$, which is possible only if E lies on the segment KM. But in this case, from $BC \parallel KE$ and $AD \parallel EM$, we derive $BC \parallel AD$. The other two sides are parallel for the same reasons.

3. We must prove that if the equalities of angles from the direct theorem are valid for a certain point H in an acute triangle, then this point is its orthocenter. Circumscribe the triangle and denote by C_1 and N, respectively, the points where the line CH meets AB and the circumcircle (fig. 3). From the first of the given equalities and the Inscribed Angle Theorem we get $\angle HBA = \angle HCA = \angle NBA$; similarly, $\angle HAB = \angle NAB$. It follows that H and N are symmetric about AB and, therefore, CC_1 is an altitude and $\angle AHB = \angle ANB = 180° - \angle ACB$. But there is no more than one point on the altitude CC_1 at which AB subtends a prescribed angle. The angle $180° - \angle ACB$, in particular, is subtended only at the orthocenter, which completes the proof.

4. Label the angles in the statement α_i, β_i, γ_i ($i = 1, 2$) as shown in figure 4. From $\alpha_1 - \alpha_2 = \beta_1 - \beta_2 = \gamma_1 - \gamma_2$, we must derive $OA = OB = OC$.

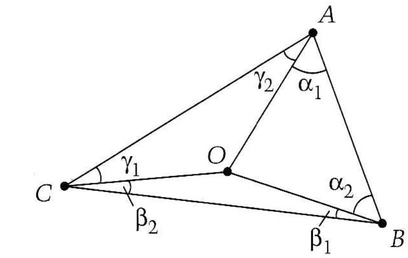


Figure 4

Suppose $OA \ne OB$—for definiteness, $OA < OB$. Then $\alpha_1 > \alpha_2$, so $\beta_1 > \beta_2$ and $OB < OC$. Similarly, $OC < OA$. But then $OA < OB < OC < OA$, which is a contradiction. So $OA = OB$, and, for the same reason, $OB = OC$. In the direct theorem, the differences are all in fact zero. Notice that for the converse theorem, it is enough to assume that all the differences $\alpha_1 - \alpha_2$, $\beta_1 - \beta_2$, $\gamma_1 - \gamma_2$ have the same sign.

5. Let O, H, I be the circumcenter, orthocenter, and incenter of triangle ABC, respectively (fig. 5). At least two of the triangle's bisectors—say, of angles A and B—do not coincide with the Euler line. If we extend these to meet the circumcircle at D and E, we find that D and E are the midpoints of the arcs BC and CA, so OD and OE are perpendicular to the sides BC and CA. Therefore, $OD \parallel HA$ and $OE \parallel HB$, and it follows that the triangles OID and OIE are similar to HIA and HIB, respectively. The ratio of similarity for both pairs of similar triangles is the same: HI/IO. Since $OD = OE$ (they are radii of the same circle), we know that $HA = HB$. Then $\angle HAB = \angle HBA$. An examination of triangles ABT, ABU now shows that $\angle CAB = \angle CBA$, so the

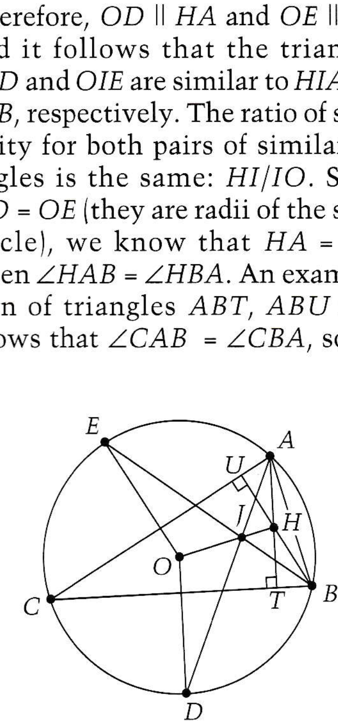


Figure 5

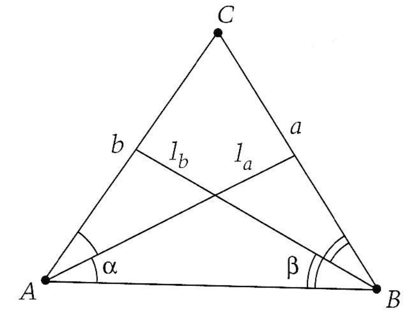


Figure 6

original triangle is isosceles.

6. Let $BC = a$, $AC = b$, $AB = c$. Then the converse theorem can be worded as follows: "If two bisectors of a triangle, l_a and l_b (fig. 6), are the same length, the triangle is isosceles ($a = b$)." Unlike similar statements about medians and altitudes, this is a really difficult theorem. It has been a matter of great interest for many geometers, who have offered many different proofs for it. The proof suggested below is by contradiction, as are many of the others.

Suppose $a > b$. Then $\alpha > \beta$. The lengths of our bisectors can be computed as

$$l_a = \frac{2bc}{b+c}\cos\frac{\alpha}{2}, \quad l_b = \frac{2ac}{a+c}\cos\frac{\beta}{2}.$$

To obtain, say, the first of these formulas, express the areas of the entire triangle and the parts into which it is cut by l_a in terms of b, c, l_a, and α; equate the entire area to the sum of its parts:

$$\frac{1}{2}bc\sin\alpha = \frac{1}{2}bl_a\sin\frac{\alpha}{2} + \frac{1}{2}cl_a\sin\frac{\alpha}{2};$$

—and cancel out $(\tfrac{1}{2}b \sin \alpha)/2$ using the fact that $\sin\alpha = 2\sin\alpha/2\cos\alpha/2$. Since $0 < \beta < \alpha < \pi$, $\cos\beta/2 > \cos\alpha/2$. In addition, $b/(b + c) < a/(a + c)$, because

$$\frac{a}{a+c} - \frac{b}{b+c} = \frac{c(a-b)}{(a+c)(b+c)} > 0.$$

This means that $l_a < l_b$, which is a contradiction.

If you liked this "conversion game," play it yourself with your favorite theorems in geometry.

A gripping story

Six cases of static friction, and how to slide past them

by Alexey Chernoutsan

HOW DO WE USUALLY SOLVE problems in dynamics? First we draw the forces and write down the equation for Newton's second law, projecting all the forces and accelerations onto the chosen axes. To solve these equations we also need to use the formulas for the laws that determine the forces acting on the object. For example, we substitute mg for the gravitational force (m is the object's mass, g is the acceleration due to gravity), kx for the elastic force (k is the elastic constant, x is the elastic displacement), and μN for the force of sliding friction (μ is the frictional coefficient, N is the normal force). When we draw the diagram, we apply the rules for determining the direction of each force: the force of gravity is always directed downward, the force of sliding friction is directed counter to the velocity of the object relative to the surface, and so on.

However, not all forces have their own laws. We can determine the normal force or the tension of a thread only because of the restrictions imposed by them on the motion of objects along a surface. For example, the normal force has just the right value so that the motion of the object is along the surface.

The force of static friction has the same qualities. The recipe for finding this force looks like this: the force of static friction has a magnitude and direction that maintains an object at rest relative to the surface on which it can move. Sometimes this force gives us headaches. We encounter our first difficulties when trying to depict this force in a diagram. We know only one thing about its direction: it's tangential to the surface. But in what direction? It's not always obvious. Also, in solving the problems we need to be sure that the resulting frictional force is within the range $0 \le F_{fr} \le \mu N$—otherwise the object begins to slide. And finally, the force of static friction sometimes appears in a strange garb (for example, as the tractive force of a train or car), so that it's difficult even to recognize it. Let's look at some examples.

1. A block at rest

Let's imagine that several forces act on the block under consideration, but it remains stationary. This means that the force of static friction has a magnitude and direction such that the sum total of all the forces is zero. So, what are these values?

In the simplest case (fig. 1) the answer is obvious: $\mathbf{F}_{fr} = -\mathbf{F}$.

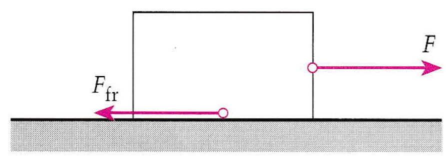


Figure 1

If the block lies on an inclined plane at an angle α, the force of friction is directed upward along the plane and is equal to $F_{fr} = mg \sin \alpha$ (where m is mass of the block). The block won't slip if $F_{fr} \le \mu N = \mu mg \cos \alpha$—that is, $\tan \alpha \le \mu$.

Now let's apply a small horizontal force to the block, directed along the plane (fig. 2), and then increase its value F. As this happens, the force of friction $\mathbf{F}_{fr}$ changes in both magnitude and direction. When the static friction force

$$F_{fr} = \sqrt{(mg \sin \alpha)^2 + F^2}$$

reaches the value $\mu N = \mu mg \cos \alpha$, the block begins to slide in the direction opposite to the direction of $\mathbf{F}_{fr}$ at this particular moment.

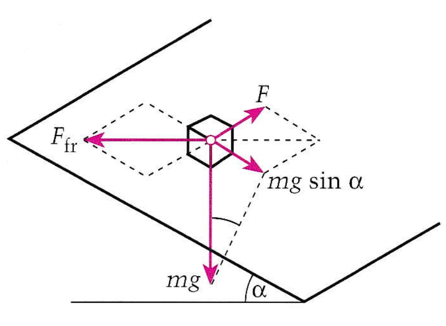


Figure 2

2. A block on a moving train

Let's imagine that a train moves horizontally with an acceleration **a** (fig. 3). In order for a block of mass m placed on its surface to move with the train, the force of static friction must impart the same acceleration **a** to it. So $\mathbf{F}_{fr}$ is directed forward and is equal to $F_{fr} = ma$. There will be no sliding if $F_{fr} \le \mu N = \mu mg$; on the other hand, when the acceleration of the train is greater than $a_0 = \mu g$, the block will slip backward relative to the train. Figure 3 also shows the force of friction $\mathbf{F}_{fr}'$ that the block exerts on the train: according to Newton's third law, $\mathbf{F}_{fr}' = -\mathbf{F}_{fr}$.

3. A block on a revolving platform

The acceleration of a block resting on a revolving platform must be directed toward the center of the platform. Since the force of friction is the only horizontal force that can impart this acceleration, it is directed toward the center and is equal to $m\omega^2 r$ (fig. 4a), where ω is the angular velocity of the platform. If the angular velocity is increased very slowly, then at the moment when the force of friction reaches the value $\mu N = \mu mg$, the block will begin to slip along the platform. If the angular acceleration is very high, then not only must the centripetal acceleration be taken into account (it's also called the radial acceleration), but also the acceleration that is directed along the velocity vector and results in a change in the speed (this *tangential acceleration* was neglected in the case above, where the angular acceleration was small). This means that the force of static friction that provides both these accelerations—or rather both components of acceleration (the resultant is a single vector, of course)—will be directed not exactly at the center, but at some angle to the radius (fig. 4b).

4. A wheel on an inclined plane

Let's imagine that a wheel is rolling down an inclined plane, but without any slippage between the

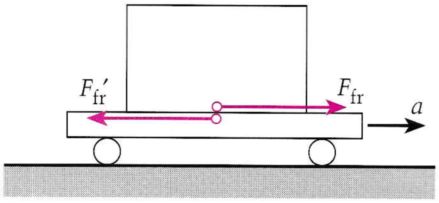


Figure 3

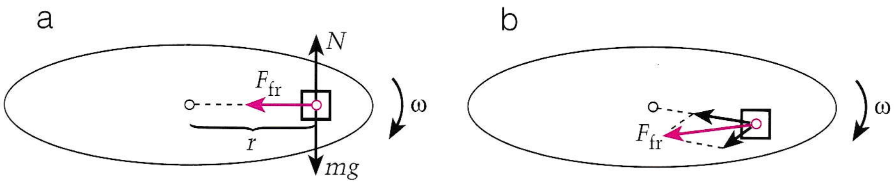


Figure 4

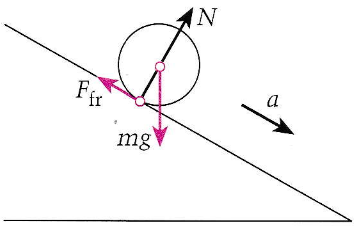


Figure 5

rim and the surface. This means that the points of the rim that are in contact with the surface are stationary at any given moment. As this occurs, the force of static friction equals the force that provides the rotational acceleration of the wheel (fig. 5). If this frictional force were absent, there would be sliding instead of rolling—the wheel would slip down the inclined plane without rotating.

5. The acceleration of a car

It should be noted that the tractive force of the engine that accelerates a car is nothing more than the force of static friction acting on the drive wheels. The engine transmits forces to the drive shaft via the transmission that try to turn the wheels clockwise (fig. 6). According to Newton's third law, a forward-directed force of static friction arises, which sets the car in motion.

And what about the passive (nondrive) wheels—are they affected by the force of static friction? Certainly, but to a far lesser extent, because this force must merely be enough to rotate the wheels, and that's all.

In addition to these forces, there is a resistive force that acts in the horizontal direction and consists of two parts: the force of rolling friction (resulting from deformation of the tires, and also from the roughness of the road) and air resistance.

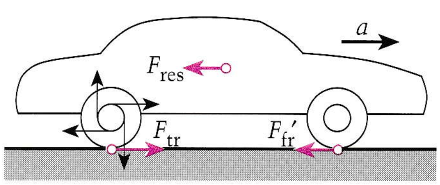


Figure 6

6. A car turning

Let's imagine a car in the middle of a turn, moving at a constant speed. Since the car's acceleration is directed toward the inside of the curve, it's perpendicular to the velocity vector. The force of static friction acting on the wheels, which roll without slipping, points in the same direction. Unfortunately, students often consider this force the force of sliding friction (the car is moving after all!) and orient it counter to the direction of the velocity vector. But then the question arises: where is the force that causes the centripetal acceleration?

It's interesting that besides the force of static friction, the car is affected by the resistive force directed counter to the velocity. Does this influence the force of static friction? In principle, yes. As the car moves at a constant speed, the resistive force must be compensated by an equal force in the forward direction—that is, an additional force of static friction oriented in the same direction as the velocity. This means that the force of static friction is directed at some angle to the radius (fig. 7): one of its components gives the car a centripetal acceleration, while the other compensates for the resistive force. On a bad road the resistive force may be considerable and can't be ignored. Indeed, skidding (and loss of control of the car!) will occur at the moment when the total force of static friction reaches the value $\mu N = \mu mg$. It's true that in theoretical problems we often ignore the resistive force. But can we do that in real life? Decidedly not!

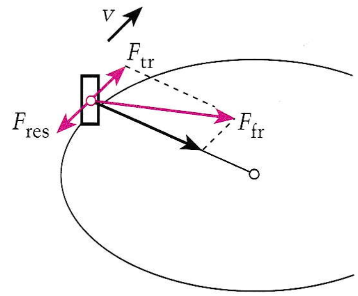


Figure 7

"SEA SOUNDS"
CONTINUED FROM PAGE 36

Interference fringes will appear only in the area where the light beams from S_1 and S_2 overlap. We can determine the size of this region by comparing similar triangles:

$$D = \frac{\delta(L + d_0)}{d_0}.$$

The number of fringes is found simply by dividing the length of the overlap region by the spacing of the fringes:

$$N = \frac{D}{x} = \frac{\delta^2}{\lambda}\frac{(L + d_0)}{L(d_0 - f) - d_0 f}.$$

With the data provided, we can calculate the number of fringes for the specific situation described:

$$f = 10 \text{ cm},\ d_0 = 20 \text{ cm},\ \delta = 0.1 \text{ cm},\ \lambda = 500 \text{ nm, and } L = 50 \text{ cm}$$

yield

$$N = 46 \text{ fringes}.$$

A different calculation must be made for D if the screen is nearer than point A in the figure. If the screen is placed nearer than point B, there will be no overlapping region and therefore no interference fringes.

The orbit of triangles

Dedicated to the memory of Leroy F. Meyers (1927–1995)

by George Berzsenyi

LAST MAY MY FORMER COLleague Brad Brock, who was at that time at the Center for Communications Research at Princeton, called my attention to an interesting posting by Kevin Brown on the Internet. Subsequently I contacted Kevin, an industrial mathematician at Boeing (in Seattle). He told me about a related posting of his and gave me permission to tell my readers about his ideas. Hence the topic for the present column. In what follows, I will briefly sketch Kevin's preliminary research on the topic, as it was developed by an "old" friend, Zachary Franco, who is presently on the faculty of nearby Butler University. In 1978–81 (while a student at Stuyvesant High School in New York) Zac used to send me wonderful solutions to the problems featured in the "Competition Corner" of the now defunct *Mathematics Student*.

To obtain the "orbit of triangles," start with an arbitrary $\triangle ABC$ with sides a, b, c such that $a + b + c = 1$. Position the triangle in the first quadrant of the coordinate plane so that side a is the segment from (0, 0) to (a, 0), and let A' be the point occupied by the vertex A. Rotate the triangle so that C is at the origin and b is along the x-axis; let B' be the point occupied by B, and let C' be defined similarly—that is, rotate the triangle so that A is at the origin, c is along the x-axis, and C' is the point occupied by C. Now "normalize" $\triangle A'B'C'$ to obtain a triangle similar to it whose perimeter is 1; repeat the above process with this new triangle, and then with the triangle obtained from it, and so on.

In his development, Zac restricted attention to the triangular region D of the ab-plane shown in figure 1, in which all points (a, b) represent a triangle with sides a, b and $1 - a - b$. In accordance with the findings of Kevin Brown, it seems that all of the points of D, with the exception of (1/3, 1/3), tend to one of the "attractors" marked P_1, P_2, P_3 in figure 1, whose coordinates are (α, β), (β, γ), (γ, α), respectively, where $\alpha = \sin 18° = (\sqrt{5} - 1)/4$, $\beta = 1/2 - \sin 18° = (3 - \sqrt{5})/4$, and $\gamma = 1/2$.

My first challenge to my readers is to **prove that all points on the boundary of D tend to one of the attractors.** My second challenge is to **recreate the "butterfly" in figure 2**, which represents the first iteration of the points of D, obtained by Zac via the software program Mathematica, using 1,250 equispaced points in the region. My third challenge is to **explore further iterations of the points of D, possibly starting with more densely packed points.** Finally, you may wish to **explore the "lines of points" approaching the attractors in figure 2** and **investigate the behavior of the points near (1/3, 1/3).** In a future column I will report on your findings.

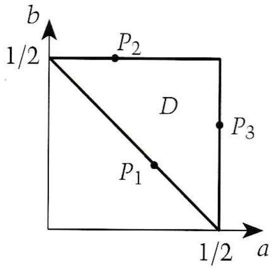


Figure 1

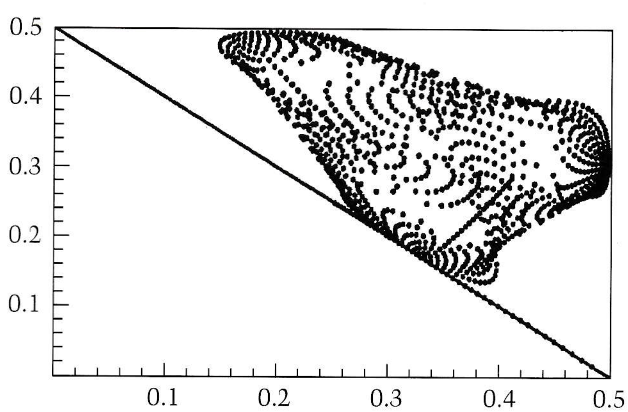


Figure 2

Feedback

As I reported at the time, my columns on triangle constructions in the July/August and September/October 1994 issues of *Quantum* were partially based on the findings of my friend Roy Meyers, who was an excellent geometer. My dedication is based on the belief that he would have enjoyed the present column as well. My last e-mail message from him was dated just a few days prior to his death; in it he shared with me some new findings about the construction of triangles. I will report on them in my next column.

Up the down incline

A double-cone roller that seems to defy gravity

by Alexander Mitrofanov

HERE IS A VERY OLD EXPERIment that looks like a trick. Take two identical cones made of wood, plastic, or metal—the material doesn't matter at all. The cones may be either solid or hollow, but they shouldn't be too light. Connect their bases firmly (use glue if necessary) and make sure the axes of the cones are in alignment. To do the experiment, you also need a thick book, as well as two identical and rather long rods (a pair of chopsticks, perhaps).

Put the book on the table and lean the rods on one edge of the book such that they form a V, with the point of the V resting on the table. Basically you've created an inclined plane in the shape of an equilateral triangle (fig. 1).

Now take your double-cone roller and put it on the rods so that its axis is horizontal. As you do, you'll see that the roller will move, all by itself and without any push, along the inclined plane—but not down, as you might expect! No, it moves upward, contrary to our everyday experience and common sense.

What's the secret of this trick? There's no magic here, of course. It turns out that under certain conditions the roller's center of gravity will drop and not rise as it moves up to the book. It's the force of gravity that causes this motion, and it looks pretty strange at first.

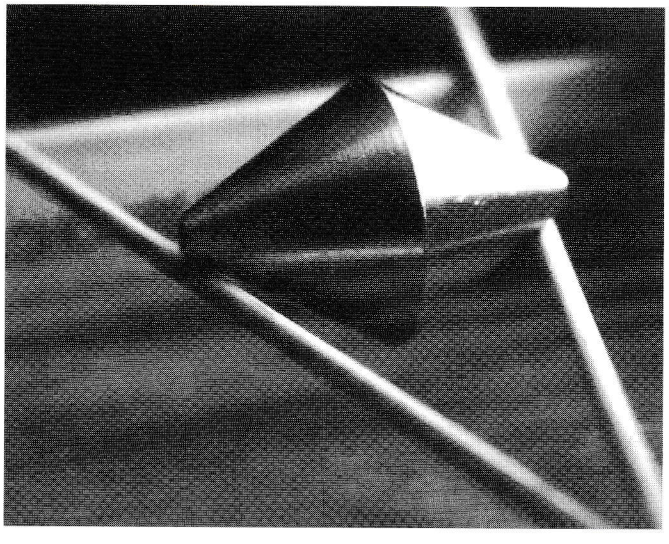


Figure 1

What conditions must be met in such an experiment? Let's look at the problem in more detail. We denote the angle of the plane with the horizontal as α, the angle between the rods as 2β, and the angle at the apex of each cone as 2γ (fig. 2). Let the roller move upward along the guide rods from position I to II, and let l_1 and l_2 be the corresponding distances between the points of tangency of the roller and the rods.

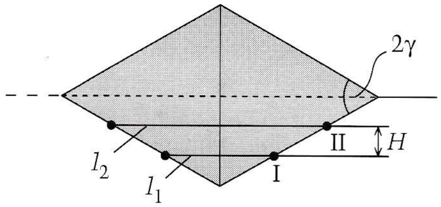


Figure 2

Figure 2 shows that relative to the points of tangency, the roller's center of gravity drops by a value of

$$H = \frac{l_1 - l_2}{2}\tan\gamma.$$

Naturally these points of tangency rise during the motion (fig. 3) by a value of

$$h = |MN|\sin\alpha = \frac{l_1 - l_2}{2}\cot\beta\sin\alpha.$$

Now we have everything we need

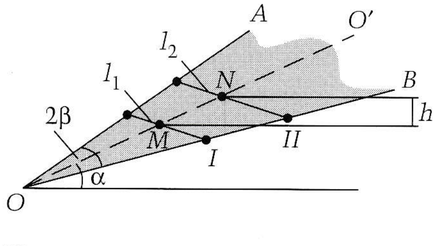


Figure 3

to formulate the condition for the roller to move up the inclined plane. Clearly this happens when $H > h$, or

$$\frac{l_2 - l_1}{2}\tan\gamma > \frac{l_2 - l_1}{2}\cot\beta\sin\alpha$$

—that is,

$$\sin\alpha < \tan\beta\tan\gamma.$$

Only in this case will the center of gravity drop as the roller moves up along the rods.

In the opposite case—that is, if

$$\sin\alpha > \tan\beta\tan\gamma$$

—the roller will "really" roll downward. When

$$\sin\alpha = \tan\beta\tan\gamma,$$

the roller will rest on the rods in a state of neutral equilibrium—that is, it will be at rest (just like a cylinder on a horizontal plane). The photo (fig. 1) illustrates just that case.

To conclude, I'll leave you with some experimental problems connected with this demonstration.

Problems

1. Check experimentally the formula $\sin\alpha = \tan\beta\tan\gamma$, which describes the condition of neutral equilibrium for the double-cone roller on an inclined plane. With what accuracy one can check this condition?

2. Make the calculations and prove experimentally that the condition for neutral equilibrium will still hold if you move the support point of the rods toward the book while keeping the points of tangency of the rods and the book's edge the same.

3. From the theoretical viewpoint one can choose the angles γ and β in such a way that the condition for the roller to rise is fulfilled even at $\alpha = 90°$ ($\sin\alpha = 1$)—that is, when the inclined plane is vertical. What experimental results will you get with such a setup?

4. Let the inclined plane be formed by rods that meet at the top rather than the bottom. What shape would an object need to have in order to roll up these guide rods? ◘

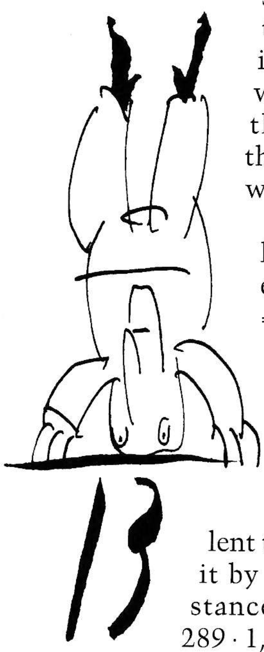
Number show

A handful of number tricks

by Ivan Depman and Naum Vilenkin

YOU CAN AMAZE YOUR friends with numerical conjuring tricks. Here's one of them.

Ask someone to write down a three-digit number. Let another friend continue this number by adding the same three digits to the right; have a third divide the six-digit number obtained by seven; ask a fourth to divide the quotient by eleven; finally, have friend number five divide the result by thirteen and pass it to the first, who will see the number that started the whole shebang.



The secret lies in the equality $1{,}001 = 7 \cdot 11 \cdot 13$. Writing two copies of a three-digit number side by side is equivalent to multiplying it by 1,001 (for instance, $289{,}289 = 289 \cdot 1{,}001$), and consecutive divisions by 7, 11, and 13 amount to one division by 1001, which restores the initial number.

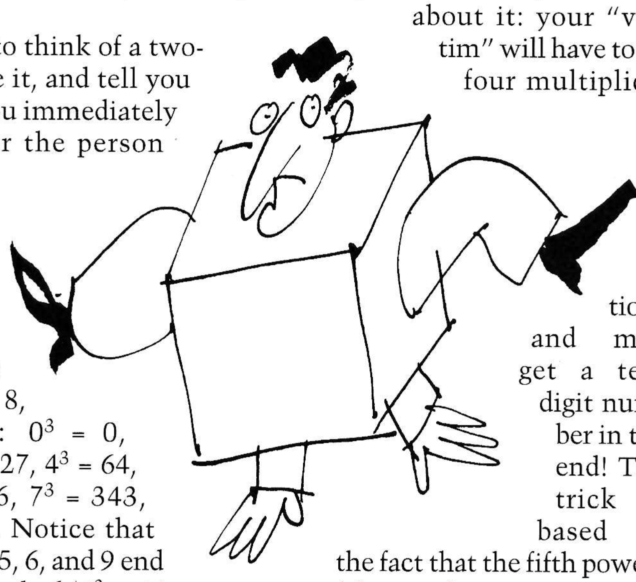
You can perform a similar trick with two-digit numbers. The number must be repeated three times (rather than two, as above) and the result is successively divided by 3, 7, 13, and 37. Here the underlying equality is $10{,}101 = 3 \cdot 7 \cdot 13 \cdot 37$. Four-digit numbers are repeated twice and divided by 73 and 137. The "secret" relation is $10{,}001 = 73 \cdot 137$.

Ask somebody to think of a two-digit number, cube it, and tell you the result. Then you immediately name the number the person thought of.

To perform this trick you'll only have to learn by heart the cubes of the one-digit numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Here they are: $0^3 = 0$, $1^3 = 1$, $2^3 = 8$, $3^3 = 27$, $4^3 = 64$, $5^3 = 125$, $6^3 = 216$, $7^3 = 343$, $8^3 = 512$, $9^3 = 729$. Notice that the cubes of 0, 1, 4, 5, 6, and 9 end in the digit being cubed ($4^3 = 64$, $9^3 = 729$), and that the numbers 2 and 8, 3 and 7 make pairs in which each number is the last digit of the other's cube.

Suppose you cube the number 67. You'll get the answer 300,763. Notice that 300 is between 216 and 343—that is, between 6^3 and 7^3. Therefore, the first digit (the tens' place) is 6. The last digit of the cube, 3, appears at the end of 7^3. So the second digit of the secret number is 7. Thus we guess the number thought of: 67. With a little practice, you'll be able to do the "guesswork" in no time flat.

Guessing a two-digit number from its fifth power is an even more impressive performance. Just think about it: your "victim" will have to do four multiplications and may get a ten-digit number in the end! The trick is based on the fact that the fifth powers of the numbers 0, 1, 2, ..., 9 all end in the digit that has been raised to the power. (For instance, $1^5 = 1$, $2^5 = 32$, $3^5 = 243$, $4^5 = 1{,}024$, and $5^5 = 3{,}125$). In addition, the conjurer has to memorize the following table showing the beginnings of the fifth powers of the multiples of ten:

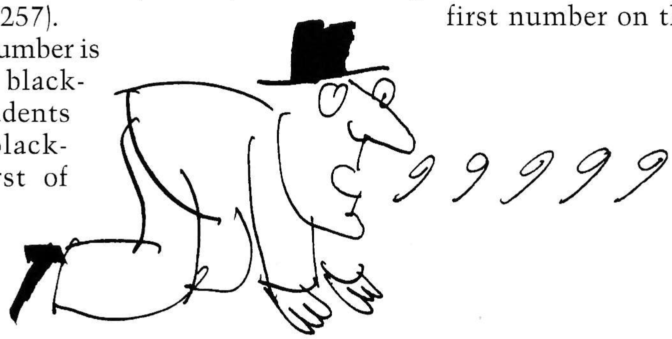


10^5 = 100 thousand
20^5 = 3 million

$30^5 = 24$ million
$40^5 = 100$ million
$50^5 = 300$ million
$60^5 = 777$ million
$70^5 = 1$ billion 500 million
$80^5 = 3$ billion
$90^5 = 6$ billion
$100^5 = 10$ billion

Having been told that the fifth power of a certain two-digit number is, say, 8 billion something, you grasp at once that this result is between 6 billion and 10 billion, so the tens' place of the unknown number is 9. And when your partner tells you that the last digit of the power, 7, you immediately give the answer: 97 (indeed, $97^3 = 8{,}587{,}340{,}257$).

A five-digit number is written on the blackboard. Two students come to the blackboard. The first of them writes an arbitrary five-digit number, the second writes another number in reply. Then they exchange five-digit numbers twice more. After that, the second student immediately writes out the sum of all seven numbers on the blackboard.



The secret of this trick is that every time the first student writes a number, the second writes a number such that the digits in each decimal place of the two numbers add up to nine. (If the first number written down was 40,817, the reply will be 59,182). The sum of two such numbers is always 99,999. After three number exchanges, the sum of the six numbers written by the students will be $3 \cdot 99{,}999 = 300{,}000 - 3$. So to write out the sum of the seven numbers, you must write the digit 3 in front of the very first number on the blackboard and subtract 3 from the six-digit number thus obtained.

To make the trick less obvious, you can decrease the first digit of one of your reply numbers by several units and accordingly decrease the corresponding digit in the total. For instance, the numbers on the blackboard can be

76,281
14,391
65,608
24,380
75,619
95,073
4,926

for a total of 356,278. Here the first digit of the third summand is decreased by 2 and the same is done with the corresponding (second) digit in the sum.

AT THE BLACKBOARD II

So what's the point?

Or rather, where is the point?

by Gary Haardeng-Pedersen

WHEN A RIGID BODY IS IN equilibrium, the net force on that body must be zero, and the net torque, calculated about any point, must also be zero.

In many cases, to provide a simple model that fits nicely into a framework of rectangular components, a single force is replaced by a pair of convenient component forces. For example, a force from a hinge may be taken as a pair of component forces: a horizontal force and a vertical force acting at the same point. A force from a rough surface may be taken as a pair of forces: a component force perpendicular to the surface (the normal force) and another component force along the surface (the frictional force).

For readers of *Quantum*, it is appropriate to consider some special cases where dealing with the vectors directly (rather than their components) will simplify the arithmetic—replacing algebraic solutions with geometric ones—and provide a better understanding of the physics involved.

In the particular cases of concern, there are exactly three forces acting on the rigid body, and these forces are not parallel. If the forces were parallel, the determination of the magnitudes of the forces would be done by simple application of the "law of levers."

As a review of the standard solution of a standard torque problem, consider—once again—the case of a uniform ladder (of mass M and length L) with its foot on a rough horizontal floor and its top against a smooth vertical wall (fig. 1) What is the minimum coefficient of static friction required to keep the ladder stationary when it's at an angle of θ with the vertical?

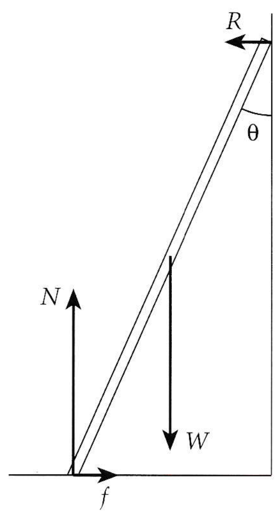


Figure 1

One of the forces acting on the ladder is its weight $\mathbf{W} = M\mathbf{g}$, acting vertically downward from the midpoint of the ladder (since the ladder is uniform). A second force on the ladder is the normal force $\mathbf{R}$ from the wall. Since the wall is smooth, this force must be directed perpendicular to the vertical wall—that is, $\mathbf{R}$ is horizontal. The force from the rough floor is usually resolved into two components: the component perpendicular to the floor (the normal force $\mathbf{N}$) and the component parallel to the floor (the frictional force $\mathbf{f}$). In terms of the total force from the floor $\mathbf{F}$, which is at an angle of ϕ with the vertical, the component forces are

$$N = F\cos\phi$$

and

$$f = F\sin\phi.$$

Balancing vertical forces gives $N = Mg$, and balancing horizontal forces gives $f = R$. Taking torques about the foot of the ladder leads to the third condition needed:

$$RL\cos\theta = Mg\frac{L}{2},$$

or

$$R = \frac{1}{2}Mg\tan\theta.$$

The minimum required coefficient of static friction is then

$$\mu = \frac{f}{N} = \frac{R}{Mg} = \frac{1}{2}\tan\theta.$$

The alternative approach is to refrain from resolving the force from the floor into its components and to calculate the net torque (which is zero about any point) about a different position. Let's choose to calculate the net torque about the position P indicated in figure 2. This point is chosen to lie vertically above the midpoint of the ladder and

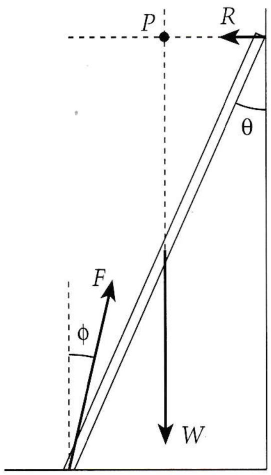


Figure 2

at the same vertical level as the contact point between the ladder and the wall.

The condition necessary for rotational equilibrium is that the sum of the torques about any point must equal zero. Denoting these torques as τ_W (the torque due to the weight vector), τ_R (the torque due to the force at the top of the ladder), and τ_F (the torque due to the force at the bottom of the ladder), then

$$\tau_W + \tau_R + \tau_F = 0.$$

Since the point P about which we have chosen to calculate the net torque is vertically above the midpoint of the ladder and the weight acts vertically through the midpoint of the ladder, P lies on the line of action of **W**. Consequently, $\tau_W = 0$. The same argument applied to τ_R shows that, because of the choice made in the location of P, this torque is also zero. The apparently complicated condition for rotational equilibrium then reduces to $\tau_F = 0$.

The interpretation of this condition is that the line of action of **F** must pass through the point P. And this means that

$$\tan\phi = \frac{\frac{1}{2}L\sin\theta}{L\cos\theta} = \frac{\tan\theta}{2}.$$

Since the direction of **F** is now known, a scale drawing could be used to determine the magnitudes of each of the forces.

Using a sketch (fig. 3) instead of a scale drawing and the fact that the vector sum of the three forces is zero allows us to draw a vector triangle. Vector **W** is a vertical vector of known magnitude—the weight of the ladder. From the end of this vector, the second vector **R** can be drawn horizontally; the vector starts at the end of **W**, but unfortunately the length of the vector is unknown. This means that we don't know where to start the vector **F**. But we do know the direction of **F** and know that it ends at the point where **W** started, since the sum of the three vectors is zero!

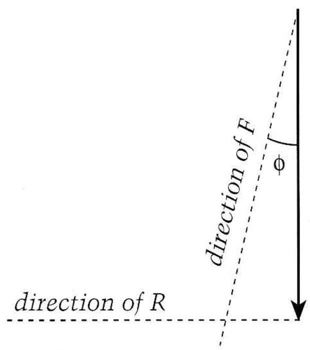


Figure 3

An added bit of information is that the angle ϕ between the force vector and the normal to the surface can be expressed as

$$\tan\phi = \frac{f}{N} \le \mu.$$

Consider a slightly modified ladder problem (fig. 4) A uniform ladder of mass M and length L has its foot on a rough horizontal floor and leans against a smooth sloping wall. The ladder, on the verge of slipping, is at an angle of $\alpha = 36.9°$ with the horizontal and the wall is at an angle of $2\alpha = 73.8°$ with the horizontal. What is the coefficient of static friction between the ladder and the floor?

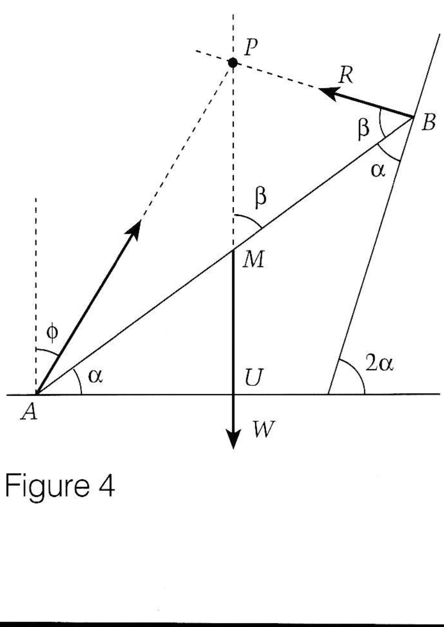


Figure 4

Label the foot of the ladder A, the top of the ladder B, and its midpoint M. The weight vector **W** acts vertically down from M, and the normal force **R** from the smooth slope acts at B, perpendicular to the incline. Once again, choose a point P about which to calculate the torques, so that it lies on the intersection of the lines of action of **W** and **R**. The argument that the net torque about P must be zero together with the facts that τ_W and τ_R are zero again leads to the conclusion that the line of action of the force acting on the foot of the ladder **F** passes through P. Once again a scale drawing would determine the angle ϕ between **F** and the vertical to be approximately 29°, and a vector triangle can be drawn to determine the magnitudes of **R** and **F**.

Alternatively, geometrical arguments show that the triangle PMB (see figure 4) is isosceles, where the angle $\beta = \pi/2 - \alpha$ and $PM = PB$.

Denote by U the point on the horizontal surface directly below the midpoint of the ladder. Then

$$\tan\phi = \frac{AU}{PU},$$

where

$$AU = \frac{L}{2}\cos\alpha$$

and

$$PU = PM + MU = \frac{L/4}{\cos\beta} + \frac{L}{2}\sin\alpha.$$

Then

$$\mu = \tan\phi = \frac{2\cos\alpha}{\csc\alpha + 2\sin\alpha} = \frac{2\cos\alpha\sin\alpha}{1 + 2\sin^2\alpha}.$$

For the value of α specified, $\mu = 24/43 = 0.558$ and $\phi = 29.2°$.

Another example of the use of the intersection point of the lines of action of the three forces is illustrated

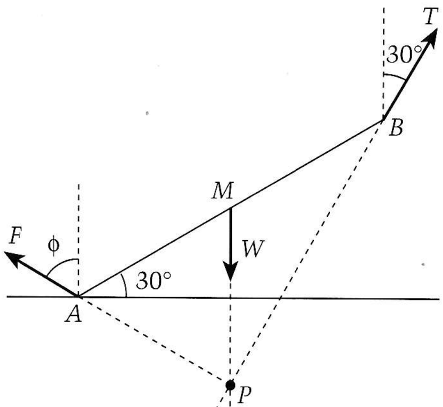


Figure 5

in figure 5. A uniform board is at rest with its left end on a rough horizontal floor. It is maintained at an angle of 30° with the horizontal by a rope attached to its right end. This rope pulls upward and to the right at an angle of 30° with the vertical. If the board is just about to slip, determine the coefficient of static friction between the board and the floor.

The line of action of the force **T** acting on the upper end of the board is in the direction of the rope, and the line of action of the weight **W** is vertical, passing through the midpoint of the board. The two lines of action intersect at a point *P* below the horizontal floor. The line of action of the force **F** exerted on the board by the floor must pass through *P*. Labeling the midpoint of the board *M*, its lower end *A*, and its upper end *B*, we find that triangle *PMB* is isosceles (the angle *BMP* = 120° and the angle *MBP* = 30°), so that *MB* = *MP*. The triangle *AMP* is then equilateral, and the line of action of **F** is therefore at an angle of 60° with the vertical. Thus $\mu = \tan 60° = \sqrt{3}$.

Finally, consider another variation on the ladder problem. The situation is identical to that posed in the first problem (illustrated in figure 1) except that the wall is rough, with the same coefficient of static friction as the floor. What is the value of this coefficient of static friction if the ladder is just about to slip?

If the coefficient of static friction is μ and the ladder is just about to slip, the force acting on the foot of the ladder is at an angle of ϕ with the vertical,

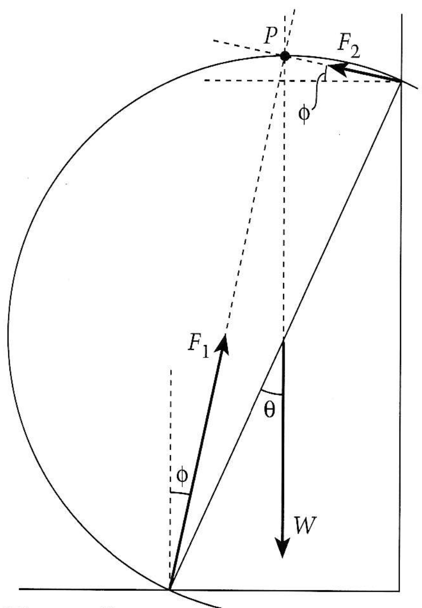


Figure 6

and the force acting on the top of the ladder is at an angle of ϕ above the horizontal, where $\phi = \tan^{-1} \mu$. The lines of action of these two forces are then perpendicular to each other as shown in figure 6. Choosing *P* again as the intersection point of the three lines of action, we find that *P* must be vertically above the midpoint of the ladder and must also lie on a circle that has the ladder as its diameter. Consequently, $\phi = \frac{1}{2}\theta$, or $\mu = \tan \theta/2$.

For a person who prefers to attack problems from a geometrical or engineering drafting perspective rather than an algebraic approach, the sketch of the rigid body with the location of the point where the three lines of action must intersect can be a powerful technique. Of course, it doesn't work when there are more than three forces acting on the object—for example, a ladder with someone standing on a rung. Well, actually, by calculating the position of the center of gravity of the person and the ladder and then using a single weight vector through the center of gravity, you can reduce the problem to a three-force problem, and then this approach will work.

Gary Haardeng-Pedersen *is an associate professor of physics at Sir Wilfred Grenfell College (a campus of Memorial University of Newfoundland), located in Corner Brook, Newfoundland, Canada.*

Bulletin Board

Let's go mathcamping

The fourth annual Canada/USA Mathcamps will be held in 1996 at the University of Washington, Seattle (July 29–August 23).

Canada/USA Mathcamps are the national summer mathematics camps for high school students. They are four-week programs to develop the mathematical ability and psychological well-being of mathematically talented students. Selection of students is by a two-tiered qualifying math quiz and by a teacher's recommendation. Trainers in the camp are nationally and internationally known expert mathematics communicators who are also research mathematicians.

Two programs will run simultaneously but separately in Seattle: an entry-level camp called Mathcamp I and an advanced program called Mathcamp II.

The camp exists for three reasons. (1) Students in a regular math class in school possess widely varying abilities in mathematics. While the teacher's instruction benefits the class taken as a whole, the talented few who could go deeper and further are not adequately challenged by the content and pace. (2) It is a widely held view that prospective mathematicians should begin at an early age. Mathematical vocation is most often awakened at about fifteen years of age. (3) The gifted have social and emotional needs that cannot be adequately met in a regular school setting. The greatest social need of the gifted is for true peers—those who are similar in ability and interests.

The mathematics training camp goes beyond problem solving. It is a threefold mathematical re-education complementing the high school syllabus: (a) acquiring a confident familiarity with those essential concepts that are not in today's high school curriculum; (b) proper writing in the language of mathematics; and (c) doing mathematical proofs employing the techniques that mathematicians use.

There are other benefits from attending this camp. Soon after the end of the camp, an opinion letter evaluating the student is filed by Mathcamps in its archives. The letter describes the student's raw intelligence, creativity, work habits, and social behavior and is sent, at the student's request, to universities and other entities that require information about the student from sources outside the school setting. Mathcamps has an accuracy control mechanism ensuring that the letters are factual portrayals of the students.

The camp fee, which includes tuition, meals, and dormitory accommodation, is US$1,683. The airfare from anywhere in Canada or the US is $195. There is also a Mathcamp I at Columbia University in New York City (June 24–July 19). It is nonresidential and charges a fee of US$795, which is the tuition part of the fee charged by the residential camp.

These camps are run by the Mathematics Foundation of America (MFOA), a nonprofit organization founded in 1995 to run the Canada/USA Mathcamps. Mathcamps began in Vancouver, British Columbia, in 1993. Five camps have been held since then. The mission of MFOA is to bring together mathematically gifted high school students from Canada and the US at no cost to the students. It is expected that this goal will be realized in less than five years.

More information on Mathcamps is available on the World Wide Web: http://www.mathcamp.org. The organizers can be contacted by e-mail at info@mathcamp.org, or by phone at 519 672-7990.

East meets West: RedShift 2

If you're a stargazer who also happens to like scanning the shelves at your favorite software store for interesting new titles, you've undoubtedly seen RedShift 2. It's a CD-ROM planetarium for your desktop computer—an enhanced version of the award-winning RedShift program introduced in 1993. But did you know that many of the developers who created RedShift 2 are veterans of the Russian space program?

Take Evgeny Kireev. The 26-year-old programmer was educated at the Moscow Aviation Institute in the department of Cosmonautics. He began his career at Russian Mission Control Center in the interplanetary spacecraft orbit determination section, where he specialized in spacecraft design and celestial mechanics.

In 1992 Kireev joined Maris Multimedia, the publishers of RedShift 2, which has offices in Moscow. He designed a celestial mechanics architecture that can simulate planetary positions with an accuracy of about two arc-seconds over a 200-year period. "You can visualize this magnitude of error," Kireev explains, "by imagining that you are looking at a planet from a point on the Earth, then walk about 500 feet and look again. The difference in position is about two arc-seconds."

In order to achieve this level of accuracy, the most authoritative astronomical data were obtained from space science centers throughout the world, including NASA and Russian Space Mission Control. For example, in order to calculate the historic positions of the planets back to 3000 B.C., the planetary position data file called DE-102 was obtained from NASA's Jet Propulsion Laboratory.

When these data were run through the Russian Space Mission Control model, the RedShift development team discovered a number of discrepancies. This led to an active dialogue between Russian Space Mission Control and the Jet Propulsion Laboratory, which resulted in a revision of the DE-102 data. So here we have a curious case of virtual reality affecting the real world: the development of RedShift 2 advanced the accuracy of data used by space scientists, and also fostered closer cooperation between once distant Russian and American colleagues.

Another Russian on the RedShift team is Dr. Yuri Kolyuka, a dean of the Soviet space program who continues to work as a scientist at Russian Space Mission Control Center. Kolyuka is in fact the Space Products team leader at Maris Multimedia, and he speaks with pride of the powerful engine in RedShift 2 for calculating both regular and irregular orbits. "Any deviation, such as the nonregular distribution of matter mass inside planets, the gravitational attraction from distant celestial bodies, atmospheric drag, and other forces acting on the space object can now be taken into account, in order to derive accurate positions of comets, asteroids, or any spacecraft—either orbiting the Earth or voyaging between planets. This gives users an incredibly realistic experience of traveling through the solar system and observing the universe from any moving object."

A test run by *Quantum* staff confirmed the power of RedShift 2, which also contains the *Penguin Dictionary of Astronomy* and numerous images and animations. While nonspecialists may feel intimidated by the computational abilities of the program—the control panel is breathtaking in its multitude of buttons and switches—the rank amateur will find easy inress into this huge topic. For example, within minutes you should be able to view the sky at the moment of your birth (in another city, perhaps) and compare it with the view tonight (at your current location).

While aimed primarily at the "backyard astronomer," RedShift 2 may find its way into the CD-ROM drives of professionals. "Because of the increased scientific sophistication of the program," Kolyuka says, "RedShift 2 will become useful also for professional researchers' work, as a means for visualization of experimental data or re-creation of fine astronomical events."

RedShift 2 is distributed by Maxis, Inc., and carries an estimated street price of $54.95. For more information, see the Maris home page at http://www.maris.com/maris.

"If today is Tuesday . . ."

". . . this must *not* be the answer to the *Quantum* CyberTeaser." More people than ever sent in answers to the March/April problem posted at the *Quantum* home page (brainteaser B166 in this issue), so it may be time for us to consider increasing the number of winners. First we need to count how many buttons we have left . . .

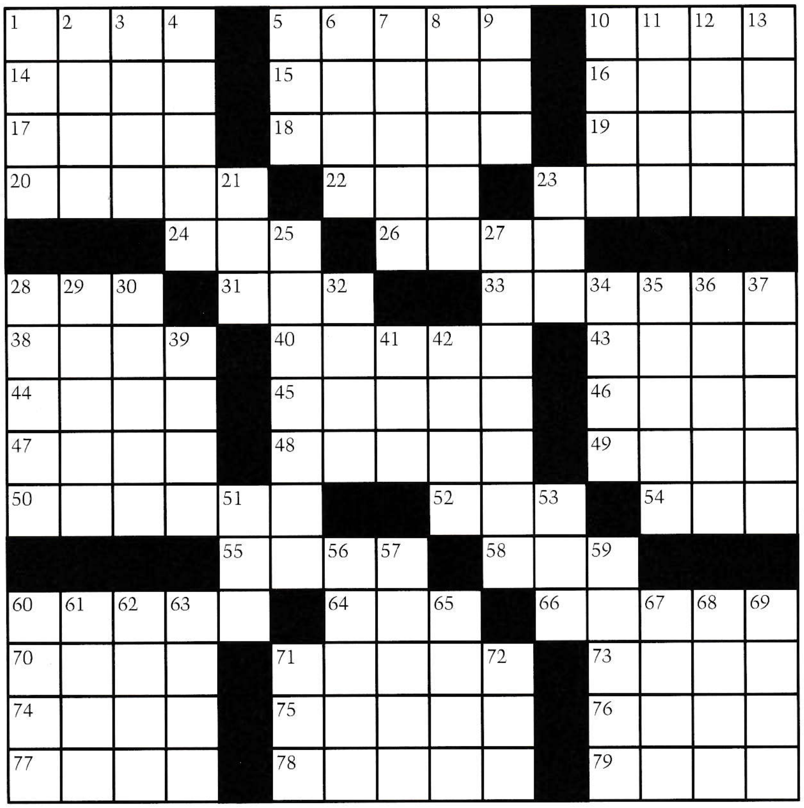
These were the first ten visitors to submit correct answers electronically:

Matthew Padilla (Lombard, Illinois)
Ed Sullivan (Herndon, Virginia)
Oleg Shpyrko (Rochester, New York)
Helio Waldman (Campinas, Brazil)
Matthew Wong (Edmonton, Alberta)
Richard Forsyth (Moorpark, California)
Julia Salzman (Pittsburgh, Pennsylvania)
Anne Marcks (Hingham, Massachusetts)
Elisa Keefe (Lawrenceville, Georgia)
Bruno Remillard (Cap-Rouge, Quebec)

Thanks to all who responded, and a special thanks to those who drop a few lines into our Guest Book. Some of your comments make us blush: "The arrival of each issue of *Quantum* is like having dessert at a fine restaurant" (from Houston, Texas). Some make us smile: "Delta(Quantum)delta(Scientific American) > Everything I Ever wish to know!" (from Stockholm, Sweden). And some pose a challenge: "Expand your magazine! I really enjoy it" (from Virginia Beach, Virginia). We read every message (though we can't reply personally to each one). So keep on writing—it's one of our most important "reality checks." Our e-mail address is quantum@nsta.org. Our home page is at http://www.nsta.org/quantum.

Criss × cross science

by David R. Martin



Across

1 Sense organs
5 ___ semiconductor
10 1949 Physiol. Nobelist ___ Moniz
14 Heavy EM field
15 Jet propulsion pioneer
16 Trig. function
17 German physicist Ernst ___ (1840–1905)
18 Vacant electron states
19 1948 Chem. Nobelist ___ Tiselius
20 1950 Chem. Nobelist
22 Curl
23 Swiss mathematician
24 Display light: abbr.
26 Unit of power
28 Greek letter
31 Trig. function
33 ___'s comet
38 Pediatrician Luther ___ (1855–1924)
40 Child: comb. form
43 44,714 (in base 16)
44 Sense organs
45 Introduction
46 Knot in wood
47 ___ process (for making German butter)
48 Shortest paths between points
49 Stared at
50 British archaeologist (1853–1942)
52 ___ product
54 Intrinsic FORTRAN function
55 Metal sources
58 Numbers: abbr.
60 Refrigerant
64 Basic logic function
66 Points of minimum disturbance
70 HIV disease
71 Gaucho's weapon
73 Speech: comb. form
74 Vaporize
75 Graded series of physiol. traits
76 1963 Physiol. Nobelist ___ Hodgkin
77 10^{-18}: pref.
78 Hebrew letter
79 ___ Descartes

Down

1 Mild oath
2 Phosphatase standardization unit
3 Outer garment
4 ___'s law (of refraction)
5 Japanese drama
6 ___ Heyerdahl
7 1977 Medicine Nobelist
8 Mary and Jesus
9 Printer's measures
10 Botanist Katherine ___
11 Female child
12 "The Diary of a Young Girl" author ___ Frank
13 Fortuneteller
21 Unit of time: abbr.
23 Greek letter
25 ___ radar
27 Joule–___ effect
28 Low cost
29 British astronomer
30 "___ my heart in San Francisco"
32 Hindu garment
34 Pigment
35 Actress Lotte ___ (1898–1981)
36 961,259 (in base 16)
37 Units of length
39 Russian ruler
41 Eternity
42 Act
51 Charged particle
53 Unit of weight
56 "___ Gay"
57 State of matter
59 Type of cell or wind
60 64,186 (in base 16)
61 Uproar
62 Revise and correct
63 European capital
65 Hyperbolic function
67 1936 Physiol. Nobelist
68 ___ vital
69 Unit of loudness
71 Superconductivity theory: abbr.
72 Understand

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M166

Denote the angular measures of the "inner" arcs AB of the given circles by 2α and 2β (for the circles ABM and ABP, respectively—see figure 1). Let us express MP in terms of α, β, and $l = AB$. Angle AMB is subtended by the arc 2α; angle APB either is subtended by the arc 2β (fig. 1) or is adjacent to an angle subtended by the conjugate ("outer") arc of measure $360° - 2\beta$ (fig. 2). In either case, by the Inscribed Angle Theorem, $\angle AMB = \angle AMP = \alpha$, $\angle APM = \beta$. Also, angle MAB, as the angle between the tangent AM to the circle ABP and its chord AB, is half the measure of the arc it intercepts: $\angle MAB = \beta$. Now, applying the Law of Sines to triangles ABM and AMP, we get

$$AM = \frac{AB\sin\angle ABM}{\sin\angle AMB} = \frac{l\sin(\alpha+\beta)}{\sin\alpha},$$

$$MP = \frac{AM\sin\angle MAP}{\sin\angle APM} = \frac{AM\sin(\alpha+\beta)}{\sin\beta} = \frac{l\sin^2(\alpha+\beta)}{\sin\alpha\sin\beta}$$

($\angle ABM = \angle MAP = 180° - \alpha - \beta$).

The length NQ can be found in the same way except that the symbols α and β are interchanged. But

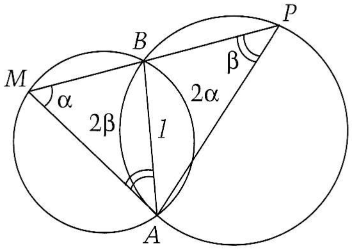


Figure 1

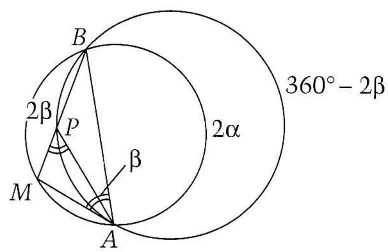


Figure 2

this permutation doesn't affect the result. So $NQ = MP$.

Our argument shows that any triangle AMP, where M and P are the points at which an arbitrary line through B meets the given circles, has angles α, β, and $180° - \alpha - \beta$. So all these triangles—in particular, AMP and AQN in the problem—are similar. This observation suggests another solution: it suffices to prove that, say, $AM = AQ$. This can be done by showing that these chords subtend equal or supplementary inscribed angles. This solution is left to the reader. (V. Dubrovsky)

M167

The answer is yes. An example can be constructed by using the method of small pertubations (see "Nudging Our Way to a Proof" in the March/April 1995 issue of *Quantum*).

Any power $P^n(x)$ of a polynomial $P(x)$ can be obtained as a product of several squares and cubes of $P(x)$: $P^{2k}(x) = [P^2(x)]^k$, $P^{2k+1}(x) = P^3(x) \cdot [P^2(x)]^{k-1}$. Therefore it suffices only to ensure that $P^2(x)$ and $P^3(x)$ have positive coefficients. If this is true for a certain $P(x)$, then any sufficiently small change in the coefficients of P will only slightly change the coefficients of P^2 and P^3, so they will remain positive. But if, in addition, P has at least one *zero* coefficient, the small perturbation can be chosen so as to make it *negative* and our goal will be achieved.

It is most natural to seek a polynomial with only one zero coefficient, all the other coefficients being ones. A quick check shows that the squares of all such polynomials of degree no greater than three have a zero coefficient. The polynomial of lowest degree satisfying our requirements is

$$P(x) = x^4 + x^3 + x + 1.$$

The coefficients of P^2 and P^3 are positive integers; written in order, they coincide with the digits of $11{,}011^2 = 121{,}242{,}121$ and $11{,}011^3 = 1{,}334{,}996{,}994{,}331$. (The algorithm for multiplying polynomials is exactly the same as the one for multiplying integers, except that no carries are made for polynomials. But in our particular case 11,011 is squared and cubed without carries anyway.)

Thus, the required polynomial can be written as

$$P_\varepsilon(x) = x^4 + x^3 - \varepsilon x^2 + x + 1,$$

where ε is positive and small enough to ensure that the coefficients of P_ε^2 and P_ε^3 differ from the coefficients of P^2 and P^3 ($P(x) = P_0(x)$) by less than one and thus remain positive.

M168

Let z be the smallest of the given side lengths and A the area of the triangle in question. Then the relation between its altitudes takes the form

$$\frac{2A}{z} = \frac{2A}{x} + \frac{2A}{y},$$

or $xy - yz - zx = 0$. But then we have

$$x^2 + y^2 + z^2 = (x + y - z)^2,$$

which is the square of an integer.

M169

(a) Clearly, any position of the

battleship leaves room for the first cruiser. To prove that the second cruiser can be placed after the first two vessels, divide the "ocean" into eight rectangular sections as shown in figure 3. Any of them gives enough room for a cruiser that doesn't border on the neighboring sections. On the other hand, any ship can have common squares with at most two sections. So after the first two ships are drawn, at least $8 - 2 \cdot 2 = 4$ sections will be free for the second cruiser.

Figure 4 shows a similar partition (into 12 rectangles) for destroyers. The battleship, two cruisers, and no more than two destroyers can occupy at most $5 \cdot 2 = 10$ of its sections and will always leave free at least $12 - 10 = 2$ sections for the next destroyer. Of course, each of these sections can hold a destroyer with its one-square-wide neighborhood (see figure 4)

Finally, the 16-section partition in figure 5 proves that room for the submarines can always be found, too. Each of these sections is the one-square neighborhood of a one-square ship (perhaps truncated by the borders of the ocean). And the battleship, two cruisers, three destroyers, and three or fewer submarines overlap with no more than $2 \cdot (1 + 2 + 3) + 3 = 15$ sections, leaving at least one section free for the next submarine.

The trick with this problem is that the approach that springs to mind—counting the total area of the neighborhoods of the ships already drawn and making sure there's some empty space left—doesn't lead to a solution, at least not directly.

(b) An arrangement that leaves no room for a battleship is easy to find. We don't even need to use all nine of the smaller ships—see figure 6.

M170

The answer is $N = n(n + 1)/2 - 1$.

Denote Tarantoga's definitions by $S_1, S_2, \ldots, S_n$. Imagine that the professor represents them by n points on the plane and marks the dissertation that shows that sepulation in the sense of S_i is sepulation in the sense of S_j (which will be denoted by $S_i \to S_j$) by drawing an arrow from point S_i to S_j, thus constructing an oriented graph with n vertices. The last condition of the problem means that the arrow $S_i \to S_j$ can never short-circuit an already drawn path of arrows leading from S_i to S_j (fig. 7).

Some of the points may be joined by a pair of opposite arrows. Let's see how many of these there can be. Consider all these double arrows. Notice that they can't form a cycle, because the single arrow in such a cycle drawn last would make a "short-circuit" (fig. 8). So, if we delete *single* arrows from Tarantoga's graph and replace each double arrow with a single (nonoriented) edge, we'll get a graph on n vertices without cycles. We can show that such a graph has no more than $n - 1$ edges.

Indeed, delete any of its edges. The graph falls into two

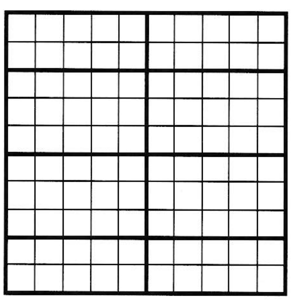


Figure 3

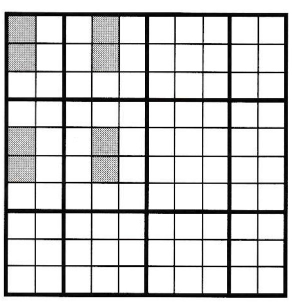


Figure 4

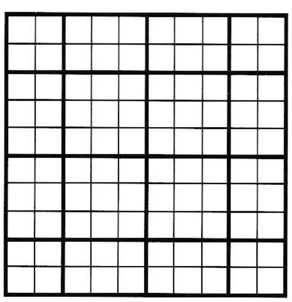


Figure 5

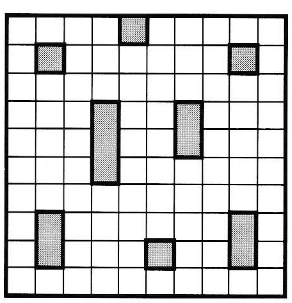


Figure 6

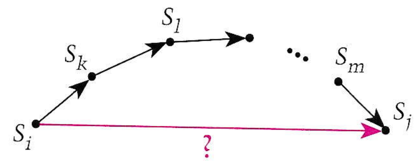


Figure 7

disconnected pieces (otherwise this edge would have to be a segment in a cycle). Delete another edge. The graphs falls into three pieces. After we delete k edges, we get $k + 1$ pieces. But the number of pieces is no greater than n, the number of vertices.

Thus in our original (oriented) graph, there are not more than $n - 1$ double arrows. Since Tarantoga's n points make $n(n - 1)/2$ pairs, the total number of arrows (defended dissertations) does not exceed

$$2(n-1) + \left[\frac{n(n-1)}{2} - (n-1)\right]$$
$$= \frac{n(n+1)}{2} - 1 = N.$$

Now let's explain how to organize the production of such a set of Ph.D.'s.

First, $n - 1$ graduate students defend the dissertations $S_1 \to S_n$, $S_2 \to S_n, \ldots, S_{n-1} \to S_n$. Then $n - 2$ students defend the dissertations $S_1 \to S_{n-1}, S_2 \to S_{n-1}, \ldots, S_{n-2} \to S_{n-1}$, and so on, until one student defends $S_1 \to S_2$. After that another $n - 1$ students defend the dissertations $S_n \to S_{n-1}, S_{n-1} \to S_{n-2}, \ldots, S_2 \to S_1$. This amounts to $n(n - 1)/2 + n - 1 = n(n + 1)/2 - 1$ dissertations, none of which follows from those defended earlier. On the other hand, all the n definitions are proved to be equivalent: this follows from $S_1 \to S_n$, $S_n \to S_{n-1}, S_{n-1} \to S_{n-2}, \ldots, S_2 \to S_1$.

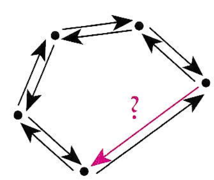


Figure 8

Physics

P166

The center of mass of the half of the snake that is inside the tube is located at a distance $l/4$ from the end of the tube. Its coordinates are as follows (see figure 9, which gives a view from above):

$$x_1 = -\frac{l}{4}, \quad y_1 = 0.$$

The center of mass of the free end of the snake, which can coil arbitrarily on the plane, can be located at any point inside a circle of radius $l/4$ whose center is at the origin of the coordinates—that is, its coordinates (x_2, y_2) conform to the inequality

$$x_2^2 + y_2^2 \le \left(\frac{l}{4}\right)^2. \qquad (1)$$

Since the masses of both parts of the snake are equal, the coordinates of the center of mass of the entire snake must be located midway between the centers of mass of the two halves and can be found from the following relationships:

$$x_{\text{cm}} = \frac{x_1 + x_2}{2} = \frac{x_2}{2} - \frac{l}{8},$$

$$y_{\text{cm}} = \frac{y_1 + y_2}{2} = \frac{y_2}{2}.$$

Using these equations, we express x_2 and y_2 in terms of x_{cm} and y_{cm}:

$$x_2 = 2x_{\text{cm}} + \frac{l}{4}, \quad y_2 = 2y_{\text{cm}}.$$

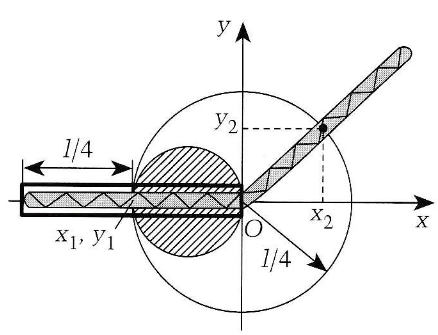


Figure 9

Inserting these expressions into equation (1) we obtain

$$\left(2x_{\text{cm}} + \frac{l}{4}\right)^2 + 4y_{\text{cm}}^2 \le \left(\frac{l}{4}\right)^2,$$

or

$$\left(x_{\text{cm}} + \frac{l}{8}\right)^2 + y_{\text{cm}}^2 \le \left(\frac{l}{8}\right)^2.$$

This means that the center of mass of the entire snake can be found at any point inside a circle of radius $l/8$ whose center is at $(-l/8, 0)$ (the shaded circle in figure 9).

P167

At the moment when the bar's position on the inclined plane is x $(x < l)$, we can write the equation for its acceleration as

$$Ma_x = Mg\sin\alpha - F_f N\left(1 - \frac{x}{l}\right), \qquad (1)$$

where $N = l/2r$ is the total number of rollers on the segment of length l, $N(1 - x/l)$ is the number of rollers that come into contact with the bar at this moment, and F_f is the frictional force acting on the bar from one roller. Assuming that the bar doesn't slide on the rollers, the frictional force F_f imparts a tangential acceleration $|a_x|$ to every roller that is in contact with the bar. Therefore,

$$F_f r = \frac{mr^2 a_x}{r},$$

or

$$F_f = ma_x. \qquad (2)$$

Inserting equation (2) into equation (1) yields the dependence of the bar's acceleration on its position on the inclined plane $(x < l)$:

$$a_x = \frac{g\sin\alpha}{1 + \frac{m}{M}\frac{l}{2r}\left(1 - \frac{x}{l}\right)}.$$

When $x > l$, the bar no longer has contact with the rollers and it slides along the plane with an acceleration

$$a = g \sin \alpha.$$

The general form of the function $a(x)$

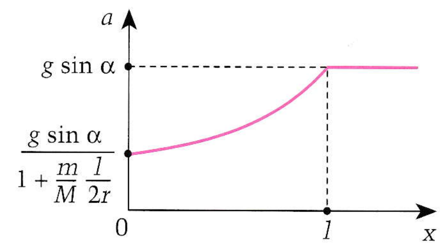


Figure 10

is shown in figure 10.

P168

This problem doesn't require a precise analytical solution but rather an estimate, because the conditions of the task clearly indicate that we can neglect some important factors. First let's find the amount of saturated water vapor in the vessel and compare it with the amount of water:

$$n_{\text{sat}} = \frac{PV}{RT} \cong 4 \cdot 10^{-4} \text{ mol}.$$

Obviously it's a very small fraction of the total mass of water in the vessel ($n_{\text{tot}} \cong 0.05$ mol), so the second question is clear: the water would never evaporate without the help of an absorbent substance.

To calculate the time needed for the absorbent to take up the water vapor, we assume that the vapor is always saturated. Then its concentration is

$$\frac{N}{V} = \frac{P}{kT}.$$

Let's choose a coordinate axis x perpendicular to the absorbent's surface, so that the number of vapor molecules colliding with the absorbent's surface area S in a time Δt is

$$N_{\text{col}} = \frac{1}{2}\frac{N}{V}Sv_x\Delta t,$$

and a similar formula describes the period τ necessary for all the water molecules N_{tot} to be absorbed:

$$N_{\text{tot}} = \frac{1}{2}\frac{N}{V}Sv_x\tau = \frac{1}{2}\frac{P}{kT}S\sqrt{\frac{kT}{m_0}}\tau.$$

After simple rearrangements we have

$$\tau = \frac{2N_{\text{tot}}}{N_A}\frac{\sqrt{\mu RT}}{PS}$$

$$= 2\frac{\sqrt{18\cdot 10^{-3}\cdot 8.3\cdot 278}}{18\cdot 870\cdot 100}\ \text{s}$$

$$\cong 10^{-5}\ \text{s},$$

where μ is the molecular mass of water.

This answer is paradoxically small. We can only say that under the conditions of the problem the absorbent immediately takes up most of the molecules. In reality it falls far short of absorbing every incoming molecule, so the process takes much longer. Also, when only a little water is left and the vapor is no longer saturated, absorption will be slower: the fewer molecules there are in the flask, the slower the absorption.

P169

When connected to a battery, the capacitor is charged to a potential difference V, and its plates acquire charges $+q$ and $-q$, where

$$q = CV = \frac{\varepsilon_0 SV}{l_0}.$$

The upper plate is attracted by the electric field produced by the lower plate, so it is pulled down by a force

$$F = qE,$$

where E is the strength of the lower plate's field. Since the linear dimensions of such capacitors are normally large compared to the distance between the plates, we can approximate this field by that of a uniformly charged infinite plane. The intensity of such a field doesn't depend on the distance from the charged plane and is given by

$$E = \frac{\sigma}{2\varepsilon_0} = \frac{q}{2\varepsilon_0 S} = \frac{V}{2l_0}.$$

The force F causes the upper plate to move and stretch the spring. Like the force of gravity, this electrical force doesn't depend on the plate's position. However, the elastic force of the spring is proportional to the plate's displacement. So a plate of mass m will oscillate harmonically about the equilibrium position (fig. 11), where

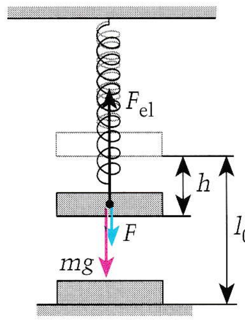


Figure 11

$$F + mg = F_{el}. \qquad (1)$$

The amplitude of the plate's oscillation is equal to the distance h between its initial and equilibrium positions. Therefore, the plates do not touch as long as this distance h is less than half the initial distance between the plates—that is,

$$h < \frac{1}{2}l_0.$$

Let's denote the spring's stretch in the initial state by Δx_0. In the new equilibrium position its stretch is $\Delta x_0 + h$, from which we get

$$F_{el} = k(\Delta x_0 + h).$$

As the upper plate was initially at the equilibrium position,

$$mg = k\Delta x_0.$$

Combining the formulas for F_{el} and mg with equation (1) yields

$$F + k\Delta x_0 = k(\Delta x_0 + h),$$

or

$$h = \frac{F}{k}.$$

Thus the plates don't touch each other if

$$\frac{F}{k} < \frac{1}{2}l_0,$$

or

$$k > \frac{2F}{l_0} = \frac{2qE}{l_0} = \frac{\varepsilon_0 SV^2}{l_0^3}.$$

P170

Light rays do not travel in a straight line in an atmosphere where the refractive index n decreases with altitude. Rotation of the wave front of the light and a corresponding bend of a light beam are the consequences of the fact that the lower the refractive index of a medium, the greater the speed of light $v = c/n$ in it.

Let's denote by Δh the width of the optical channel where the light rays travel around the planet at a constant altitude. Consider two rays at the inner and outer edges of this channel.

The beam that travels at a constant altitude h_0 (fig. 12) circles the planet in the time

$$t = \frac{2\pi(R+h_0)}{v_1} = 2\pi(R+h_0)\frac{n_0 - \alpha h_0}{c}.$$

The other beam, traveling in this channel at a distance $\Delta h \ll h_0$ above the first beam, must circle the planet at a height $h_0 + \Delta h$ in exactly the same time—only in this case will the wave front passing through the channel be perpendicular to the circle of radius $R + h_0$:

$$t = \frac{2\pi(R+h_0+\Delta h)}{v_2}$$

$$= 2\pi(R+h_0+\Delta h)\frac{n_0 - \alpha(h_0+\Delta h)}{c}.$$

Since both times are equal, and $\Delta h \ll h_0$, we get

$$h_0 = \frac{1}{2}\left(\frac{n_0}{\alpha} - R\right).$$

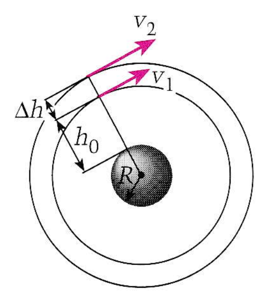


Figure 12

This phenomenon is called circular refraction, and there is evidence that it may occur, for example, in the atmosphere of Venus.

Brainteasers

B166

If Ciliegia works at the slower rate, then he finishes 6 stools between Friday and Sunday, and his total number of stools is a multiple of 5 as well as 3. Any such number is a multiple of 15, and in fact 15 itself works, if we start Ciliegia working on Wednesday. And there are no other solutions possible: if he started working n days before Friday, then $5n = 3(n + 2)$, so $n = 2$.

But there is another possible interpretation of the problem. Perhaps Ciliegia, working at his fast pace, will finish on Friday; and at his slow pace, on a Sunday more than one week later. Then, if n is the number of days before Friday that he started, and m is the number of weeks between the Friday and the Sunday, then Ciliegia works $n + 1$ days at the fast rate and $7m + (n + 1) + 2$ days at the slow rate, so we have $5(n + 1) = 3(7m + n + 3)$, or $21m = 2n - 4$. We need a solution of this equation in integers. There are general and standard ways to do this, which the reader is invited to consult in any book on number theory. Meanwhile, we will note that since $2n - 4$ is even, m must also be even. Letting $m = 2k$, we have $42k = 2n - 4$, or $21k = n - 2$, or $n = 21k + 2$. But $21k$, being a multiple of 7, represents an integer number of weeks. So n days before Friday is still a Wednesday, no matter how many weeks intervene.

B167

The required rearrangement can be achieved by the following three steps:

$$15(624)37 \to 16(245)37 \to 12(456)37 \to 1234567.$$

It's not difficult to see that this is the shortest solution. We leave it to the reader to show that any permutation can be obtained in a similar way.

B168

The water level went down.

B169

Label the veins on the left side in the initial pattern a, b, c and the veins on the right A, B, C, as shown in figure 13a. Now we note that each subsequent pattern is obtained from the previous one by moving one left vein in the order b, a, c, b, a, c, ..., one position down (the vein in the lowest position jumps to the top) and, at the same time, one right vein (in the order B, C, A, B, C, A, ...) moves one position up (or from the very top to the bottom). The unknown pattern is shown in figure 13b.

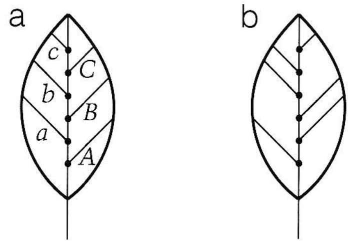


Figure 13

B170

Denote the sides of one of the squares traced clockwise and the straight lines containing them by a, b, c, d and of the other square by a_1, b_1, c_1, d_1. Then we have to prove that the lines AC and BD are perpendicular, where A is the intersection point of a and a_1, B is the intersection of b and b_1, and so on. We'll see that this is true for any two squares, whether they form an octagon or not (fig. 14). It's clear that if we shift the band bounded by a_1 and c_1 in parallel to itself, then the new segment $A'C'$ will be parallel (and equal in

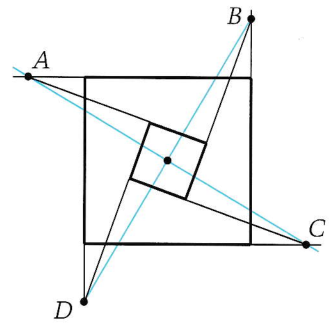


Figure 15

length) to AC (fig. 14). The same is true for the band b_1d_1.

Therefore, it will suffice to prove our statement for any convenient position of the second square obtained from the initial position by parallel translation—for instance, in the case where the centers of the squares coincide (fig. 15). But in this case BD is obtained from AC simply by a 90° clockwise rotation about the common center of the squares. Thus, we have proved not only that $AC \perp BD$, but also that $AC = BD$. (V. Dubrovsky)

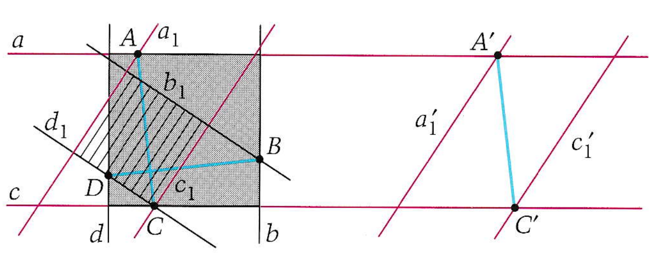


Figure 14

Toy Store

1. The recursive equations for r_k and u_k remain the same, but the "initial values" of these numbers must be changed: now we have $r_2 = 2$, $u_1 = u_2 = 1$. The correspondingly modified calculation yields $r_k = 3 \cdot 2^{k-2} - 1$ $(k \geq 2)$, $u_k = 2^{k-1}$ for odd $k \geq 1$, and $u_k = 2^{k-1} - 1$ for even k.

2. For any 0–1 string $A = a_1 \ldots a_k$, denote by $N(A)$ its number in our table—that is, its "distance," in moves, from the zero string; by $r(A)$ the corresponding "shaded digit"—that is, the sum $a_1 + \ldots + a_k \pmod 2$; and by $\overline{A} = a_1 \ldots a_k$ the string obtained by drawing a bar over every other digit one in the string (for instance, $(\overline{110101}) = 1\bar{1}010\bar{1}$). We have to prove that

$$N(A)=\left(\overline{a_1\ldots a_k r}\right)_2,$$

where $r = r(A)$. Notice that the string $a_1 \ldots a_k r$ always has evenly many ones, and therefore

$$\left(\overline{a_1\ldots a_k r}\right)_2=\left(\overline{a_1\ldots a_k 0}\right)_2-r,$$

no matter if $r = 0$ or $r = 1$.

Let's apply induction over the number n of ones in a string. Let $n = 1$. We can ignore any digits 0 that occur to the left of the single digit 1 in our string. Then the string $A = 10\ldots0$, and $N(A) = 2^k - 1$, where k is the number of digits in string A. On the other hand, here $r(A) = 1$. So

$$\left(\overline{Ar(A)}\right)_2=\left(10\ldots0\bar{1}\right)_2=2^k-1$$
$$=N(A),$$

which complies with our formula. Now suppose the formula is valid for all strings with fewer than n units and consider an arbitrary string A with n units. Again, we can ignore leading digits 0 and write our string as $A = 1a_1 \ldots a_k$. In the table this string occurs before $10\ldots0$ (with k zeros). The distance between these strings equals the distance between the k-digit strings $0\ldots0$ and $A' = a_1 \ldots a_k$, which in turn, by the induction hypothesis, is equal to $N(A') = (\overline{a_1 \ldots a_k r'})_2$, where $r' = r(A')$. Now, using the fact that $r = r(A) = 1 - r'$, we get

$$\begin{aligned}N\left(1a_1\ldots a_k\right)&=N(10\ldots0)-N\left(a_1\ldots a_k\right)\\&=2^{k+1}-1-\left(\overline{a_1\ldots a_k r'}\right)_2\\&=2^{k+1}-\left(\overline{a_1\ldots a_k 0}\right)_2-\left(1-r'\right)\\&=\left(\overline{1a_1\ldots a_k 0}\right)_2-r\\&=\left(\overline{1a_1\ldots a_k r}\right)_2,\end{aligned}$$

completing the proof.

3. It's not hard to see that the folded binary representations of $-n$ are obtained from those of n by "switching the bars." For example, 5 can be written as $(1\bar{1}01)_2$ or as $(1\bar{1}1\bar{1})_2$, while -5 has the representations $(\bar{1}10\bar{1})_2$ or $(\bar{1}1\bar{1}1)_2$. So it will suffice to consider positive integers. We'll call a folded binary representation even or odd depending on whether there are an even or odd number of terms in the alternating sum. It's easy to see that both kinds of representation of an even number $n = 2k$ are obtained by adding zero at the right of the corresponding representation of k; the odd representation of an odd number $n = 2k + 1$ must end in the digit 1 and is obtained by adding this digit to the even representation of k; and the even representation of $n = 2k + 1 = 2(k + 1) - 1$ ends in $\bar{1}$ and is obtained by adding $\bar{1}$ to the odd representation of $k + 1$. The number one has two representations: 1 and $1\bar{1}$. Now the proof can be completed by induction.

4. Denote by N_k the sequence of the numbers of the rings (or shields) moved in the first $2^k - 1$ steps of the optimal solution to the corresponding puzzle. The argument used in the article to derive the equation for r_k shows that $N_k = N_{k-1}kN_{k-1}$, so that $N_1 = 1$, $N_2 = 121$, $N_3 = 1213121$, and so on. But the similar sequence for the Tower of Hanoi satisfies the same equation, so the two sequences coincide.

5. Number the dolls $0, 1, \ldots, k - 1$ in order of size starting from the biggest. Represent any possible arrangement of the dolls by a 0–1 string such that the ith digit in it (counting from the right) is 0 if the ith doll is hidden in the next doll and 1 otherwise. The zeroth digit, corresponding to the largest doll, will always be 1, so we can simply delete it. By the description of our game, the remaining $(k - 1)$-digit string will change in exactly the same way as our binary model. We have to transform the zero string into (a) the string $110\ldots0$ (the first moment when the leftmost digit turns into one); (b) the string $11\ldots1$. It follows that the answer is (a) 2^{k-1}; (b) u_{k-1} (which was calculated in the article).

Surprises

1. If f_{n+1}/f_n has a limit y, then

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

has the limit $1 + 1/y$. Since a sequence can have only one limit, we see that $y = 1 + 1/y$, or $y^2 - y - 1 = 0$, and thus

$$y=\frac{\sqrt{5}+1}{2}.$$

2. This requires

$$\frac{W^2}{W+L}+W-L=0,$$

so $W^2 = L^2 - W^2$, which reduces to $L = \sqrt{2}W$. Then the winning percentage is

$$\frac{W}{W+L}=\frac{1}{1+\sqrt{2}}=\sqrt{2}-1.$$

3. If $P = W/(W + L)$, then $L = W(1 - P)/P$. So we have

$$\begin{aligned} W\cdot\frac{W}{W+L}+W-L &= WP+W-\frac{1-P}{P}W \\ &= WP+W-\frac{1}{P}W+W \\ &= W\left(P+2-\frac{1}{P}\right) \\ &= W\frac{W+L}{W}P\left(P+2-\frac{1}{P}\right) \\ &= (W+L)\left(P^2+2P-1\right). \end{aligned}$$

Magic of 3 × 3

(See the January/February issue of *Quantum*)

1. See figure 16.

2. See figure 17. This is the *lo shu* with 1 taken from each number.

3. There are two solutions. Shift the entire bottom row of cards to the top of the square, or move the entire leftmost column of cards to the square's right side.

219	224	223
226	222	218
221	220	225

Figure 16

1	8	3
6	4	2
5	0	7

Figure17

The latest magic

This will update my offer of \$100 in "The Magic of 3 × 3" (*Quantum*, January/February 1996) for an order-3 magic square made with nine distinct square numbers. Lee Sallows, mentioned in my article, wrote a program that found many almost magic squares in which only one diagonal failed to give the magic sum. His square with the lowest constant is shown in figure 1.

Such semimagic squares exist, as John Robertson of Berwyn, Pennsylvania, has shown, if and only if they consist of three triplets of numbers in arithmetic progression, all with the same differences between adjacent terms. Corresponding terms in the triplets need not be in arithmetic progression, as required for the square to be fully magic. Robertson has also shown that finding all such squares is equivalent to finding all the rational points on certain elliptic curves.

In most cases found by Sallows, the constant is also a square, as in the example given (fig. 1). However, this is not true of all partial magic squares, as shown by the counterexample in figure 2, discovered by Michael Schweitzer, a Göttingen mathematician.

The constant for rows, columns, and one diagonal is the nonsquare 20966014. In my article I said that order-3 squares made of squares are possible with zero in one cell. I should have added that squares of this type are magic only in rows and columns.

Robertson sent a variety of 4 × 4 magic squares made with distinct squares, and called my attention to R. D. Carmichael's *Diophantine Analysis*. Order-3 magic squares for powers of n are impossible unless three powers can be in arithmetic progression. For this to be true, the equation $a^n + b^n = 2c^n$ must have solutions with distinct integers for a, b, and c. Leonhard Euler proved there are no solutions for $n = 3$. This rules out order-3 squares made with cubes or multiples of cubes. Carmichael also shows impossibility for $n = 4$ and multiples of fourth powers. I have been informed by Noam Elkies that if Andrew Wiles's proof of Fermat's Last Theorem is valid, as almost all number theorists now believe, then it can be shown that $a^n + b^n = 2c^n$ has no solution for n greater than 2.

Even though three squares can be in arithmetic progression, there may be no way to construct a 3 × 3 fully magic square with nine distinct squares. Schweitzer has shown that if such a square exists, the central term must have at least nine digits, and if the entries have no common divisor greater than one, all entries must be odd.

—Martin Gardner

127^2	46^2	58^2
2^2	113^2	94^2
74^2	82^2	97^2

constant = 147^2

Figure 1

35^2	3495^2	2958^2
3642^2	2125^2	1785^2
2775^2	2058^2	3005^2

constant = 20966014

Figure 2

Nesting puzzles

Part II: Chinese rings produce a Chinese monster

by Vladimir Dubrovsky

A VERY OLD LEGEND (AT least, some books say it's very old) has it that the puzzle discussed in this article was invented in ancient China by a soldier. As was required by his demanding profession, he often had to take up arms and leave home, marching great distances. And so he was separated from his family for many weeks and months. His young wife missed him very much, and every time he returned home from another war he found her upset and depressed, her beautiful eyes getting sadder and sadder. The warrior gave her wonderful flowers of wild plum, made funny figures from rice stalks and sprigs to dispel her melancholy for a while, but the long, dark nights brought grief to her heart again. One day, after a fierce battle in which our brave warrior was badly wounded, an idea for an amusing new toy occurred to him, one that would help the young woman in her long wait. Using the bamboo shaft of his spear

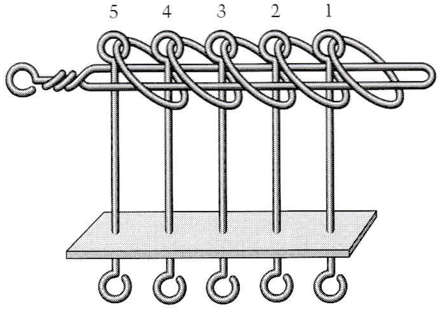


Figure 1

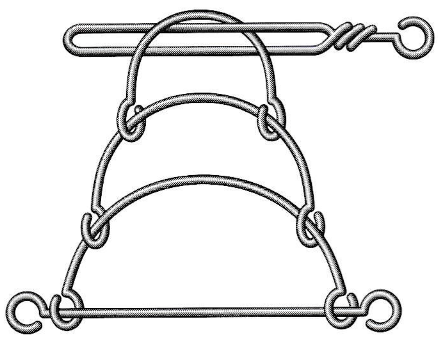


Figure 2

and silk threads that he had unraveled from his headband, he made a game that could be played for days on end. He presented it to his wife, and soon she had again become prettier than everybody else, her eyes shining brighter than ever before.

Whatever you might think about the plausibility of this legend, the puzzle, which is usually made of wire as shown in figure 1, is very old indeed. Suffice it to say that its Russian name, *meleda*, stems from a verb that has long been out of use in the Russian language. (The verb means "to dawdle or loiter." By the way, the French word for this toy, *baguenodier*, has a very close meaning, and in modern China it's called "the horror of guests," because it can be played *ad infinitum*.)

It's known for certain that the ancient Scandinavians used this mechanism as a lock for their trunks. Most probably they were the first to bring the puzzle to Europe. The Europeans gave it its most common and internationally accepted name of the "Chinese rings." The puzzle won the honor of being described and studied by such outstanding mathematicians as Cardano (in 1550) and Wallis (in 1693), and they failed to find its complete solution! Apparently, a full solution was first published by the French mathematician L. Gros in 1893. However, the puzzle is still being reinvented from time to time. The last time it was patented in Europe was in 1931 (in Hungary), and it was patented in the US in 1977. The same or similar idea was used in many other puzzles (see figures 2 and 3).

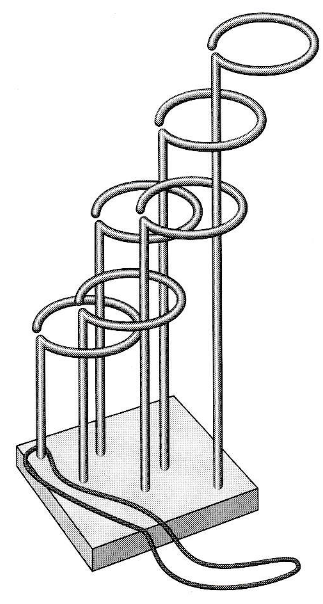


Figure 3

How it works

The modern version of the "soldier's puzzle" (fig. 1) consists of a long and narrow wire loop; a number of rods with rings hinged at their top ends and small bulbs or hooks at their bottom ends; and a metal plate with a row of holes into which the rods are inserted. Each ring except one (at the right end of the puzzle viewed as in the figure) is hooked around the rod next to the right. The loop is passed through all the rings so that it embraces all the rods. All these elements can loosely move with respect to one another, but while the rods and plate make one piece that is mobile but cannot be disassembled, the loop can be separated from the rings. And this is exactly what you have to do with this puzzle.

To solve it, you first have to understand what's possible with this tricky construction. This requires a lot of imagination if you don't actually have the toy, but even if you do, it's a rather difficult task—the puzzle acts like a living thing in your hands, almost, and it's hard to find any system in its behavior. I'm sure, however, that *Quantum* readers could complete this task both ways—in your head and with your hands. Unfortunately I can't take a break while writing this article and wait for your results, so I'll have to give out the answer in order to proceed.

The first ring (numbered as in figure 1) can freely be taken off or put on the loop at any time. For any kth ring ($k > 1$) this is possible only if ring $k - 1$ is put on the loop and all the previous rings (with smaller numbers) are taken off. (In fact, the construction allows rings 2 and 1 to be put on or taken off simultaneously, in one move, but it will be more convenient to ignore this possibility for the time being.)

Now we can create a mathematical model of the Chinese rings and complete the solution with pencil and paper. And we will, after we get to know a younger sister-puzzle of the Chinese rings.

The locking-disk puzzle

This puzzle was invented by William Keister of New York not too long ago. It's a rather enigmatic-looking object (fig. 4) resembling a slide rule.[1] As you see in the figure, it consists of a gray "sheath" that holds a red sliding "shuttle blade" with a row of white interlocking rotary pieces shaped like a heraldic shield (fig. 5). (Sorry, I can't help all these chivalric associations!) Initially all the "shields" point upward (fig. 6a). Your task is to "unsheathe the blade." This can be done only through the right end of the sheath (as viewed in the figures)—the left end is made narrower. What's more important, this is possible only when all the shields are oriented horizontally (fig. 6b)—more exactly, they must point to the left (not to the right—mind the rightmost shield!). In principle the construction allows you to turn some of the shields to the right, but this will only add redundant moves to your solution (think of why, or just take my word for it). Thus we can without hesitation confine ourselves to two positions of a shield—upward and leftward.

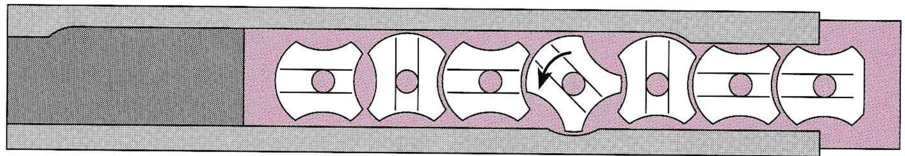


Figure 4

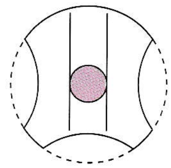


Figure 5

Once we agree to this, it becomes quite easy to describe all possible "elementary" transformations of the puzzle. We see in the figures that the edge of the opening in the sheath is curved in such a way that there is only one place where shields can be turned. We can either push the shuttle all way to the left (as in figure 6a) and reorient the rightmost shield, or pull it to the right until it gets stuck (fig. 4), which brings to the "turntable" the *left neighbor of the first vertical shield counting from the right*. In other words, we can turn either shield 1 or shield k ($k > 1$), if shields $1, 2, \ldots, k - 2$ are horizontal and shield $k - 1$ is vertical.

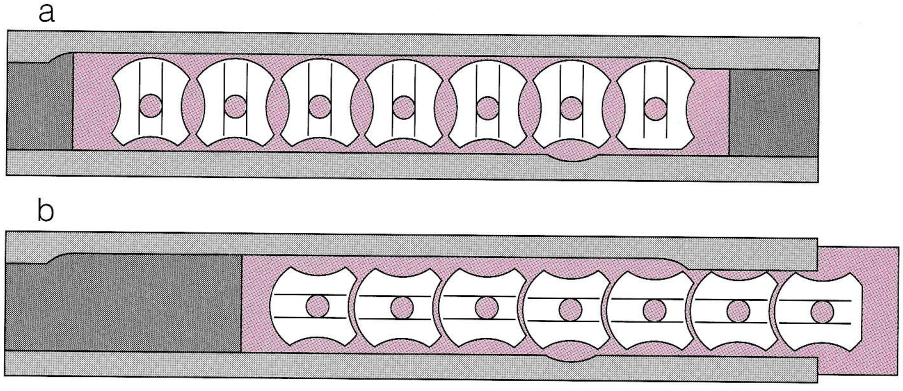
Now the similarity between the locking-disk puzzle and the Chinese rings becomes apparent. Based on this similarity, we'll introduce a common mathematical model for both puzzles.

Transformations of zeros and ones

Suppose each of our two puzzles has m basic elements—rings in the one case and shields in the other. Each of these elements in either puzzle can be in one of two possible positions. These two positions can

[1]I'm afraid that for younger readers of *Quantum* the slide rule is an equally enigmatic object. The Kaleidoscope in this issue contains a description of this once common tool.

a



b

Figure 6

be labeled 0 and 1. Then the entire state of either puzzle will be represented by an m-digit string of zeros and ones. To be more definite, we'll write 1 in the kth place of a string if the kth "Chinese ring" is put on the loop and 0 otherwise; for the second puzzle, ones will designate vertical shields, zeros horizontal. Now the transformation rules for both puzzles amount to the same rule for transforming 0–1 strings:

Given a sequence of zeros and ones, we can either alter its last (rightmost) digit—from 0 to 1 or from 1 to 0—or the left neighbor of the first 1 counting from the right.

For instance, 101100 can be changed to 101101 or 100100. The first of the two operations can be applied to any 0–1 string, the second to all strings except 00…0 and 10…0 (they don't have any "digits to the left of a 1" at all). In these terms, our goal is to turn the *unit* string 11…1 into the *zero* string 00…0 using only these operations. It will be more convenient, however, to start with the inverse problem of turning 00…0 into 11…1.

It should be said that our model deprives the puzzles of some of their special features (for instance, it ignores the third possible orientation of a shield—pointing to the right). These features make the real puzzles trickier than their "mathematical skeletons," but don't affect their optimal solutions.

So let's take the zero string 00…00 and start to transform it. The first move is determined uniquely: 00…00 → 00…01. The second is also unique: 00…01 → 00…11 (because we certainly don't want to return to the initial string by reversing the first move). In fact, for every position we achieve, there are two possible next moves—but one of them will take us back to the position we came from. So if we want to make progress in the puzzle—and assuming that the puzzle can be solved—we will find a unique chain of positions that solves it. In particular, the sequence of positions cannot contain "loops" (such as those illustrated in figures 7a and 7b). We can arrange them in a straight path joining 00…00 to 10…00 (fig 7c). Let's show that this path includes all possible 0–1 strings of the same length.

Denote by r_k the smallest number of moves needed to transform the k-digit zero string into 100…0. By the rules of our game, the first move that replaces the kth (leftmost) zero in the initial string by a digit one is 010…0 → 110…0. So the entire transformation falls into three stages. The first stage takes 00…00 into 0100…00. Since it essentially changes the $(k-1)$-digit string 00…00 into the $(k-1)$-digit string 10…00, this stage consists of r_{k-1} moves. The second stage consists of the single move 010…00 → 1100…00. The third stage turns 110…00 into 1100…00, and is actually the inverse of the first stage. So it, too, consists of r_{k-1} moves. Thus we get the equation

$$r_k = r_{k-1} + 1 + r_{k-1} = 2r_{k-1} + 1,$$

with $r_1 = 1$ (0 → 1). Does this ring a bell? We already solved it in the first part of this article (in the last issue). The formula for r_k is

$$r_k = 2^k - 1.$$

So the path from 00…0 to 10…0 has 2^k "stations" on it (including the endpoints)—that is, all 2^k possible k-digit 0–1 strings. The accompanying table illustrates this result for the case of 5-digit strings.

Now we can solve the problem of finding the smallest number of moves that take the string 00…00 into the string 11…11. For an initial string of k zeros, denote this number by u_k. Repeating the above argument,

n	5	4	3	2	1	0	$d(n)$
0	0	0	0	0	0	0	–
1	0	0	0	0	1	1	1
2	0	0	0	1	1	0	1
3	0	0	0	1	0	1	–1
4	0	0	1	1	0	0	1
5	0	0	1	1	1	1	1
6	0	0	1	0	1	0	–1
7	0	0	1	0	0	1	–1
8	0	1	1	0	0	0	1
9	0	1	1	0	1	1	1
10	0	1	1	1	1	0	1
11	0	1	1	1	0	1	–1
12	0	1	0	1	0	0	–1
13	0	1	0	1	1	1	1
14	0	1	0	0	1	0	–1
15	0	1	0	0	0	1	–1
16	1	1	0	0	0	0	1
17	1	1	0	0	1	1	1
18	1	1	0	1	1	0	1
19	1	1	0	1	0	1	–1
20	1	1	1	1	0	0	1
21	1	1	1	1	1	1	1
22	1	1	1	0	1	0	–1
23	1	1	1	0	0	1	–1
24	1	0	1	0	0	0	–1
25	1	0	1	0	1	1	1
26	1	0	1	1	1	0	1
27	1	0	1	1	0	1	–1
28	1	0	0	1	0	0	–1
29	1	0	0	1	1	1	1
30	1	0	0	0	1	0	–1
31	1	0	0	0	0	1	–1

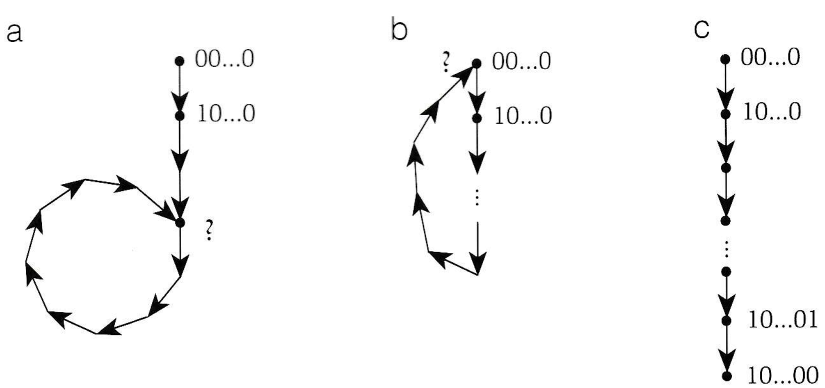


Figure 7

we find that this sequence of moves consists of two parts: 000...0 → 110...0 (as we've seen, it takes $r_{k-1} + 1 = 2^{k-1}$ moves) and 110...0 → 111...1, which coincides with

$$\underbrace{0...0}_{k-2\text{ zeros}} \to \underbrace{1...1}_{k-2\text{ ones}},$$

and so takes u_{k-2} moves. Thus we have

$$u_k = 2^{k-1} + u_{k-2},$$

with $u_1 = r_1 = 1$, $u_2 = 2$. A sharp look at the first three rows of the table (and the columns labeled 1 and 2) will show that $u_1 = 1$ and $u_2 = 2$. This information gives u_k recursively, but a final closed formula depends on the parity of k. For odd k we get

$$u_k = 2^{k-1} + 2^{k-3} + \ldots + 2^2 + 1$$
$$= \frac{2^{k+1} - 1}{3},$$

and for even k

$$u_k = 2^{k-1} + 2^{k-3} + \ldots + 2 = \frac{2^{k+1} - 2}{3}.$$

These are the numbers of moves needed to take apart a locking-disk puzzle with k shields. As for the Chinese rings, as was mentioned above, the first two rings can be taken off or put on the wire loop in one move.

Problem 1. Recalculate the numbers r_k and u_k for the Chinese rings given this condition.

Well, now we know how long the shortest solution is. But notice that we still don't know how to construct it. But this is not so difficult. When we transformed the zero string into the unit string, the only thing we had to take care of was not to reverse the same digit twice in a row—this condition determined all the moves uniquely. But if we want to start with 11...1, which is actually the case with the real puzzles, we'll have to choose between two possible initial moves. Of course, even if we make the wrong choice, we'll solve the puzzle anyway: we'll move down the table, reach the 10...0 string, do an about-face, and go all the way back up the table until we reach the zero string. However, it would be better to start "upward" from the very beginning. The shaded column in our table will help us start in the right direction.

Notice that every move changes the parity of the number of ones in a string. Zeros and ones in the shaded column mark the strings where this number is even or odd, respectively. On the other hand, moves that alter the rightmost digit always alternate with those altering one of the other digits. These remarks imply the following simple "parity rule": *if we start with a string consisting of k ones (and a number of zeros) and want to move up the table in order to obtain the zero string, we must begin by reversing the rightmost digit if k is odd; for even k we must first reverse the left neighbor of the first digit one from the right.* (Of course, this rule must be inverted if we are going to move down the table, to 10...0.)

So the shortest solution of the k-piece locking-disk puzzle begins with turning the first shield for an odd k and the second shield for an even k.

This rule completes the solution, but not the investigation of our puzzles.

Folded binary system

Suppose we are given two arbitrary positions of one of our puzzles. Can we determine the "distance" (in moves) between them and the shortest transformation of one into the other (that is, the first move in this transformation)? This would be easy if our table were big enough to include both positions. The distance would then be equal to the difference of their numbers in the table (found in its first column), and the first move would be determined by their order (plus the parity rule). So all we need to answer both questions is to learn how to calculate the number of a position in the table without writing down the table itself.

The rule for this calculation is really remarkable. We take the "binary code" of the given position (let it be, for example, 10111), adjoin the corresponding "shaded digit"—the sum modulo 2 of its digits—at its right end (in the example, we get 101110), draw a bar over every other digit one ($10\bar{1}1\bar{1}0$), and read the string thus obtained as if it were the ordinary binary notation of a certain number in which the signs of the powers of two corresponding to the barred ones are reversed ($2^5 - 2^3 + 2^2 - 2^1$). The value of this alternating sum of powers of two (in our example, $32 - 8 + 4 - 2 = 26$) is the number of the given position in our table—that is, the smallest number of moves needed to obtain this position from 00...0, and vice versa.

Problem 2. Prove that this rule holds for all 0–1 strings. (*Hint*: a 0–1 string—say, 10111, which we considered above—can be obtained from the zero string in a series of steps in which the units appear in the string, one at each step, from left to right: 00000 → 10000, 10000 → 10100, 10100 → 10110, 10110 → 10111. Trace these steps using the table of transformations.)

The representation of an integer as the alternating sum of decreasing powers of two is called its *folded binary representation* and is denoted by a 0–1 string in which every other digit one has a bar. We write, for instance, $26 = (10\bar{1}1\bar{1}0)_2$. Notice, however, that this number also has another folded binary representation: $26 = (10\bar{1}010)_2$.

Problem 3. Prove that any nonzero integer (positive or negative) has exactly two folded binary representations, with an even number of units in one of them and an odd number of units in the other.

As an exercise, you can check the following formulas for $(k + 1)$-digit strings:

$$(10...0\bar{1})_2 = r_k = 2^k - 1,$$

$$(1\bar{1}1\bar{1}...1\bar{1}) = u_k \text{ (for odd } k),$$

$$(1\bar{1}1\bar{1}...1\bar{1}0) = u_k \text{ (for even } k),$$

where u_k is defined as above.

Return of the dragon curve

Look again at our table. Let's move from top to bottom, writing out the numbers of digits changed as we pass from each line to the next (these numbers are found in the top line of the second column):

$$1, 2, 1, 3, 1, 2, 1, 4, 1, 2, 1, 3, 1, 2, 1, 5, \ldots.$$

Do you recognize this sequence? Yes, you saw it in the first part of this article: it's the sequence of the numbers of disks successively moved in solving the Tower of Hanoi puzzle!

Problem 4. Prove that the sequences describing optimal solutions of the two kinds of puzzles will coincide no matter how long they are continued.

So the Tower of Hanoi is, in a certain sense, isomorphic to the Chinese rings puzzle (and to its "slide rule" and "digital" relatives).

It would be interesting to develop this observation and establish the correspondence between the ordinary binary codes of the states of the Hanoi Tower and the folded binary codes we studied above. But our table reveals another, much more interesting relationship.

Each move in the table alters one digit in the current string—0 to 1 or 1 to 0—and so changes the sum of all digits by 1 or –1. These changes (denoted by $d(n)$, where n is the number of a move) are written out in the last column. Imagine a bug crawling on the coordinate plane: it starts at the origin; crawls along a unit segment joining it to point (1, 0); turns by $d(1) \cdot 90° = +90°$ (in the positive, counterclockwise direction—that is, to the left); crawls another unit segment, ending at (1, 1); then turns by $d(2) \cdot 90° = 90°$ again; crawls to (0, 1); makes a $d(3) \cdot 90° = -90°$ turn (to the right); moves to (0, 2); and continues in the same way, reading the sequence $d(n)$ and making the corresponding turn at the end of each unit segment it covers. Draw the bug's path. You must have seen it in *Quantum*: it's the so-called *main dragon design* (see the article "Dragon Curves" in the September/October 1995 issue). A proof of this remarkable fact is not difficult, but lies beyond the scope of this article. It's based on one of the definitions of the dragon design and the following equation for the sequence $d(n)$:

$$D_{k+1} = D_k 1 \overline{D}_k,$$

where D_k denotes the segment $\{d(1), d(2), \ldots, d(2^k - 1)\}$ of the sequence (corresponding to the transformation of the k-digit zero string into 10...0) and $\overline{D}_k$ is obtained from D_k by reversing the order and signs of its terms. The equation is proved by the same argument as the equation for r_k derived above.

This connection between our "binary puzzle" and the dragon curve may seem rather artificial. But it goes far beyond the formal analogy between the recursive relations defining the sequences of moves in the puzzle and turns on the dragon path. As we've seen, given a 0–1 string, we can compute the number n of the line in our table where it appears. It turns out that the location of the nth turn of the main dragon design on the plane can be computed from this binary string in much the same way. We merely have to replace the powers of two in the folded binary representation of n by the same powers of the complex number $1 + i$, and the alternating coefficients 1 and –1 by . . . But no, I don't want to reveal all the secrets right away. That would be too much for one article. Besides, I'm sure you'll have much more fun finding them yourself, which is not so difficult now. After that, you'll be able to establish, perhaps, the most remarkable property of the main dragon design (see problem 11c in "Dragon Curves"): extended to infinity, this polygonal path—together with three copies of it obtained by rotating it through 90°, 180°, and 270° about its origin—fill without gaps and overlaps the complete unit-square grid.

I began the first part of this article by comparing nesting puzzles with the popular Russian *matryoshka* doll (which has a number of similar nested dolls inside). I'll finish with a problem in which the *matryoshka* itself is used as a nested puzzle.

Problem 5. A *matryoshka* toy consists of k nested dolls. You are allowed to open either the largest or the next largest unhidden doll, remove the next smaller doll from it or put the next smaller doll in it, and close it. Initially all the dolls are hidden in the largest one. What is the least number of moves required to (a) extract the smallest doll; (b) completely disassemble the toy?