QUANTUM

JULY/AUGUST 1992

$5.00

The student magazine of math and science

SPRINGER INTERNATIONAL

National Gallery of Art, Washington (Gift of Edgar William and Bernice Chrysler Garbisch) © NGA

Peaceable Kingdom by Edward Hicks (ca. 1834)

WHEN EDWARD HICKS DIED IN 1849, THOUSANDS attended his funeral. It's safe to say that most of the mourners came to honor not a painter but a preacher. A devout man and a powerful speaker, Hicks was better known in his lifetime for his pious words than his peaceful landscapes. He even had his doubts whether art is compatible with religion. But he managed to convince himself that painting can bring meaning to life by illustrating moral ideas—and proceeded to paint like the devil!

Hicks was a Quaker, and his pacifist beliefs are evident in "Peaceable Kingdom." He believed that the state of Pennsylvania in the New World was the fulfillment of the Biblical prophecy: "The wolf also shall dwell with the lamb, and the leopard shall lie down with the kid; the calf and the young lion and the fatling together; and a little child shall lead them . . ." (Is. 11:6–9). Hicks depicted this favorite passage again and again, perhaps as many as 100 times. More than two dozen versions of "Peaceable Kingdom" painted by Hicks are known to exist. (The theme has been treated by others—for instance, Henri Rousseau.)

Hicks began his artistic career applying paint to coaches and signs, not canvas. He was not formally trained, so his work is variously labeled "primitive," "folk," or "naive." Although the terms sound condescending, art historians use them simply to denote a genre, not to indicate their relative merit. In fact, you can find entire wings in museums devoted to "primitive" art.

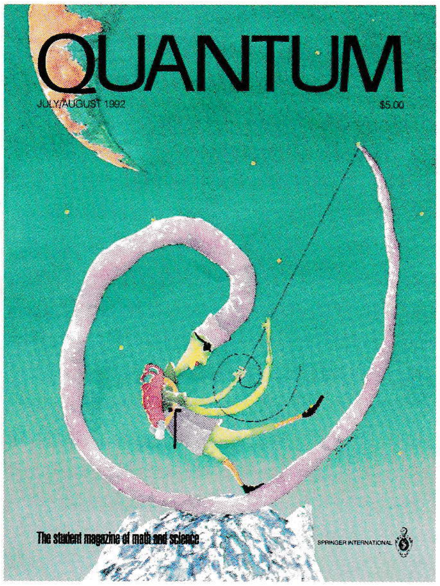
On page 18, you'll encounter an altogether different picture of the animal kindgom. Whereas the prophet has the lion "eat straw like the ox," the sharks and mackerels in Constantine Bogdanov's article behave quite naturally.

And did you notice the river meandering in the background? Any child knows that rivers never flow in a straight line. But why do they curve the way they do? Turn to page 34 for a glass of tea and an answer.

QUANTUM

JULY/AUGUST 1992

VOLUME 2, NUMBER 6



Cover art by Sergey Ivanov

At the very beginning of *The Curves of Life* (Dover Publications), Thomas Andrea Cook freely admits that his knowledge of botany and biology was "as slight as his skill in mathematics or his erudition in the development of art" when he began his monumental study of the spiral. Since he could not hope to know four or more scientific disciplines thoroughly, he thought it more prudent to specialize in none but to have "the keenest sympathy with each." He invites experts to correct his errors and proceeds in the hope that "the artist or the architect will consider the biologist in a kindlier light, and that the mathematician and the botanist may lie down together."

This generous scholar has kindly pointed us in at least three directions: to the cover story, beginning on page 4; to Gallery Q; and to the Publisher's Page. Cook goes on to write: "When Captain Scott was in winter quarters near the South Pole, he overheard a biologist of his party offering their geologist a pair of socks for a little sound instruction in geology. So fruitful an attitude of mind need not be limited to the Antarctic region." Amen!

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Photosynthesism

The "interleaving" of scientific and mathematical disciplines

ASK RESEARCH SCIENTISTS or mathematicians to identify themselves. After giving their names, they will most likely add a clarifying label: nuclear theorist; ergodic theoretician; topologist; theorist in combinatorial analysis; low-temperature, solid-state experimental physicist; geneticist; molecular biologist; and so on. Likewise, we have the traditional, broad subject areas like physics, chemistry, biology, and the Earth and space sciences (astronomy, geology, meteorology, oceanography). There is no such thing as the "compleat scientist," it seems.

Are the various scientific disciplines and subjects really distinct? Are they *disjoint*? And what about science as a whole? What are the relationships among the branches of science?

There's a lot of talk today about the importance of learning science from the perspective of overarching themes, such as energy; or of integrating the sciences to avoid the "arbitrary boundaries" of the many disciplines and subdisciplines. Yet researchers inevitably narrow their focus. Do they in fact restrict their work to their own subfields?

In mathematics, research specialization is almost always the route to successful results. A specialist in combinatorics is working in a hybrid subfield between partition theory in the mathematical subfield of number theory and a supposedly separate subfield called combinatorial analysis. How then should the serious student learn mathematics as an integrated or theme-oriented subject? When does the specialization occur? And does one ever really work exclusively in a subfield?

It's not hard to identify certain general processes used by scientists to produce what we might call scientific products. In varying sequences, we observe, classify, measure, infer, control variables, hypothesize, develop models and theories, and make predictions. As a result of these processes, we create terms, determine facts, invent concepts, find certain empirical laws, produce models and theories, and apply these in a variety of situations.

A law in science is a definitive assertion of a relationship that has been repeatedly challenged but has never been denied. A careful analysis of elementary textbooks in science shows that there are very few fundamental laws of science. Most of these are in physics—for example, Newton's laws of motion ($\mathbf{F} = m\mathbf{a}$ [second law]), the law of universal gravitation ($F = G\frac{m_1 m_2}{r^2}$), Coulomb's law ($F \propto \frac{q_1 q_2}{r^2}$), and so on. The quantitative character of physical law allows a natural correspondence between pure mathematics and the sciences, enabling science to use math as a tool. The other sciences are largely applications of those fundamental laws in a sequence that broadens across the fields as you move among physics and chemistry and biology and the Earth sciences.

Now consider a specific area of science that is regarded as fundamental knowledge: photosynthesis. Here is a subject that is supposedly "biology." Yet it is inextricably linked with, and usually studied as part of, other subjects (for example, biophysics, physical chemistry, biochemistry).

The pure *biology* of photosynthesis is quite narrow. Terminology and classification predominate: autotrophs, chlorophyll, carotenes, accessory pigments, chloroplasts, grana, stroma . . . At higher levels of abstraction, there are enormously complex biological processes that also occur in photosynthesis, some indirectly linked with nitrogen-fixing bacteria.

The *chemistry* of photosynthesis involves complex organic reactions, requiring inputs of carbon dioxide and water and producing, after a series of steps, sugars and oxygen. These reactions require the active involvement of two different kinds of chlorophyll and can occur only from exposure to artificial or natural light. When we begin to talk about light, we enter the realm of *physics*. The incidence of certain wavelengths of the electromagnetic spectrum is essential to the process of photosynthesis. The interaction of light of these wavelengths with the chlorophyll involves electron excitation by photons. The "light reactions" of photosynthesis, as well as the "dark reactions," involve energy exchange and the synthesis of molecules in the plant.

The overall linked process of photosynthesis reduces carbon dioxide with hydrogen taken from water and thus releases oxygen from the water. The removal of carbon dioxide from the atmosphere when the carbon atoms are tied up in carbohydrates, such as sugars and cellulose, and the subsequent decay and deposition of carbon compounds in the land and waters of the Earth are part of the carbon cycle in the *Earth sciences*.

The point is this: how can you learn the biology of photosynthesis without also applying laws of physics and chemistry, and how could you study the carbon cycle in the Earth sciences without photosynthesis? Where is the narrow discipline in all this?

Now the crucial questions: What's the best way to learn science? Should the physics be learned concurrently, and as you need it, while you learn chemistry? And should the chemistry and additional physics be learned concurrently, as you need it, while you learn biology? And should the physics, chemistry, and biology be learned concurrently, as you need it, as you learn the Earth sciences?

These questions get at the essence of the issue of learning separate disciplines versus learning integrated science. How do you decide if the empirical gas laws—laws used in all four broad subject areas—belong to physics, chemistry, biology, or one of the Earth sciences? If the same laws are applicable in all areas, what difference can it possibly make?

And what about specialization for the serious student? As you learn the sciences and the mathematics needed to understand them, there must be areas where interesting but unanswered questions arise in your mind. When those questions are connected to our understanding of an important area of science like photosynthesis, narrow, specialized research can often provide the missing link to a better understanding of the phenomenon. Without the larger perspective, the significance of narrow research is often lost. Indeed there have been times in the past when such research results have been found in the literature years after its need was identified.

I haven't tried to answer the questions I've raised. I'm committed to the notion of learning fundamental science first and using it in applications later. It would be interesting to know how bright, inquisitive students feel about such matters. Write to *Quantum* and tell us what you think.

—Bill G. Aldridge

QUANTUM

THE STUDENT MAGAZINE OF MATH AND SCIENCE

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

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York Inc. Volume 3 (6 issues) will be published in 1992–1993. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by the Russian Academy of Sciences and the Russian Academy of Pedagogical Sciences, as well as original material in English. **Editorial offices:** NSTA, 3140 North Washington Boulevard, 2nd Floor, Arlington, VA 22201. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. Production Editor: Madeline R. Kraner

Advertising:

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

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

Subscription Information: North America

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

All Countries Outside North America

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

This project was supported, in part, by the
National Science Foundation
Opinions expressed are those of the authors and not necessarily those of the Foundation

Friction, fear, friends, and falling

Contemplations of a climbing physicist

by John Wylie

I LIKE FEW THINGS BETTER than to discover that wildly diverse elements in nature have a thread of physics in common. It doesn't matter if I am the first to make such a discovery—most of the time I find that I'm about 200–300 years after the fact. The satisfaction comes from doing it on my own and relating it to my own experiences.

I've been learning the rudiments of rock climbing and mountaineering. Now here is an arena ripe with good science, from atmospheric physics to the dynamics of avalanches to the mechanics of the body while climbing to the technology of modern rope production. On a recent trip to the mountains, I found myself at one point standing on a very steep rock slope surveying the glacier some 600 meters below when, being a physicist, I had cause to reflect on the friction that was responsible for my state of well being.

Friction

Recently there has been a surge in interest in the sport of rock climbing. As with many other sports that have become popular, its growth has been accompanied by a technological revolution in the gear required (although, as in many sports, "desired" is perhaps a better word). Arguably the biggest change in the sport has been the introduction of high-friction climbing shoes. Unlike the heavy lug-soled hiking boots that have been used for years, these shoes are very lightweight and have a smooth rubber sole that has much in common with the slick tires used on racing cars. The impact these shoes have had on the sport can be illustrated by the traditional block-on-an-inclined-plane problem.

Consider one of the new-generation rock shoes on a rock plane, inclined at an angle θ to the horizontal. A standard treatment would allow one to calculate the maximum angle for which the shoe will not slide down the plane. Balancing the components of the forces along the inclined plane (fig. 1), one finds

$$mg \sin \theta = \mu N = \mu mg \cos \theta,$$

where μ is the coefficient of static friction between the sole of the shoe and the surface of the inclined plane. This gives $\tan \theta = \mu$. This isn't a startling result until one realizes that $\mu \cong 1.2$ for this new rubber on smooth granite. This corresponds to an angle of 50 degrees! It's disconcerting for the novice but commonplace for an experienced climber to walk up (or down) a slope approaching this angle.

We can now understand the physics behind every good climber's style. They keep their bodies out and away from the rock. A natural inclination, on the other hand, is to cling desperately to the rock face and pull oneself into the wall (fig. 2). In pushing her body away from the rock, our climber puts most, if not all, of her weight onto her feet—onto her expensive rubber climbing shoes. The greater normal force on the climber's feet results in an increased frictional force. The poor climber distributes too much weight onto her hands, where

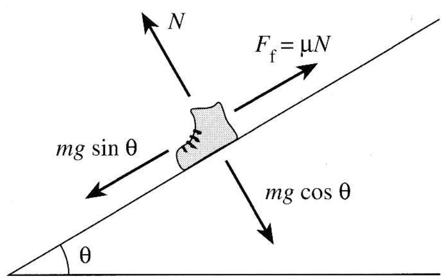


Figure 1
A climbing shoe on an inclined plane. The shoe won't slide down if the component of the force of gravity directed down the plane is balanced by the frictional force directed up the plane.

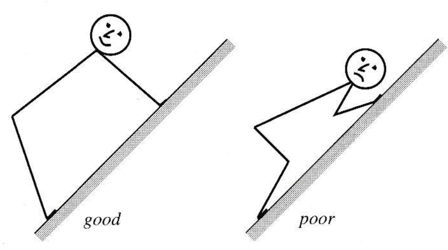


Figure 2
Two possible stances for a climber on a steep slope. The poor stance doesn't distribute the bulk of the climber's weight onto her feet, and the reduced frictional force is likely to cause her to slip.

the friction isn't nearly as good . . . and is likely to get worse as the climber gets nervous.

Fortunately not all climbing is on such easy ground. The most exciting climbing is done on more vertical terrain. The rule is still "keep your feet on the rock." A good example of this is chimney or corner climbing. In these situations a climber finds himself between two opposing rock walls and must use his legs to stem between them. The climber in figure 3 must ascend between two parallel walls. This chimney is wide enough to allow the stance shown; narrower chimneys may require the climber to adopt a style in which his back and one foot are on one wall and the other foot is pressed against the opposing wall. In any case, the principle of opposition is the same. The climber in figure 4 is using the stemming stance to climb a steep corner. To understand the physics behind either situation, let's look at the forces acting on the chimney climber's leg.

For the climber in figure 3 to remain in a comfortable equilibrium, the resultant force on his leg due to the rock wall should be directed along the long bones of his straight leg. As in the last problem, we find that this requires that $\tan\theta = \mu$. Here the climber faces some decisions, however. At an angle of less than this amount, the frictional force is larger and the climber is more secure, but at the expense of increased stresses in the climber's legs and hips. Certainly the climber's general flexibility is a factor as well.

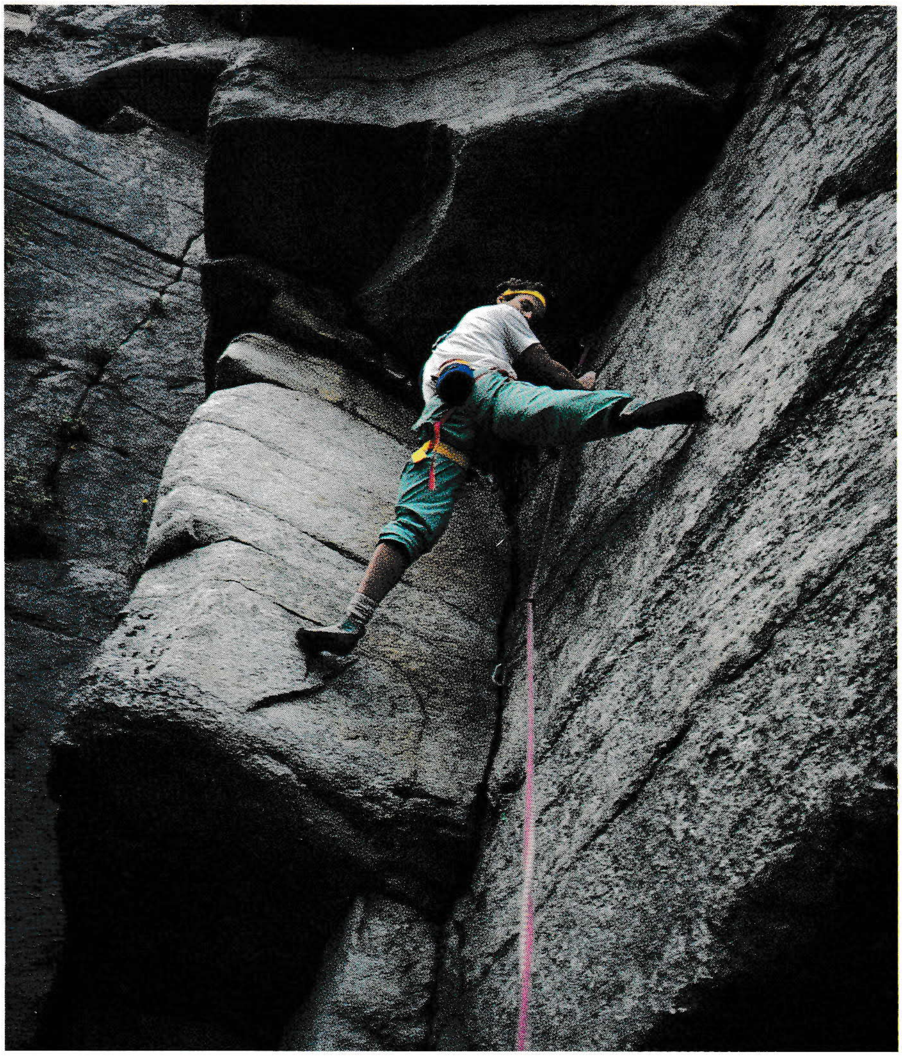


Figure 4
The author using a stemming stance to ascend a steep corner. His hands and arms are being used only for balance. The author's partner (or "second") is below and out of the picture, paying out rope as needed and ready to catch a fall.

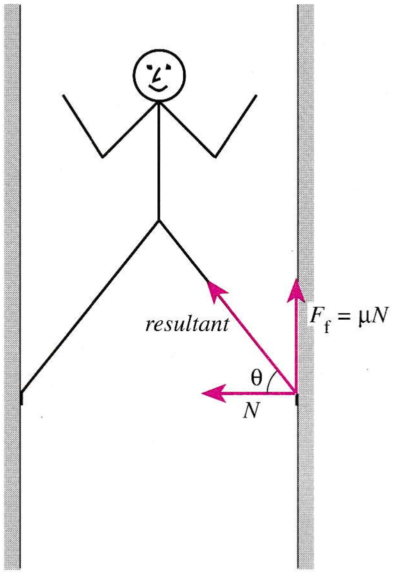


Figure 3
A climber using a stemming stance in a wide chimney. The stance is very comfortable and can be easily maintained if the resultant force of the rock wall on the climber's foot is directed along the length of the climber's leg.

Each climber has an optimal angle, arrived at naturally through experience, with which they are most comfortable with regard to their flexibility, security, and stress. To climb a wider (or narrower) chimney, however, a climber is forced to vary from this angle. Wouldn't it be nice if a climber could extend or contract the length of his legs so as always to be using his optimal angle while climbing? This is the impetus (sort of) for the next most important revolution in climbing technology, but first a little bit about fear and climbing methodology.

Fear

With solid climbing technique, strength, and absolute confidence, climbers would have no use for ropes and all the other bits of gear that are used for safety. The truth is that all competent climbers go "solo" to one extent or another, but most prefer the safety of a rope, particularly when on new or challenging routes or when falling rocks present a hazard.

As a climber leads upward, trailing

a rope out behind her, she must somehow attach the rope to the rock so that if she falls, she will tumble only twice the distance from her last point of attachment. A climbing partner or "second" is below, paying out the rope as required, ready to catch a fall. One of the traditional methods for attaching the rope to the rock has been with pitons. A piton (pronounced PEE-tahn) is a steel spike hammered into a crack or crevice in the rock. A snap-link aluminum ring called a carabiner (kah-rah-BEE-ner) is attached to a hole in the protruding head of the piton, and the rope is clipped through it. Climbers have always appreciated that this method is not an entirely acceptable arrangement.

First, it's hard work to drive pitons into rock; second, they damage the rock to a significant degree. The past decade has seen the energetic development of clean "protection"—that is, small aluminum chocks and nuts that serve much the same purpose as pitons but can be placed and removed without undue effort and without damaging the rock. Most of these pieces of protection must be wedged into small cracks and crevices in the rock, and the rope must be attached to them with carabiners. Figure 5 shows how a rope could be anchored to the rock by these methods. (If you look back at figure 4 again, you can see the carabiner attached to a sling, which in turn is attached to the protection in the rock wall between the climber's legs.)

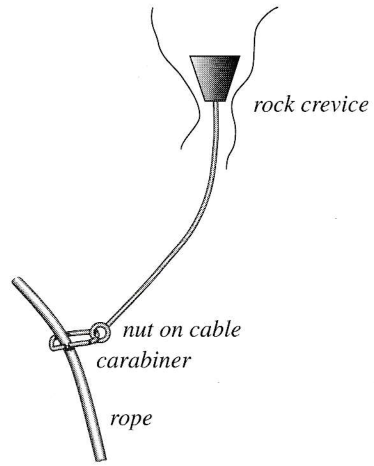


Figure 5
An aluminum nut on a wire cable, wedged into a constriction in a crack. The rope passes through a carabiner clipped to the wire cable. In the event of a fall, the downward load on the nut will only cause it to wedge more tightly, yet it's easily removed. In most cases a sling will be used to connect the carabiner to the cable to prevent the nut from working loose as the rope bounces about.

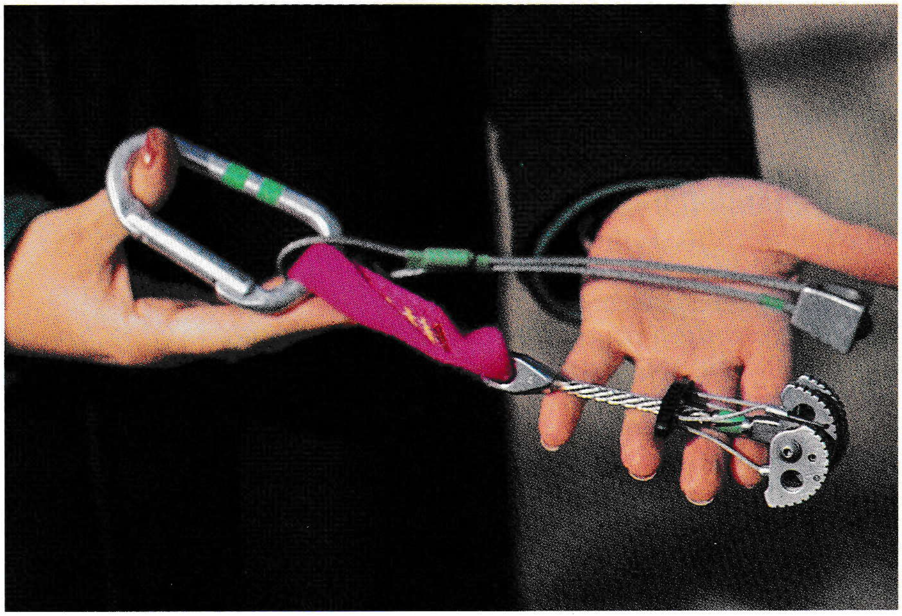


Figure 6
A Friend, a nut, and a carabiner. Notice the springs on the Friend. These merely hold the device in place in the absence of a large load.

One bugaboo that remained unsolved for quite some time was protecting parallel-sided cracks into which neither pitons nor these new pieces of protection could be wedged. These cracks are essentially very narrow chimneys (1–10 cm wide). Many such possible crack climbs remained unscaled as few climbers were bold enough to lead them without protection. An elegant solution would be to invent a device that mimics a climber's solution to the problem: a piece that uses friction to remain in place under a load. And this brings us to one of the most beautiful applications of physics in all of climbing.

Friends

The solution might go something like this. Let's create a device with two cams (legs) hinged on a central shaft (body) that mimics the climber in figure 3. The climber, you'll recall, had his legs at an optimal angle only for one particular width of crack. Our device should be usable in a range of crack widths and always function at its optimal angle θ. In the case of our device (see figures 6 and 7), this optimal angle will depend on its design, the materials it's made of, and the stress its axle can withstand. Certainly the cams must be designed with an expanding radius r so as to maintain a constant angle θ as the

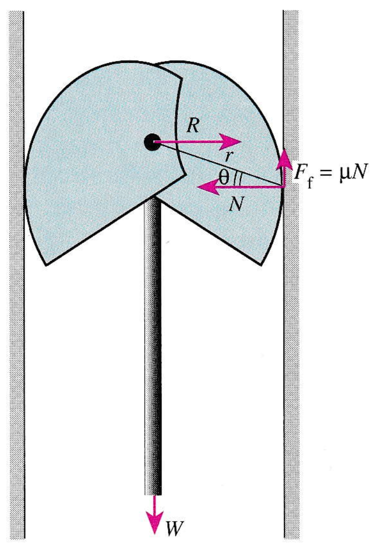


Figure 7
A Friend in a vertical, parallel-sided crack. This is the classic placement for a Friend. The figure shows the forces acting on the Friend under a load. The most important consideration in designing a Friend is the angle θ between the cam's radial arm and the normal to the rock wall (the tangent to the cam's curve).

width of the crack widens. Such devices have been developed and carry the generic label "spring-loaded camming devices." One of the most popular of these is marketed under the name Friend.™ The name is so apt I'll use it in the rest of this article. (In figure 4, you can see a small Friend placed in the vertical crack in the rock wall at leg level and another in a roof crack just above eye level. A properly placed Friend is out of sight!)

Let's look in some detail at the forces that act on a Friend when it's placed in a vertical crack as in figure 7. There is a normal force N due to the wall, a reaction force R acting on the axle, a load weight W, and a frictional force $F_f = \mu N$ between the rock and the cam. Balancing horizontal and vertical components of these forces gives

$$R = N,$$
$$W = \mu N.$$

Choosing a pivot at the point of contact between the rock and cam allows us to write the torque equation

$$Rr \sin\theta = Wr \cos\theta.$$

Combining these three equations yields the fundamental design requirement of a Friend:

$$\tan\theta = \mu \equiv \text{constant}.$$

What must the shape of the cam be to satisfy this requirement? Any mathematician will know the answer immediately. (Being a physicist, I first worked out the solution and then felt silly when I found out how well known the answer is!)

The shape of a cam is the logarithmic spiral

$$r = r_0 \exp(-\alpha \tan\theta),$$

as expressed in the polar coordinates (r, α). Here θ is the angle that the curve's radial arm makes with the normal to the tangent at any point on the curve. This angle is a constant! Since we have already found that our Friend must satisfy the expression $\tan\theta = \mu$, we have the equation of a Friend's spiral in terms of the coefficient of friction:

$$r = r_0 \exp(-\alpha\mu).$$

This curve is shown in figure 8. For this graph, a value of $\mu = 0.3$ was chosen. Physically, this corresponds to an aluminum cam on smooth granite. Different brands of camming devices have slightly varying contact angles, which give them each a unique character. Superimposed on the plot is a sketch of an actual cam.

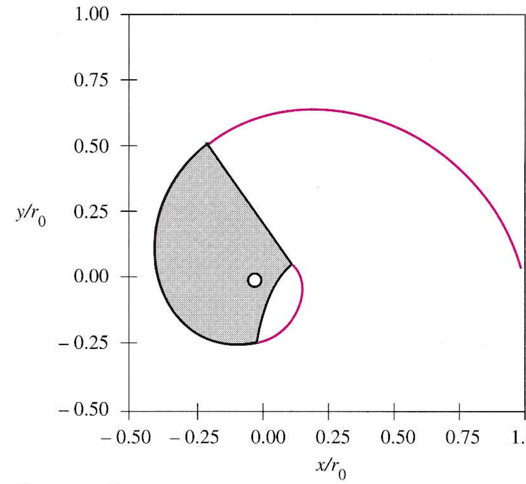


Figure 8

A plot of a logarithmic spiral. A contact angle $\theta \cong 17°$ is used. This corresponds to a coefficient of friction $\mu \cong 0.3$. The shape of an actual cam is superimposed on the plot; its axle corresponds to the origin of the curve. Cams of various sizes are all sections of the same curve.

Mathematicians know this curve well—it has been studied since 1638, first by René Descartes and later by Jakob Bernoulli and Evangelista Torricelli. The logarithmic (or equiangular) spiral has a number of other interesting properties that are worthy of further investigation. The shape of the spiral is independent of the scaling parameter r_0, and the origin is an asymptotic point. The spiral continuously approaches the origin but never arrives. A student could then look at the scaling properties of this spiral by plotting ever increasing magnifications of the neighborhood around the origin. The shape of the spiral is independent of scale! In this sense, the spiral has much in common with a fractal.[1]

Friends are in fact produced in a range of sizes for cracks of varying widths; the cam shapes are all sections of the same spiral. A Friend, properly placed, will undergo structural failure before it slips out of a parallel-sided crack. This failure occurs at a load of well over 1,000 kg.

So there I am on the rock face, contemplating all of this, when it occurs to me that in a serious fall, I could very easily generate a load well in excess of this. That's where the rope comes in.

Falling

Imagine taking a fall onto a rope with little or no dynamic characteristics—that is, a rope that doesn't stretch. Suppose that after falling some distance, a climber attains a downward momentum that is arrested and reduced to zero in a time Δt by the action of the rope. Denoting the change in the climber's momentum by Δp, we see that the average force $F_{av} = \Delta p/\Delta t$ exerted on the climber and on the protection attaching the rope to the rock can be immense if the time is small and fall is great. Quite apart from being a crippling experience, the load placed on the protection may cause it to fail or pull free of the rock, which could render the experience fatal.

That's why climbers use dynamic ropes. The initial potential energy that a climber carries into a fall is, to a very good approximation, completely dissipated as heat as the rope stretches. Indeed, one of the dangers of a fall is the heat generated, which can damage a nylon rope. Static ropes, in contrast, are used when one must climb on the rope itself. Cavers use static ropes, as do rescue personnel. Remember that rock climbers don't often use their ropes to aid in the climb—they bring them along to provide protection in case of a fall. Ropes are manufactured to very specific tolerances, and a wise climber is always relating these to the climbing situation at hand.

[1]See the feature article on fractals in the last issue of *Quantum*.—*Ed.*

The UIAA (an international alpine standards association) sets the following criteria for a rope to bear their approval. A mass of 80 kg held by 2.8 m of rope falling 5 m must generate an impact force of no more than 1,200 kg on the first such test fall. We forgive them for using kilograms as a unit of force; the figure should be about 12 kN. Moreover, a rope is rated by the number of such consecutive test falls it can survive (usually eight or nine). A little physics helps illustrate the design considerations that go into a modern climbing rope.

The quality of a rope is characterized by its rope modulus $M = EA$—the product of its Young's modulus E and its cross-sectional area A. Typically, $M = 40$ kN. Young's modulus is defined as

$$E = \frac{Fl}{A\Delta l},$$

where F is the tension in a rope of length l that induces an extension of the rope by Δl. The ratio of the height of a fall to the length of the rope between a leader and his second is called the fall factor and describes the seriousness of the fall. At a fall factor of $\phi = 2$, the height of the fall is twice the length of the rope (that is, there can be no intermediate points of protection between a leader and his second).

Energy considerations require the work done by the rope in arresting the fall to match the fall's potential energy. That is, $mgh = F\Delta l$, where m is the mass of the climber and h is the height of the fall. The tension in the rope is then $F = mg + mgh/\Delta l$, since the rope must arrest the fall and support the climber's weight. From the definition of the rope modulus we have $\Delta l = Fl/M$, so that

$$F = mg + \frac{Mmgh}{Fl}$$
$$= mg + \frac{Mmg\phi}{F}.$$

In solving for F we choose the $F > 0$ solution from the resulting quadratic

$$F = \frac{mg}{2} + \sqrt{\left(\frac{mg}{2}\right)^2 + \phi Mmg}.$$

With the UIAA test data, we calculate that $F \cong 8{,}000$ N and $\Delta l \cong 56$ cm for a fall of factor $\phi = 1.8$. This corresponds to a percentage stretch of about 20%. The most serious, or "factor two," fall puts both the leader and the second in peril. The nylon fibers in the core of a climbing rope may be stretched beyond their elastic limit, or they may tear altogether. Under less threatening conditions, a climber can count on a 5–10% give in the rope to ease the impact. In all of this we have ignored the mass of the rope itself (about 73 g/m). It makes an interesting problem to determine its significance in climbing practice.

What do we learn from this? A "safe" fall isn't necessarily a short one; it's one with a small fall factor. A long length of rope between the leader and the second, with protection placed just below the leader, would result in a small impact force on the climber and the protection. So climbers will tend to take more risks or "run it out" when they are far above their second than otherwise. One must move very conservatively and place protection frequently at the beginning of a pitch of climbing (when just starting out to climb higher than one's second).

So how trustworthy is a Friend? Clearly, under the circumstances of a severe fall, not very. But with an appropriate understanding of rope and friction physics, it's truly one of a climber's best friends!

Questions

Two rock climbers are climbing a steep face. The leader is secured by a rope that is held by the attentive second positioned directly below. The rope between them is 10 m long and passes through a carabiner connected to a Friend placed in a crack in the rock 6 m below the leader (fig.9). The mass of the leader's body and gear is 80 kg. The leader, tired after an inspired stretch of climbing, loses her grip and falls off the rock wall. The rope tied to a harness around her waist catches her fall (straight down). The jerk she experiences is called a catching shock and is lessened by the elasticity of the rope.

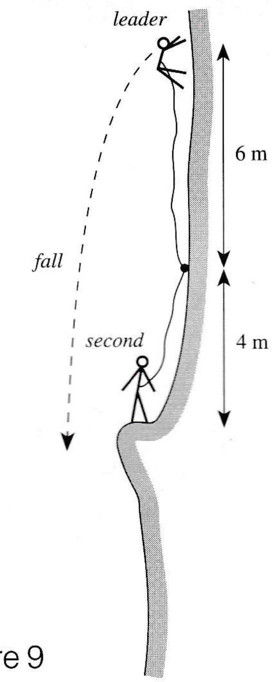


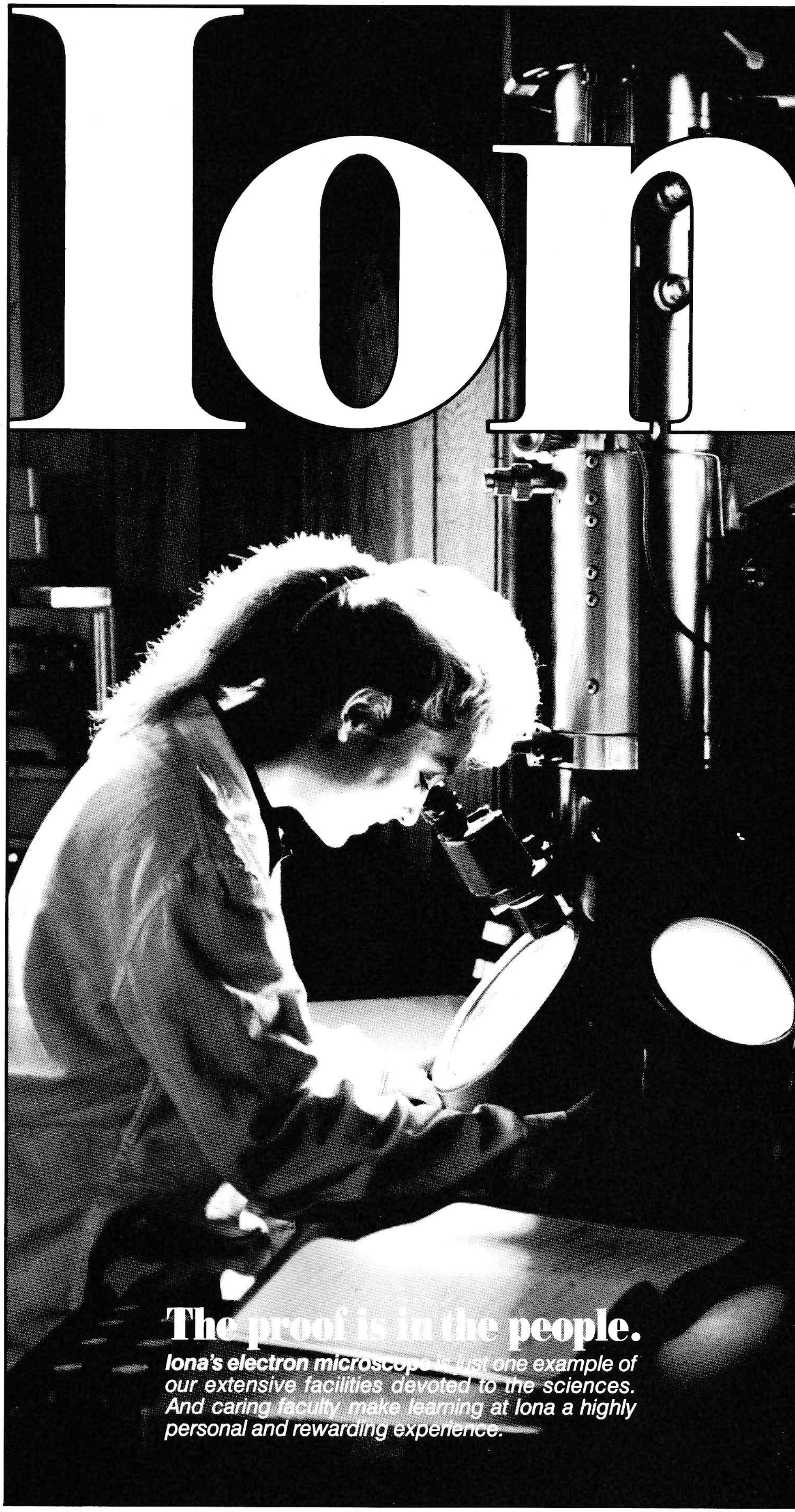
Figure 9

1. What is the fall factor ϕ for this accident?
2. What is the value of the catching shock?
3. What is the deceleration of the leader due to the rope as a fraction of the acceleration due to gravity?
4. How much does the rope stretch?
5. Describe how a fall with a fall factor $\phi = 0$ may occur and calculate the catching shock in this case.

Dr. **John Wylie** *teaches at The Toronto French School and is the director of the Canadian Chemistry and Physics Olympiad.*

ANSWERS ON PAGE 60

Iona



Be your best at Iona.

Our "One-on-One Science Program" pairs selected science majors with faculty mentors in important, original research in such studies as water quality, lead poisoning and computer modeling.

Our science and math programs offer 10 majors —from biochemistry to ecology to physics. Iona has also made a major commitment to math and science studies.

Scholarship and financial aid for qualified students are part of Iona's tradition: making quality education affordable.

National honor societies in the sciences and math are active at Iona. And our honors programs recognize and reward outstanding students, encouraging original research and independent study.

For an application call 1-800-231-IONA.

Iona College
715 North Avenue
New Rochelle, NY 10801-1890

Circle No. 9 on Reader Service Card

Jewels in the crown

The beauty of inductive reasoning

by Mark Saul

IT HAS BEEN SAID THAT mathematics is the queen of the sciences. If this is true, then she must wear a crown. What jewels shall we place in her crown? Mathematics abounds in beautiful results. Rather than selecting any specific beauty, we should find a prominent place in the crown for beauties on a larger, more general scale. Among these might be the method of mathematical induction.

Mathematical induction is a method of proof allowed us by the very definition of the natural numbers: each number is followed by a "next" number. A proof by induction begins with the observation that a certain proposition depends on a variable (we'll call it N) that takes on positive integral values. If this proposition is true when $N = 1$, and if, for every positive integral value of k, the truth of the proposition for $N = k$ implies its truth for $N = k + 1$, then we say that the proposition has been proved by mathematical induction.

It can be shown that the principle of mathematical induction—the fact that such a procedure establishes the truth of a statement—is equivalent to several very basic propositions about the natural numbers. It's amazing how many intricate proofs can be based on this simple elemental principle.

And how does this gem appear in the average high school textbook? Typically, we find a problem like this.

Problem 1. Show that, for any natural number N, $1 + 2 + 3 + \ldots + N = N(N + 1)/2$.

Proof by induction. Part 1:

$$1 = 1(2)/2,$$
$$1 + 2 = 2(3)/2$$
$$1 + 2 + 3 = 3(4)/2.$$

Part 2: Suppose that the proposition is true for

$N = k$; that is, suppose $1 + 2 + 3 + \ldots + k + (k + 1) = k(k + 1)/2$. We can add the term $(k + 1)$ to each side of this equality. We find that

$$\begin{aligned} 1 + 2 + 3 + \ldots + k + (k+1) &= k(k+1)/2 + (k+1) \\ &= (k+1)(k/2 + 1) \\ &= (k+1)[(k+2)/2]. \end{aligned}$$

Since this is exactly the form predicted by the original assertion, the proof by induction is complete.

One practices on more and more complicated problems of this nature. One finds that the "hard part" of the proof lies in establishing that the expression obtained by adding the next term is identical to the prediction of the proposition. This is accomplished by algebraic manipulation, and one quickly gets the idea that mathematical induction is a thinly disguised version of those dull exercises in simplifying algebraic expressions that one must practice but offer no delight or surprise.

And so the bright gem of the mathematical crown is used to cut glass.

This dreary sort of introduction fails to exhibit two key facets of mathematical induction. One of these facets is the element of discovery that an inductive situation can provide. For example, suppose the same problem were posed like this.

Problem 2. The first "triangular" number is 1:

•

The second triangular number is 3:

•

• •

The third triangular number is 6:

•

• •

• • •

Find the 100th triangular number.

In this presentation the problem is still open: the reader must observe a pattern, formulate a hypothesis, and then explain what makes the pattern reproduce itself (the inductive step). For example, we may try to factor the triangular numbers:

$$\begin{aligned} 1 &= 1, \\ 3 &= 1 \times 3, \\ 6 &= 2 \times 3, \\ 10 &= 2 \times 5, \\ 15 &= 3 \times 5, \end{aligned}$$

and note a pattern emerging in the factors. A shrewd guess at the algebraic form of the factors will give the game away, and the routine of induction, as discussed above, takes over. But before this, we have had the joy of discovery, of putting together the pieces of a puzzle. It is this joy that George Polya celebrated in his monumental work *Induction and Analogy in Mathematics*, in which you can find many more such examples.

A second facet of mathematical induction that's concealed by the usual textbook treatment is its generality. Mathematical induction is not simply a technique used to sum series. It can be used in a wide variety of situations and in discussing diverse types of mathematical objects.

The collection of problems offered below can serve as an introduction to induction as discovery and to the many situations in which induction can be applied. I may have omitted your favorite bit of induction. In fact, I have deliberately omitted what is perhaps the greatest inductive tour-de-force in the history of mathematics: Cauchy's proof of the arithmetic mean–geometric mean inequality. This you can look up elsewhere (for example, see Beckenbach and Bellman).[1]

Each example presents an opportunity to make a general conjecture out of an observed pattern. In this sort of induction problem, the first part of the proof ("grounding" the induction in the first few cases) is simply a repetition of the process that generated the initial observations. But since it's part of the formal proof, I feel compelled to mention it each time.

Example 1. Compute the numerical value of

$$100 - (99 - (98 - (97 - (\ldots - (3 - (2 - 1)) \ldots)))).$$

[1]See also "What Did the Conductor Say?" and math challenge M58 in this issue.—*Ed.*

Solution. We note that

$$\begin{aligned} 2 - 1 &= 1, \\ 3 - (2 - 1) &= 2, \\ 4 - (3 - (2 - 1)) &= 2, \\ 5 - (4 - (3 - (2 - 1))) &= 3 \end{aligned}$$

and make a guess.

Conjecture. For any natural number N,

$$\begin{aligned} N - (N - 1 - (N - 2 - (N - 3 - &(\ldots - (3 - (2 - 1)) \ldots)))) \\ &= N/2 \text{ if } N \text{ is even,} \\ &\quad (N+1)/2 \text{ if } N \text{ is odd.} \end{aligned}$$

Proof by induction. Part 1: This part was completed in forming the conjecture.

Part 2: Suppose the proposition is true for $N = k$. That is,

$$\begin{aligned} k - (k - 1 - (k - 2 - (k - 3 - &(\ldots - (3 - (2 - 1)) \ldots)))) \\ &= k/2 \text{ if } k \text{ is even,} \\ &\quad (k+1)/2 \text{ if } k \text{ is odd.} \end{aligned}$$

Then we look at the expression for $N = k + 1$:

$$k + 1 - (k - (k - 1 - (k - 2 - (k - 3 - (\ldots - (3 - (2 - 1)) \ldots))))).$$

If k is even, this is equal to $k + 1 - (k/2) = (k + 2)/2$, which is what the proposition would predict (since $k + 1$ is odd). If k is odd, this is equal to $k + 1 - (k + 1)/2 = (k + 1)/2$, which again accords with the prediction of the proposition.

Another way to formalize this problem is to create a sequence:

$$\begin{aligned} a_1 &= 1, \\ a_2 &= 2 - 1, \\ a_3 &= 3 - (2 - 1), \\ a_4 &= 4 - (3 - (2 - 1)), \end{aligned}$$

and, in general, $a_n = n - a_{n-1}$. This is a *recursive* definition: each term of the sequence (except the first) is defined using the previous term. Problem 2 above is asking for the value of a_{100}. As we'll see, inductive proofs work well with such recursive definitions.

Challenge 1. Compute the numerical value of

$$100^2 - (99^2 - (98^2 - (97^2 - (\ldots - (3^2 - (2^2 - 1^2)) \ldots)))).$$

Art by Sergey Ivanov

Example 2. The Fibonacci numbers are defined as follows:

$$F_1 = F_2 = 1,$$
$$F_n = F_{n-1} + F_{n-2} \text{ for } n > 2.$$

Find all the *even* Fibonacci numbers.

Solution. Let's calculate the first ten Fibonacci numbers:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55.$$

By observation, we find that the subscripts of the even Fibonacci numbers are exactly the multiples of 3.

Conjecture. F_n is even if and only if n is a multiple of 3.

Proof by induction. Part 1: This is established above, for $n = 1, 2, 3$.

Part 2: Suppose we know that the conjecture is true for every value of N less than $3k$. I'll show that it's also true for every value of N less than $3k + 3$. By the induction hypothesis F_{3k-2} and F_{3k-1} are both odd. So their sum, which is F_{3k+3}, is even. Also by the induction hypothesis F_{3k} is even, while F_{3k-1} is odd. So their sum, which is F_{3k+1}, must be odd. The same type of argument will show that F_{3k+2} is odd.

This proof displays two variants of the principle of mathematical induction. First, we include in the induction hypothesis not only the assumption that the proposition is true for the case $N = k$ but also the assumption that it's true for certain natural number less than or equal to k. It turns out that this stronger assumption, even if it includes *all* the natural numbers less than or equal to k, also allows us to proceed inductively.

The second variant in this proof is the fact that the induction proceeds in a "skip-step" fashion: we prove the conjecture for F_{3k+1}, F_{3k+2}, and F_{3k+3} separately. Together, the three cases will cover all the positive integers.

Challenge 2. Which Fibonacci numbers are multiples of 5? Which are multiples of 7?

Challenge 3. Make a general conjecture about which Fibonacci numbers are divisible by a given number d. Test your conjecture for $n = 4, 6, 8, 9, 10, 11$. How general a statement can you make about when d divides F_n?

Challenge 4. What would happen if we changed the first two terms of the Fibonacci sequence but left the rule of formation the same? That is, suppose we define the "Gibonacci numbers" as follows:

$$g_1 = 3,$$
$$g_2 = -2,$$
$$g_n = g_{n-1} + g_{n-2} \text{ for } n > 2.$$

Or suppose we chose two other numbers for the first two terms of the sequence, then proceeded "fibonaccily" to form the rest of the sequence.

How many of the observations given above still hold?

Example 3. Let $f(x) = 2x + 1$. Compute $f(f(x))$, $f(f(f(x)))$, and $f(f(f(f(x))))$. Each of these functions is of the form $ax + b$. What do you notice about the coefficients? Does the pattern continue?

Solution. By direct computation we find that

$$f(f(x)) = 4x + 3,$$
$$f(f(f(x))) = 8x + 7,$$
$$f(f(f(f(x)))) = 16x + 15.$$

Conjecture. If $f_1(x) = 2x + 1$ and $f_n(x) = f_1(f_{n-1}(x))$, then $f_n(x) = 2^n x + 2^n - 1$.

Proof by induction. Part 1: This is demonstrated above.

Part 2: Suppose $f_n = 2^n x + 2^n - 1$. Then $f_{n+1}(x) = f_1(f_n(x)) = 2(2^n x + 2^n - 1) + 1 = 2^{n+1}x + 2^{n+1} - 2 + 1 = 2^{n+1}x + 2^{n+1} - 1$, which is what the conjecture predicted.

Challenge 5. Define

$$f_1(x) = \frac{4x-3}{2x+1}$$

and form the functions f_2, f_3, ... as before. These are now somewhat more complicated since they depend on four real numbers, not just two. Can you give a rule of formation for the four sequences of coefficients now? (Hint: if you know about matrix multiplication, compare the first few terms of the series with the powers of the matrix $\begin{pmatrix} 4 & -3 \\ 2 & 1 \end{pmatrix}$ using ordinary matrix multiplication.)

Example 4. In the kingdom of Aveugle, the Bourgnes are kings. That is, King Bourgne XXIII (the twenty-third of that name) was preceded by King Bourgne XXII, who was preceded by Bourgne XXI. In fact, all the 23 kings of the present dynasty had the same name. Before that, they also had the name Bourgne, but they didn't keep count of their number. The Bourgnes have been kings in the kingdom of Aveugle since the late Jurassic.

The first King Bourgne (he had already lost track of his number) had a box full of diamonds. When he died, one diamond was given to his favorite servant, exactly half the remaining diamonds were sold to help the poor, and the rest became the property of the king's (unique) heir, who became the new King Bourgne. This established a tradition. Over the years, the remaining diamonds were never touched except when King Bourgne N died. When this happened, one diamond was given to a servant and half of the remaining diamonds were donated to charity. The rest of the diamonds became the property of King Bourgne $N + 1$. This tradition has continued down to the present. If King Bourgne XXIII has exactly one diamond, how many diamonds did Bourgne I have? [Wisconsin Mathematics, Engineering, and Science Talent Search (1982)]

Solution. If Bourgne XXIII has one diamond, then Bourgne XXII must have had $1 + 1 + 1 = 3$ diamonds; Bourgne XXI must have had $3 + 3 + 1 = 7$ diamonds; and Bourgne XX had $7 + 7 + 1 = 15$ diamonds.

Conjecture. N Bourgnes ago, the king had $2^{n+1} - 1$ diamonds.

Proof by induction. Part 1: This portion of the proof is demonstrated in the calculations above.

Part 2: Suppose that k Bourgnes ago, the king had $2^{k+1} - 1$ diamonds. How did he get his diamonds? The king before him must have had $2(2^{k+1} - 1) + 1 = 2^{k+2} - 1$ diamonds. This is exactly what the conjecture would predict.

In the specific case described, the first king is 20 kings before the present one, so he had $2^{21} - 1$ diamonds.

Challenge 6. What can you say about the number of diamonds owned by King Bourgne I if the present King Bourgne (whose number is forgotten) has r diamonds left?

Example 5. A farmer has harvested a number of apples. He is paid for them in the following peculiar fashion. He is allowed to divide the apples into two piles, in any way he chooses. The number of apples in each pile is counted and the two numbers are multiplied together. This product is converted into dollars, and the farmer receives this number of dollars as a partial payment.

The farmer is then allowed to choose one of the two piles he has just created and separate this pile into two smaller piles. Again, these two new, smaller piles are counted and the numbers are multiplied together. This product is converted into dollars, and the farmer receives this number of dollars as continued payment for his apples.

Now there are three piles of apples. The farmer is allowed to choose one of the piles and separate it into two piles. These are counted, the product is taken, and the farmer receives this number of dollars as continued payment.

This elaborate process is continued until the piles each have a single apple. In particular, any pile of two apples must eventually be separated into two "piles" of one apple each. This final separation of the two apples will contribute $1 \times 1 = 1$ dollar to the farmer's payment.

If the farmer has 100 apples, how should he perform the divisions to get the best price for his apples? What is that price? [*Kvant*, problem M100 (1987, No. 1)]

Solution. Trying a few simple cases (two or three apples), we find that it doesn't matter how the subdivision is done and that the payment is always $n(n+1)/2$ for n apples (the nth "triangular" number).

Conjecture. The payment for n apples is always the nth triangular number.

Proof by induction. Part 1: This portion is left to the reader.

Part 2: We choose a "strong" induction hypothesis, as in example 2. Suppose that the payment for d apples is $d(d+1)/2$, independent of the mode of division, for every number $d < k$. We take $k + 1$ apple and divide them into two piles in any way at all. Suppose these piles have a apples and b apples. Then $a + b = k$, and both a and b are less than k. So the induction hypothesis applies to the division of each of these piles. From subdividing these piles, we get $a(a+1)/2$ and $b(b+1)/2$ dollars. We also get ab dollars from the product of the first two piles. This gives us a total of $a(a+1)/2 + b(b+1)/2 + ab$ dollars for the payment. We then have

$$\begin{aligned} &a(a+1)/2 + b(b+1)/2 + ab \\ &\quad = (a^2 + a + b^2 + b + 2ab)/2 \\ &\quad = [(a^2 + 2ab + b^2) + (a+b)]/2 \\ &\quad = (a+b)(a+b+1)/2 \\ &\quad = k(k+1)/2, \end{aligned}$$

as predicted by the conjecture.

This proof demonstrates the power of a well-chosen induction hypothesis.

Example 6. Three lines that are not coincident (that is, do not pass through the same point) and no two of which are parallel divide a plane into 7 regions. Into how many regions do n lines divide a plane if they are in general position (no two are parallel and no three are coincident)?

Solution. We find the following values:

Lines	Regions
1	2
2	4
3	7
4	11

It isn't hard to conjecture that the number of regions created by n planes is given by $(n^2 + n + 2)/2$ regions—for example, by comparing this sequence with the "triangular" numbers discussed above.

To prove this, we assume that any k lines in general position divide a plane into $(k^2 + k + 2)/2$ regions and see what happens when we add a new line. Clearly, the new line intersects each of the k old lines once (to be in general position) and so is divided into $k + 1$ parts. Each of these parts divides one of the old regions in two, thus creating a new region. So there will be $k + 1$ new regions. We have now reduced the problem to an algebraic identity:

$$\begin{aligned} &(k^2 + k + 2)/2 + (k+1) \\ &\quad = [(k+1)^2 + (k+1) + 2]/2, \end{aligned}$$

which is easy to verify.

Challenge 7. A set of three circles is such that every pair of circles intersects in two distinct points. Into how many regions do these three circles divide the plane? Into how many regions do n circles divide the plane if every pair of them intersects in two distinct points?

Challenge 8. The same as challenge 7, but replace the circles with triangles.

Challenge 9. The same as the previous challenge, but replace the triangles with parabolas.

Example 7. Find a formula for the sum of the first n Fibonacci numbers.

Solution. We note that

$$\begin{aligned} 1 &= 1, \\ 1 + 1 &= 2, \\ 1 + 1 + 2 &= 4, \\ 1 + 1 + 2 + 3 &= 7, \\ 1 + 1 + 2 + 3 + 5 &= 12, \end{aligned}$$

and note that each result is one less than a Fibonacci number.

Conjecture. $F_1 + F_2 + \ldots + F_n = F_{n+2} - 1$.

The first few cases are already established. To complete the induction, we add F_{n+1} to each side of the equation above. On the right we get $F_{n+1} + F_{n+2} - 1$, which is equal to $F_{n+3} - 1$ by the definition of F_{n+3}. The induction is complete.

Challenge 10. What is the sum of the first n Gibonacci numbers? (See example 2 and challenge 4.)

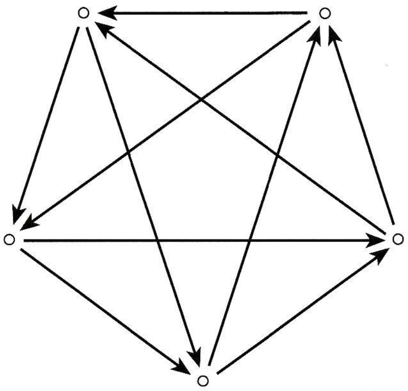


Figure 1

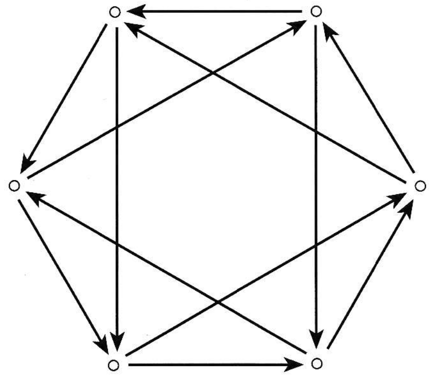


Figure 2

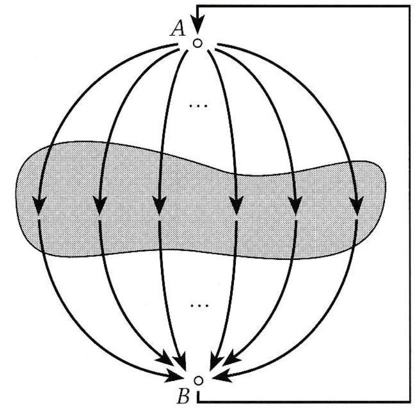


Figure 3

Challenge 11. Find a formula for the sum of the *squares* of the first n Fibonacci numbers, or a more general Fibonacci-type sequence. Prove that your formula is valid.

Example 8. Starting with the set $S_{10} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, form as many subsets as possible, subject to the restriction that no two elements of any subset differ by 1. For example, the subset $\{1, 3, 5\}$ is allowed, but not the subset $\{1, 3, 4\}$. In particular, the "singleton" subsets are all allowed, as is the null set (since no two of their elements differ by 1). How many such subsets are there? [Canadian Mathematical Olympiad (1985)]

Solution. We choose to answer a more general question. How may such subsets of the set $\{1, 2, 3, ..., n\}$ are there? Call this number S_n. Then, by direct count, $S_1 = 2$, $S_2 = 3$, $S_4 = 5$, $S_5 = 8$. It looks like the S_n's are really the F_n's (the Fibonacci numbers) in disguise. But why?

It's not hard to show that the sequence S_n will satisfy the Fibonacci recursion. Suppose we knew S_8 and S_9. How would we count S_{10}? Surely any "good" subset counted by S_8 is still counted by S_{10}, and this will account for all the good subsets that don't include the number 10. How many subsets are there that *do* include 10? Well, if we cross out the number 10 in each such subset, we'll have a good subset counted by S_8 because the initial subset (which contained 10) didn't contain 9—it was "good." So S_8 counts just those good subsets of S_{10} that *include* the new element 10. Therefore, $S_8 + S_9 = S_{10}$ can be quickly made general, and S_{10} turns out to be 144.

Challenge 12. Pierre Le Fou, the chef at Le Quincaillerie, knows how to prepare only two dishes: porcupine *en colere* and koala *bonne femme*. He prepares one of these dishes every evening. House rules allow him to serve the same dish twice in a row, but never three times in a row.

Pierre is charged with planning the menus for five straight days. How many such menus can he plan? How many menus can he plan for n days in a row?

Example 9. A rural county contains several towns connected by one-lane roads. To avoid head-on collisions, the county government has decided that every road can be traveled in only one direction. To save money, they decide that no two towns should be connected by more than one road. To save road signs, the roads intersect only at the towns they connect: wherever two roads cross, the county has built overpasses, so that you can "change"roads only at a town. In addition, roads must be built so that one can get from any town to any other town, passing through at most one more town.

1. Show how the required roads can be built for a county with three towns.

2. Show that the required roads *cannot* be built if the county has two or four towns.

3. Show that the required roads can be built if the county contains any (natural) number of towns greater than four.

Solution. See figure 1 for $n = 5$ and figure 2 for $n = 6$.

Using mathematical induction, we assume that the required network can be drawn for any county of k towns. Figure 3 illustrates the situation for a county with $k + 2$ towns. We can show that the required network exists by designating two of these as "new" towns (here they are labeled A and B) and thinking of the other k towns as "old" towns. By the induction hypothesis, the old towns can be connected by a network as required. Suppose this network has been drawn in, under the shaded "cloud" in the figure. We can include the two new towns by connecting new town A to each old town and each old town to new town B. A final connection of B to A completes the network. We have shown that if the required network can be drawn for any k towns, then it can be drawn for any $k + 2$ towns. Since we have proved the assertion for two consecutive integers, this completes the induction.

Additional reading

Beckenbach, Edwin, and Richard Bellman. *An Introduction to Inequalities*. Washington: Mathematical Association of America, 1961.

Golovina, L. I. and I. M. Yaglom. *Induction in Geometry*. Moscow: Mir Publishers, 1979.

Niven, Ivan. *Mathematics of Choice: How to Count Without Counting*. Washington: Mathematical Association of America, 1964.

Polya, George. *Induction and Analogy in Mathematics* (volume 1 of *Mathematics and Plausible Reasoning*). Princeton: Princeton University Press, 1954.

Rosenbaum, Louise Johnson. *Induction in Mathematics*. New York: Houghton Mifflin Company, 1966.

Sominsky, I. S. *The Method of Mathematical Induction*. Moscow: Mir Publishers, 1983.

MATHEMATICAL SURPRISES

Fibonacci strikes again!

And again, and again . . .

by Elliott Ostler and Neal Grandgenett

NUMBER SEQUENCES HAVE long provided mathematicians with thought-provoking problems and interesting applications to the real world. One particular sequence, the Fibonacci series, is especially interesting and powerful in its mathematical applications. After a short brainstorming session, we decided to write down just a few of our favorite applications of this amazing number sequence.

The Fibonacci series is easily produced by starting with 1 and adding the last two numbers to produce the next number in the sequence: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, … . The series was first discovered by Leonardo of Pisa (better known as Fibonacci) in the early 1200s. It was given as a solution to a now famous problem posed by Fibonacci, commonly known as the "rabbit problem." It reads:

> How many pairs of rabbits will be produced in a year, beginning with a single pair, if in every month each pair bears a new pair that becomes reproductive from the second month on?

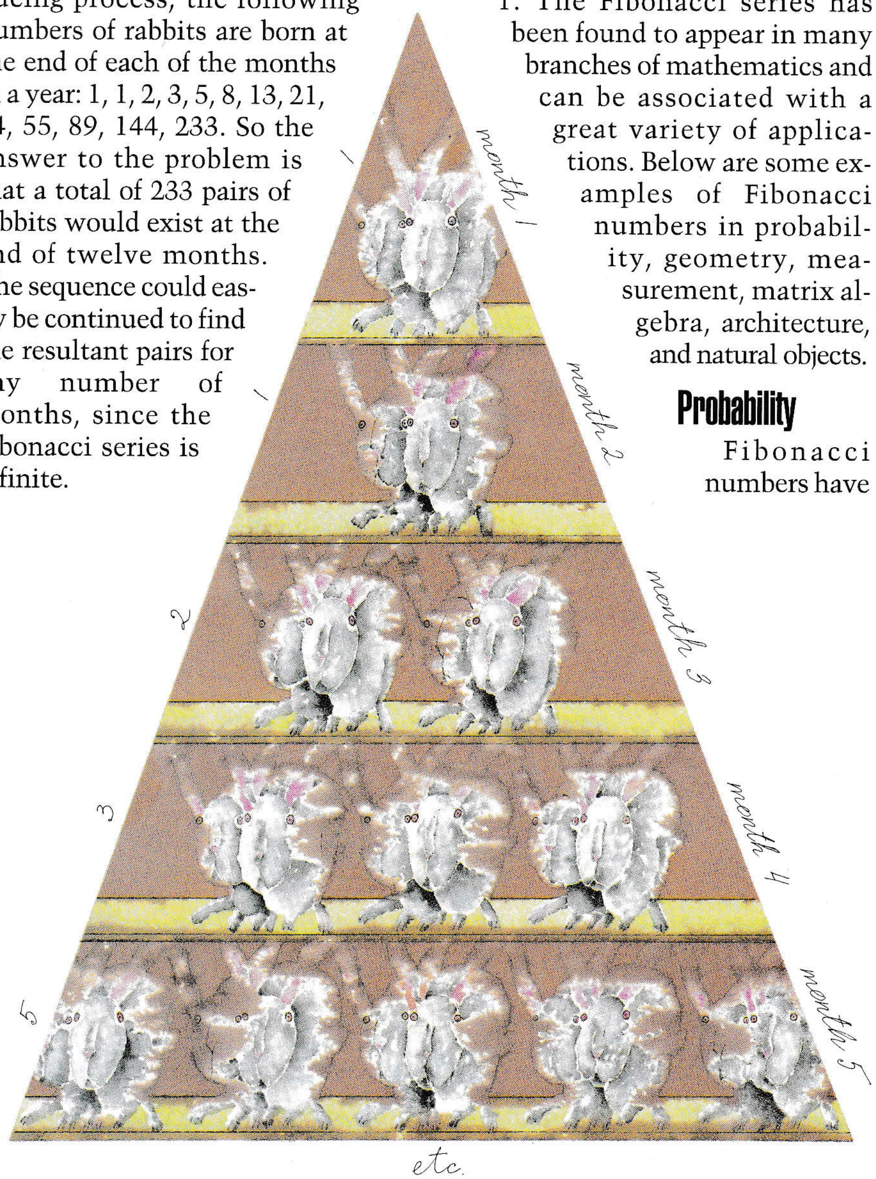


Figure 1

Fibonacci rabbit propagation. We start with one pair of rabbits at the beginning of the first month. At the end of month 1 there is still only one pair, since the rabbits are not reproductive until the second month. At the end of month 2 there are now two pairs, since the first pair has become reproductive, while the second pair is not yet reproductive. The figure could be continued indefinitely.

Art by Sergey Ivanov

The first few months of the solution are shown in figure 1. Notice how the pairs present at each month correspond to the Fibonacci series.

By continuing the rabbit-producing process, the following numbers of rabbits are born at the end of each of the months in a year: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233. So the answer to the problem is that a total of 233 pairs of rabbits would exist at the end of twelve months. The sequence could easily be continued to find the resultant pairs for any number of months, since the Fibonacci series is infinite.

The functional notation for the series was first noted by Johannes Kepler in 1611. It was given as $f_n + f_{n+1} = f_{n+2}$, where f_n is the nth Fibonacci number, $f_0 = 0$, and $f_1 = 1$. The Fibonacci series has been found to appear in many branches of mathematics and can be associated with a great variety of applications. Below are some examples of Fibonacci numbers in probability, geometry, measurement, matrix algebra, architecture, and natural objects.

Probability

Fibonacci numbers have

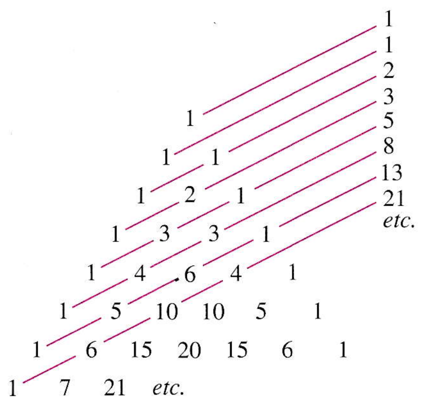


Figure 2
The Fibonacci series from the diagonals of Pascal's triangle.

been found to appear in a variety of probability and combinatorial problems. Many of the classic problems in these areas are binomial in nature, such as coin-toss problems. The distributions of the solutions to these problems typically relate to the binomial theorem and can be derived from the coefficients in the expansion of Pascal's triangle. Close observation of this triangle shows that the Fibonacci series can be obtained by summing the elements of the diagonals, as illustrated in figure 2.

The fact that this sequence is generated in the diagonals of Pascal's triangle indicates that it may also be found in the solutions to problems related to binomial distributions. Here is one such problem:

> If a coin is flipped *N* times and the resulting patterns are recorded in order as the coin is flipped, how many unique sequences of heads and tails are possible if the coin can never land on heads two or more times in a row?

Table 1 lists all possible sequences for each given number of flips. The Fibonacci series can be seen as the number of appropriate sequences at each of the consecutive number of flips. Using this continuing pattern, the number of unique sequences for any number of flips can be determined.

Geometry

Another area in which the Fibonacci series commonly appears is geometry. In the following example, the four diagrams illustrate how the Fibonacci numbers appear in increasingly larger five-point stars constructed from regular pentagons. When the length of the side of the initial pentagon is a Fibonacci number of 5 or greater, approximates of Fibonacci numbers are generated as the stars are constructed. Figure 3 illustrates this process.

In figure 3a, a triangle is drawn between points *A*, *B*, and *C*. It can be shown that the length of segments *AC* and *AB* is approximately eight. An identical triangle can then be drawn between points *B*, *C*, and *D*, creating the first point of the star. Triangles *ABC* and *ABD* are congruent and symmetric with respect to segment *BC*. The same process is then followed for each of the remaining sides of the original pentagon, producing figure 3b.

Table 1

Number of flips	Sequences without consecutive heads	Total
1	(H), (T)	2
2	(HT), (TH), (TT)	3
3	(HTT), (THT), (TTH), (HTH), (TTT)	5
4	(HTTT), (THTT), (TTHT), (TTTH), (HTHT), (HTTH), (THTH), (TTTT)	8
5	(HTTTT), (THTTT), (TTHTT), (TTTHT), (TTTTH), (HTHTT), (HTTHT), (HTTTH), (THTHT), (THTTH), (TTHTH), (HTHTH), (TTTTT)	13

At this point, you can see that all of the resulting measurements approximate Fibonacci numbers. Figure 3c shows the points of the star joined to make a new pentagon. The measurements approximate Fibonacci numbers, but are not exact, as shown in figure 3d. Repeating the process from figure 3a, one can construct a new star from the new pentagon shown in figure 3c. This process will infinitely generate stars and pentagons that more and more closely approximate Fibonacci numbers as measurements, provided that the original pentagon had a Fibonacci number of at least 5 as the length of its sides.

Measurement

In the area of measurement, the Fibonacci series can be used as a quick conversion table for changing between kilometers and miles. The series can be used to estimate the conversion of miles to kilometers by using any pair of consecutive numbers in the series. For example, 3 miles is roughly 5 kilometers; 5 miles is roughly 8 kilometers; 8 miles is roughly 13 kilometers; and so on. This conversion table works because of the similarity in value between the ratio of two consecutive Fibonacci numbers (1.618 as the numbers increase) and the general conversion

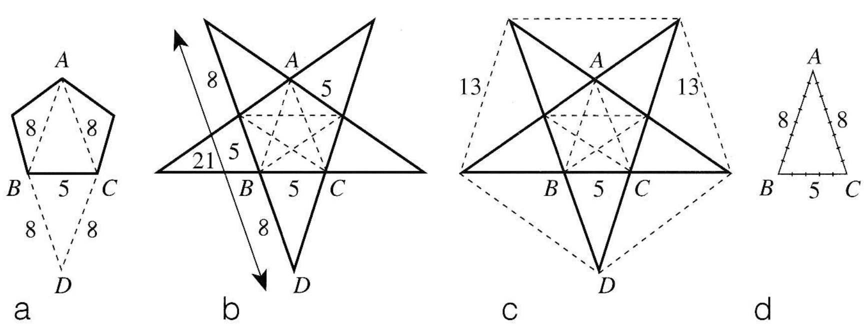


Figure 3
The Fibonacci series in the measurements of recursive pentagons and stars. Note that the relationship is only an approximate one. Since angle ABC = 72° and AB = 8, cos 72° should equal the ratio of 2.5 : 8 (0.3125). In actuality, this ratio only approximates the cosine of 72°, which is 0.309017. If larger Fibonacci numbers are used, such as 89 and 144, the ratio obtained (0.309028) more closely approximates cos 72°.

Table 2

kilometers	1	1	2	3	5	8	13	21	34	55	...	
miles		1	1	2	3	5	8	13	21	34	...	
ratio		1	2	1.5	1.66	1.6	1.625	1.615	1.619	1.617	...	1.618

factor for changing miles to kilometers (1.609). Since the conversion factor of miles to kilometers has a value that's only 0.009 different from the Fibonacci ratio of 1.618, the Fibonacci series can act as a quick and reasonably accurate conversion table for miles to kilometers (see table 2).

Architecture

The ratio observed between consecutive Fibonacci numbers isn't just useful as a mere conversion table. This ratio is a famous number that's commonly referred to as the "golden ratio" or "golden section" and has greatly influenced architecture and art through the ages.[1] This ratio, often associated with the so-called "golden rectangle," has intrigued scholars for many years. The Parthenon in Athens fits within the golden rectangle—its dimensions are related to the golden ratio (fig. 4). This ratio is found in many other ancient and modern buildings as well.

Matrix algebra

On a more theoretical level, the Fibonacci series can be seen in certain problems in linear algebra. For instance, the sequence is generated by using a standard 2×2 matrix that's raised to the power of N. This generating matrix consists of three identical nonzero elements in the first three positions and a zero in the last position:

$$F = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

All the elements produced will be Fibonacci numbers. Here is the generating matrix raised to the fifth power:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} (1+1) & (1+0) \\ (1+0) & (1+0) \end{pmatrix} = \begin{pmatrix} \mathbf{2} & 1 \\ 1 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} (2+1) & (2+0) \\ (1+1) & (1+0) \end{pmatrix} = \begin{pmatrix} \mathbf{3} & 2 \\ 2 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} (3+2) & (3+0) \\ (2+1) & (2+0) \end{pmatrix} = \begin{pmatrix} \mathbf{5} & 3 \\ 3 & 2 \end{pmatrix},$$

$$\begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} (5+3) & (5+0) \\ (3+2) & (3+0) \end{pmatrix} = \begin{pmatrix} \mathbf{8} & 5 \\ 5 & 3 \end{pmatrix}.$$

Notice that all the numbers produced in each resulting matrix are Fibonacci numbers, but in particular the top left element (printed in bold) gives each consecutive term of the Fibonacci series.

Nature

In nature, Fibonacci numbers have been found to be directly associated with the natural spirals in objects such as pine cones, pineapples, and daisy blossoms. Figure 5 illustrates how these numbers appear in the head of most daisy blossoms. The individual florets of the blossom create two spirals—one clockwise set and one counterclockwise set. The clockwise set contains 21 spirals—a Fibonacci number; and the counterclockwise set contains 34 spirals—the next consecutive Fibonacci number.

This double spiraling pattern with a ratio of 21 : 34 representing consecutive Fibonacci numbers also appears in pine cones, where the ratio is 5 : 8, and in pineapples, where the ratio is 8 : 13. The Fibonacci series occurs in many other plant patterns, such as the leaves of a cherry tree, the petals of a tulip, and the long branches of a willow.

In this article we've explored just a few of the applications of the Fibonacci series in mathematics. Its appearance in the world of art is no less impressive. This series, with its associated golden ratio, can be found in a wide variety of classic works—for instance, in paintings by Leonardo da Vinci. With a little research one can become truly fascinated by the diversity of its manifestations. The mathematical applications alone are so varied and extensive that an entire journal is devoted to them: *The Fibonacci Quarterly*.

Why does the Fibonacci series occur so often in our world?

CONTINUED ON PAGE 30

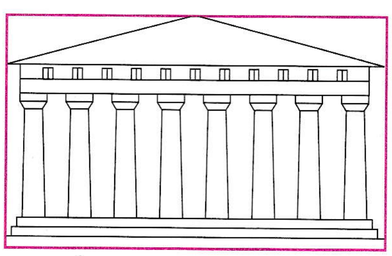
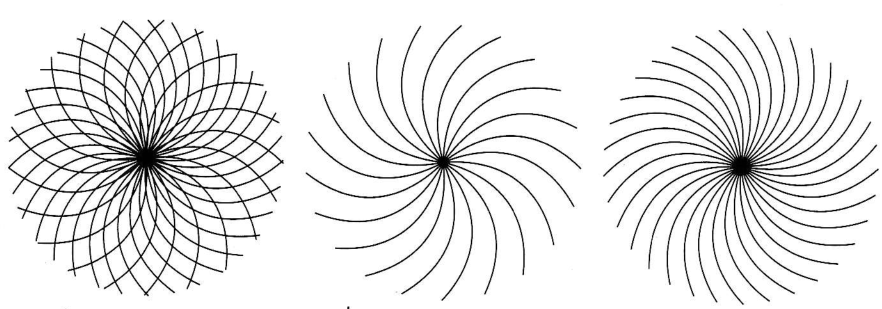


Figure 4
The Parthenon at Athens inside a rectangle whose dimensions match the golden ratio.



a b c

Figure 5
Fibonacci numbers occurring in the 21 : 34 double spiraling pattern of a daisy blossom. The daisy's spiraling pattern (a) usually consists of 21 clockwise spirals (b) and 34 counterclockwise spirals (c).

---

[1]See the Kaleidoscope and Gallery Q in the January/February 1991 issue of *Quantum*.—*Ed.*

The world according to Malthus and Volterra

The mathematical theory of the struggle for existence

by Constantine Bogdanov

ATTEMPTS AT A MATHEmatical description of the dynamics of densities of certain biological populations and associations have a remarkable history. One of the first models of the dynamics of population growth belongs to Thomas Robert Malthus (1766–1834), an English cleric and economist.

In *An Essay on the Principle of Population* (1798), Malthus stated that in human society, and in the living world as a whole, there exists an absolute law of limitless reproduction of individuals. Furthermore, the population of the Earth grows in a geometric progression, while the means of subsistence increase arithmetically.

Generalizing from the role of biological factors in the reproduction of a population, Malthus painted the severe consequences of the law of population he discovered. He thought that human society too often interferes with the workings of nature, and he argued for the abolition of laws in England that provide for the sustenance of the poor. "After the public notice which I have proposed had been given, and the system of poor-laws had ceased with regard to the rising generation," Malthus wrote in the revised edition of his *Essay*, "if a man chose to marry, without a prospect of being able to support a family, he should have the most perfect liberty so to do. Though to marry, in this case, is, in my opinion, clearly an immoral act, yet it is not one which society can justly take upon itself to prevent or punish; because the punishment provided for it by the laws of nature falls directly and most severely upon the individual who commits the act, and through him, only more remotely and feebly, on the society. When nature shall govern and punish for us, it is a very miserable ambition to wish to snatch the rod from her hand and draw upon ourselves the odium of executioner." In Malthus's view, such a person "erred in the face of the most clear and precise warning" and has no right to complain to anyone—he has "no claim of *right* on society for the smallest portion of food, beyond that which his labor would fairly purchase."

In its mathematical form, the Malthusian model is quite simple. Let $N(t)$ be the total number of a population at time t. According to Malthus, the growth rate of the population is directly proportional to the population, or

$$dN/dt = aN,$$

where a is the difference between the birth rate and the death rate. Integrating this equation, we get

$$N(t) = N(0)e^{at},$$

where $N(0)$ is the population density at time $t = 0$. It's evident that Malthus's model with $a > 0$ gives an infinite growth of population, which is never observed in natural populations, where the resources providing for the growth are always limited.

Subsequent research has shown that the population changes in the plant and animal kingdoms cannot be described by Malthus's law; rather, the reproduction of each species changes so that the species survives in the process of evolution.

The first success of mathematical ecology—the science of the relationships among living organisms and the associations they create among themselves and with the environment—was the Volterra–Lotke model proposed by Vito Volterra in his book *The Mathematical Theory of the Struggle for Existence* (1931). The biography of this scientist, known for his classic studies in integral calculus and functional analysis, is interesting in its own right. In many respects it exemplifies the title of his famous book.

Vito Volterra was born in Italy in 1860. His father died when Vito was two, and the family was left with practically no means of sustenance. Yet, in spite of the difficulties, Vito

Art by Leonid Tishkov

succeeded in obtaining an education.

He had learned differential calculus while still a teenager; not knowing integral calculus, he discovered it on his own. He graduated with honors from the science department of Florence University. Volterra quickly became world-famous for his publications in various fields of pure mathematics. But he was also interested in various problems of applied mathematics.

In 1925, in conversations with Umberto D'Ancona, a young zoologist, he learned a curious fact from the statistics of Adriatic fish markets. It turned out that during World War I and immediately after, when the fishing trade had been sharply curtailed, the relative number of predatory fish in the catch had risen. To explain this, Volterra proposed a mathematical model describing the relations between predator and prey and the changes in their populations over time. Mathematical ecology became his main interest, and he devoted himself to this subject for the rest of his life.

Volterra's personality combined the talent of a researcher and the temperament of a political activist. In 1905 he was the youngest senator in the Kingdom of Italy. Expressing progressive views, he actively opposed fascism and was the only senator to vote against the transfer of power to Mussolini in 1922. After that, he became a political émigré in France. In an attempt to enhance the image of the fascist dictatorship, Mussolini invited Volterra to return to Italy, promising honors and titles. But the scientist refused to return, offering an enduring example of honesty and principle in political life. Volterra died in 1940.

Predator and prey

One portion of Volterra's mathematical ecology is devoted to analyzing the "interrelations" of predator and prey. I think you'll find it interesting to learn how Volterra himself solved this problem. Then we'll try to solve the same problem "without even thinking"—by means of a computer. So—let's get started!

Let there be two types of animals, one of which feeds on the other (predator and prey). Under this condition, the relative growth of the number of prey living in isolation (in the absence of predators) per unit time equals e_1; at the same time, the predators, living apart from their prey, gradually die of starvation, and the relative decrease in their population per unit time equals e_2.

As soon as the predators and the prey begin to live in close contact with one another, the changes in their populations become interdependent. In this case, the rate of growth of the prey population will depend on the size of the predator population and will decrease as that population increases. The inverse relation will occur when there is a relative increase in the predator population; the increase can be considered proportional to the prey population. All this can be written as

$$dN_1/dt = N_1(e_1 - a_1N_2),$$
$$dN_2/dt = -N_2(e_2 - a_2N_1), \qquad (1)$$

where N_1 and N_2 are the numbers of prey and predators, respectively, at the time t; a_1 and a_2 are constants.

Unfortunately, it's impossible to solve the system of equations (1)—that is, to find an analytic expression for $N_1(t)$ and $N_2(t)$. But we shouldn't despair, since we can always make the problem just a little bit easier. Take a close look at system (1) and you'll have no trouble finding one of the solutions of the system—the stationary one.

If we assume that the numbers of predators and prey don't change over time, the left sides of (1) become zero, while the right sides tell us that such an equilibrium is possible only if $N_1 = e_2/a_2$ and $N_2 = e_1/a_1$. So we've found one of the solutions of system (1).

Now let's assume that the predator–prey system has somehow gotten close to equilibrium and the predator and prey populations deviate only slightly from the corresponding stationary values. Let $n = N_1 - e_2/a_2$, $x = N_2 - e_1/a_1$; then, after substituting n and x for N_1 and N_2 in system (1) and ignoring nx in comparison with the other terms, we get

$$dn/dt = -xa_1e_2/a_2,$$
$$dx/dt = ne_1a_2/a_1. \qquad (2)$$

Let's introduce a new variable $v = na_2e_1/a_1$. After substitution, system (2) will be changed into

$$dv/dt = -e_1e_2x,$$
$$dx/dt = v. \qquad (3)$$

But now let's recall the system of equations describing the motion of a mass on a spring. Let x be the displacement of the center of mass from the equilibrium position and v be its velocity. Of course these equations can describe the motion of such a mass on a spring if e_1e_2 is equal to the ratio of the spring constant to the mass. This means that the solutions of our system of equations coincide with the solution of the "textbook problem" for the oscillation of a mass on a spring.

The coincidence of the equations describing the oscillation of a mass on a spring and the numbers of individuals in the predator–prey system enables us to state that the numbers of predators and prey should oscillate with a period of $2\pi/\sqrt{e_1e_2}$. In addition, it's known that the oscillation of the velocity of a mass on a spring leads the oscillation of its coordinate by one fourth of the period. Therefore, the oscillations of the prey population must also lead the oscillations of the predator population by one fourth of the period.

So the solutions of the Volterra–Lotke system of equations are oscillations of predator and prey populations, shifted in phase relative to one another, with a period of $2\pi/\sqrt{e_1e_2}$. Of course, when the amplitude of these oscillations increases, they stop being sinusoidal; their period, however, stays nearly the same.

You must admit, though, that the predator–prey system is unlikely to serve as such an unfading generator of oscillations! Do you think perhaps that modeling the relations between the predator and prey by means of the system of equations (1) greatly oversimplifies the situation?

Let's forget the equations

Indeed, let's forget the equations. Let's imagine that we have a hypothetical two-dimensional ocean cut into equal squares by perpendicular lines (fig. 1). Our ocean is inhabited by only two species: harmless mackerels and sharks that devour them. In addition, at each intersection of the lines (node) there can be only one of these species at a given time or none of them. Now I'll describe the behavior of the animals I've put into our ocean.

1. Mackerels and sharks can swim, moving from one node to an adjacent node per unit time. A mackerel moves in a probabilistic way to one of the unoccupied adjacent nodes. A shark, on the other hand, first determines whether there is a mackerel next to it; if there is, the shark swims

to that node and devours the mackerel. If there are no mackerels at the adjacent nodes, the shark swims in a probabilistic way to any of the adjacent nodes.

2. The sharks and mackerels "mature," and their age increases by one unit after each cycle takes place. (What this cycle is will be explained a bit later.) Upon reaching maturity (T_m for a mackerel and T_s for a shark), each fish brings one offspring into the world, and the next offspring can be born only after a time interval equal to the age of maturity. A newborn is initially located at any of the nodes neighboring on the mother's node; after that it obeys all the same laws as the other fish.

3. If a shark has caught no mackerels during a certain number S of consecutive periods, it dies of starvation. A mackerel in our ocean can die only in the mouth of a shark, since it feeds on plankton that is always available in excess.

4. The ocean is rectangular and finite; a fish that happens to be near the shore never throws itself ashore, while any fish that wants to do so out of despair appears immediately on the opposite side of the ocean. In other words, our ocean covers the surface of a toroidal planet.

So the conditions of life for the ocean inhabitants have been defined. And now life begins! At random we (1) distribute sharks and mackerels throughout the ocean, (2) set the age of each fish, and (3) define how long each shark can live without eating a mackerel before it dies of starvation. Of course, all this will be done by a computer, which will keep track of life in our imaginary ocean.

The first cycle of ocean life begins. Let the mackerels first move one step and, if the time has come, breed; then the sharks begin to hunt. At the end of the first cycle we'll update our numbers, subtracting the sharks that have died of starvation and the mackerels that have been eaten, and adding the newborn fish. After that, the next cycle can begin; and so on. As a result, we (that is, the computer) can follow how the shark and mackerel populations in the ocean change as time passes.

Figure 2 shows the results of such a computer model for different values of T_s and T_m (the values of S, as well as the initial shark and mackerel populations, were constant and equal to 5, 20, and 200, respectively). We can see that the numbers of sharks and mackerels in the ocean oscillate with a certain frequency, and that the maximum population of mackerels always occurs slightly before that of the sharks.

Also, by analyzing the changes in the parameters in figures 2a–2g, we can conclude that the period of oscillation of the numbers of fish is proportional to $(T_m T_s)^{1/2}$. Indeed, a fourfold increase in T_m (compare figures 2a and 2b) has led to twofold growth in the oscillation period. The same changes take place if T_s increases (compare figures 2a and 2c) and if T_s and T_m increase simultaneously (see figure 2d).

However, the oscillations are not always as smooth as those in figures 2a–2f. Often the oscillations cease, or their periods start to vary widely (see, for example, figure 2e). In some cases, all the sharks by a twist of fate happen to be far away from their prey and perish, and the mackerel population begins to grow steadily until they occupy the entire ocean.

Figure 2f shows the results of the modeling if we assume that the fish have become "cautious"—that is, they look around before they make their next move. If there is a shark next to the fish, it will swim in the opposite direction. Under such an algorithm for fish behavior, significant and regular oscillations of population occur much less often.

Thus, a computer model of the "real" life of a predator–prey system has given almost the same results as

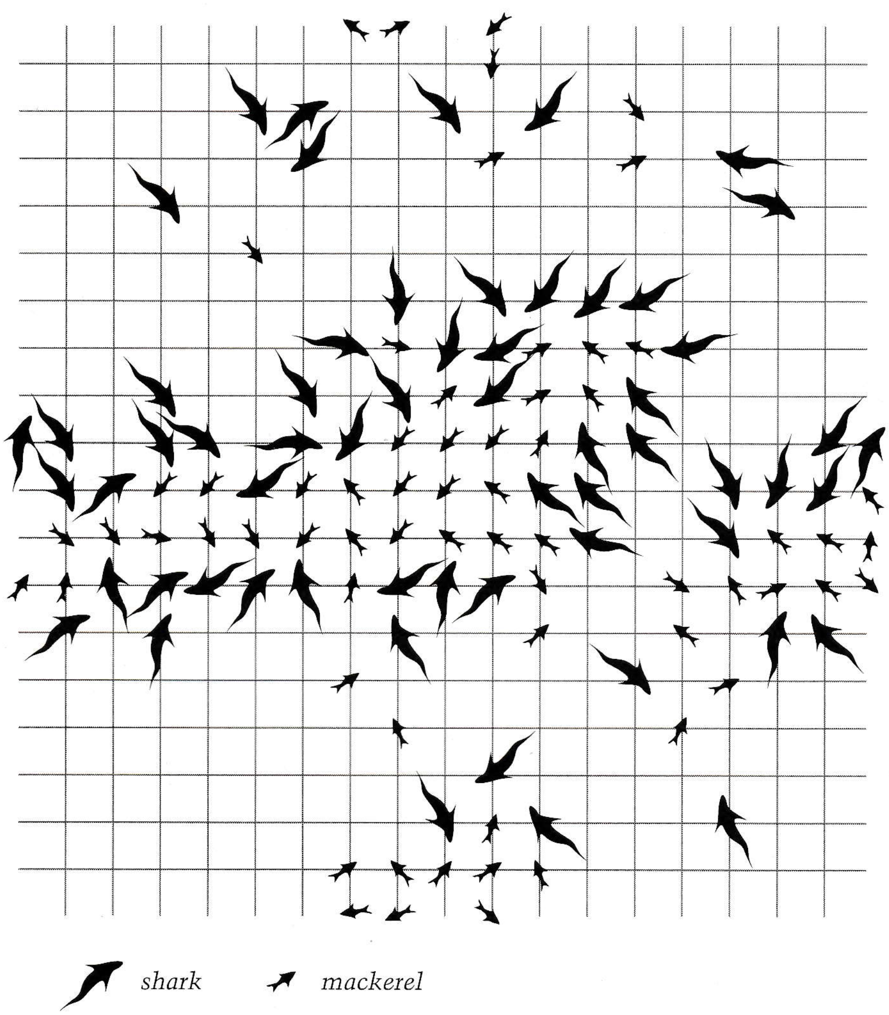


Figure 1
Two-dimensional ocean inhabited by sharks and mackerels.

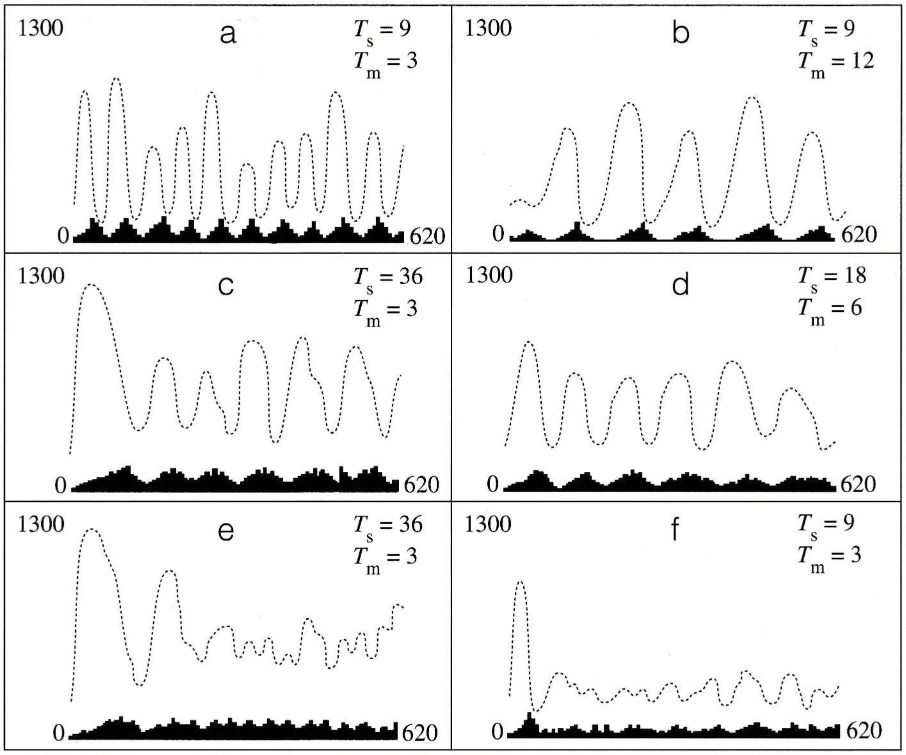


Figure 2

Population dynamics of the imaginary ocean (results of computer modeling). The population sizes are arrayed along the ordinate; along the abscissa—time (in relative units). T_m and T_s are the periods of time (in relative units) that mackerels and sharks, respectively, must live before each has a single offspring. The upper curves are the mackerel population, the lower curves—the shark population.

Volterra's equations, although it has highlighted several situations not described by these equations.

Why don't we notice such acute changes in the animal populations around us? After all, judging from the graphs in figures 2a–2b, we would conclude that the number of predators and prey must change by factors of 10 or more! The answer is simple. Volterra's equations and our model described the life of an isolated society consisting of one species of predator feeding on just one species of prey. Very rarely does this happen. More often several species of predators live in one territory, and they feed on several species of animals, including predators.

Each predator–prey system has its own frequency and phase of oscillations. If there are many such systems, and if they interfere with one another, the oscillations of population of animals are lessened. This happens by the same mechanism as when pendulums oscillating with different periods damp one another.

Nevertheless, it does happen that, in a large territory, one species of predator encounters only one species of prey. As a result, their populations change drastically as time passes, which is in complete agreement with the Volterra–Lotke model. A classic example of this would be the lynx–hare society in the Hudson Bay region of North America. Figure 3 shows how a North American company's annual catch of lynx and hares changed over 50 years.

Ecological chaos

We've gotten used to thinking that if a process is described by an equation (that is, we have found, for instance, the relation between forces and displacement), we can predict with absolute precision what will happen in the future. Classical examples of this predictability are the harmonic oscillator and the motion of celestial bodies.

Yet the existence of an equation doesn't always allow us to predict everything. Far from it. Several dynamical processes, even though they're described by seemingly explicit equations, are chaotic by nature. I'll illustrate this idea by solving one of Volterra's ecological equations.

Let a population of one species of animals inhabit a certain territory. If the territory were arbitrarily large, then according to Malthus's law the rate of growth of the population per unit time would be a constant a. But the finite dimensions of the territory mean that, as the population grows, the animals begin to suffer from a shortage of food and the birth rate decreases as a result.

So in a linear approximation, the equation describing changes in the population N of animals can be written as

$$dN/dt = (a - bN)N, \qquad (4)$$

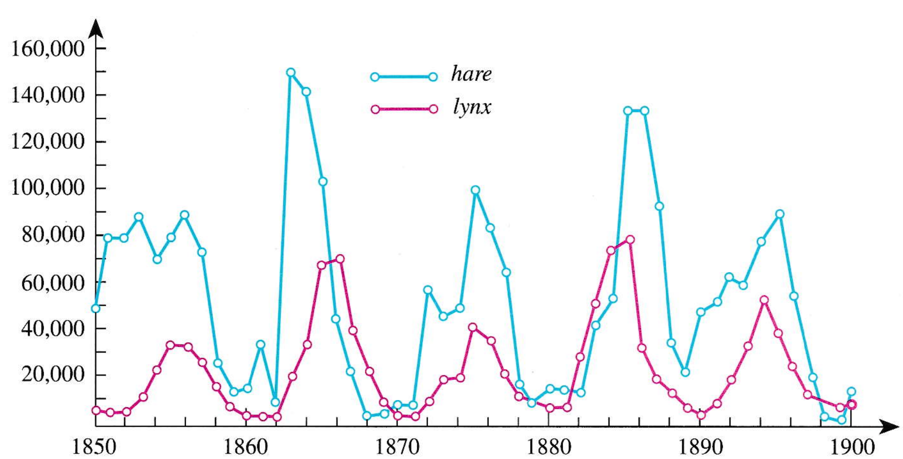


Figure 3

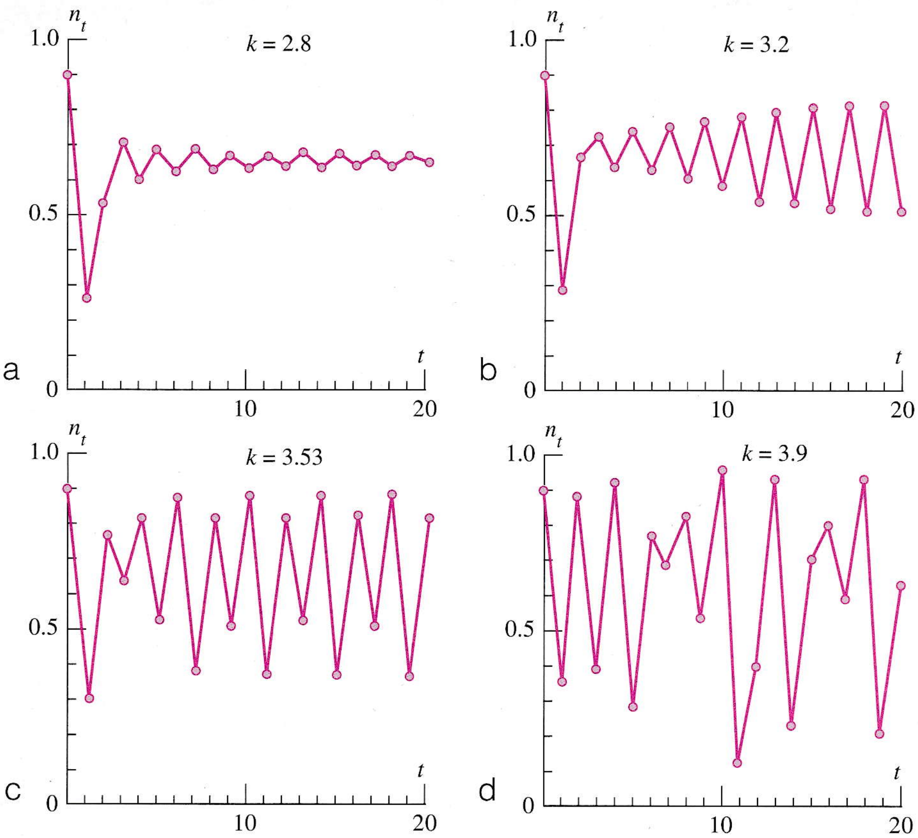


Figure 4

where b is a coefficient taking into account the decrease in the birth rate that occurs when the population grows on a limited territory.

Although it's impossible to solve equation (4) analytically, nothing is impossible for a computer. The only thing left is to transform (4) into an equation the computer understands better. Let the life of this animal population be divided into equal periods that we'll take to be units. If N_t and N_{t+1} are the animal populations during two consecutive periods, it follows from equation (4) that

$$N_{t+1} - N_t = (a - bN_t)N_t.$$

After the substitution $N = n(1 + a)/b$, we get

$$n_{t+1} = kn_t(1 - n_t), \qquad (5)$$

where $k = 1 + a$.

Equation (5) enables our computer to track the changes in the population period by period, but we have to remember that after we've normed the population n, it varies from 0 to 1.

The range of solutions to equation (5) is very broad and depends on the value of k. So, if $0 < k < 1$, animals are condemned to extinction regardless of the size of their initial population. If $1 < k < 3$, the number of animals n will eventually converge to a certain limit, equal to $(k - 1)/k$, that does not depend on the initial conditions (fig. 4a).

But the most interesting things happen to our animal population when k becomes greater than 3. For example, when $3 < k < 3.4$, the population, after a transitional period, begins to oscillate between two fixed values, and these oscillations are not damped over time; they resemble the oscillations in the predator–prey system (fig. 4b).

When k becomes slightly greater than 3.4, the animal population begins to oscillate regularly between four fixed values (fig. 4c). As the parameter k increases further, the number of fixed levels increases continuously and the changes in population become chaotic (fig. 4d).

One feature that distinguishes the chaotic process from a normal, predictable process is the very strong dependence on the initial values. As was mentioned above, the behavior of the system for $k < 3.57$ does not depend very much on the initial values; when there is chaos, $(k > 3.57)$, even a 1% difference in the initial values makes the processes entirely different after a certain time.

And one last question is: do chaotic processes occur in nature? Of course, and plenty of them. The changes in the populations of many animals, from one-celled creatures to mammals, are chaotic (see, for example, figures 2e and 2f). But maybe the most impressive illustrations of chaotic processes are the periodic epidemics that befall us. Figure 5 is drawn from data on continuous monthly observations of measles cases in New York City from 1928 to 1964.

The point is, the number of cases can be assumed to be proportional to the virus population in the given territory. Yet the reproduction rate of viruses in that territory depends entirely on the number of people in it and the degree of contact among them. In addition, every winter human resistance to infection decreases, while the frequency of contacts among people increases.

CONTINUED ON PAGE 30

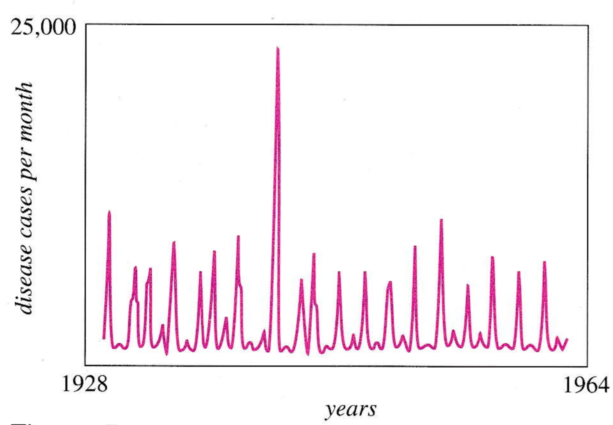


Figure 5

PHYSICS CONTEST

Sources, sinks, and gaussian spheres

"I do not perceive in any part of space, whether vacant or filled with matter, anything but forces and the lines in which they are exerted."
—Michael Faraday

by Arthur Eisenkraft and Larry D. Kirkpatrick

WHEN YOU VISIT AN ART gallery, you can gain an enhanced appreciation of a sculpture by comparing views from different angles. Likewise, it's often useful in physics to compare two different physical systems with the same mathematics in order to develop a better physical feeling for both. A prime example of this occurs with electrostatics and hydrodynamics.

Let's consider the steady-state flow of an incompressible fluid. Water is a very good approximation. Conservation of mass requires that the flow of mass into the volume be equal to the flow of mass out of the volume. Another way of stating this is that the net flow of mass—or flux—through the surface must be zero.

This is true unless there are sources or sinks inside the volume. If there are sources, the net flow of fluid through the surface must be positive—that is, there must be a net flow out of the volume. On the other hand, if the region contains a sink, the net flow will be inward, or negative. In general, the net flow will just be the sum of the positive contributions due to the sources and the negative contributions due to the sinks. If we let ρ represent the density of the fluid and $\mathbf{v}$ the velocity of the fluid at each point in space, this can be written in mathematical terms as

$$\sum \rho v_{\text{n}} \Delta A = \sum (\text{sources} + \text{sinks}),$$

where ΔA is a small piece of the surface, v_{n} is the component of the velocity normal to the surface element (v_{n} is positive when it points out of the volume). Only the sources and sinks *inside* the volume are included on the right side. In the notation of calculus we get

$$\int \rho v_{\text{n}} dA = \int (\text{sources} + \text{sinks}) dV.$$

Even though nothing is flowing in the case of the electrostatic field surrounding a distribution of electric charge, the mathematical equation is the same. We can think of the electric field E as the analogue of the flow of mass ($\rho\mathbf{v}$) in the fluid. In the electric case, the sources and sinks are charges; positive charges are sources of the electric field, and negative charges are sinks. This is all stated in Gauss's law:

$$\sum E_{\text{n}} dA = \frac{q_{\text{enc}}}{\varepsilon_0},$$

where ε_0 is a proportionality constant that depends on the units and q_{enc} is the net charge *inside* the volume. Note that it is easy to understand that only the charges inside the volume affect the sum by appealing to the case of the fluid. By analogy with fluid flow, the sum on the left side of the equation is called the "flux."

In cases of high symmetry, Gauss's law is very useful for finding the electric field due to a collection of point charges or a distribution of charge. As an example, let's calculate the electric field surrounding a positive point charge q located at the origin. Let's choose our surface to be a sphere of radius r centered on the origin to match the symmetry of the charge distribution. Because of the symmetry, we expect that the electric field will point radially outward and have the same value at all points on the surface. Therefore, the sum on the left side is easy to calculate:

$$\sum E_{\text{n}} \Delta A = E \sum \Delta A = EA = E 4\pi r^2.$$

Setting this equal to the right side and solving for E, we get

$$E = \frac{1}{4\pi\varepsilon_0} \frac{q}{r^2},$$

which we recognize as Coulomb's law.

For our contest problem, we apply Gauss's law to find the electric field for several different cases with spherical symmetry. The key to all but the last part is figuring out how much charge is enclosed in the gaussian sphere.

A. Find the electric field at all points outside a sphere of radius a that contains a uniform density of charge ρ and show that it has the same form as Coulomb's law.

B. Find the electric field at all points inside this sphere. Does the value at the surface agree with the value found in part A?

Art by Tomas Bunk


VOLTS
ATTENTION
DANGER
HIGH
VOLTAGE
ELECTRIC
BILL
STAFF

C. Now assume that a spherical region of radius b has been removed from the center of the sphere. What is the electric field at all points in space?

D. As a final challenge, find the electric field at all points inside the hole when the hole is moved off center. Assume that it is moved a distance c in the $+x$ direction, but not so far that the hole penetrates the surface. (Hint: consider the superposition of a complete sphere with charge density ρ and a "hole" formed by a smaller sphere with charge density $-\rho$.)

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

What goes up . . .

Here's the solution to the problem posed in the January/February issue.

A. In order for a satellite probe to escape from the Earth, the sum of the kinetic energy and potential energy must be greater than or equal to zero:

$$\frac{1}{2}m_p v_p^2 - \frac{Gm_p m_E}{R_O} = 0,$$

$$v_p = \sqrt{\frac{2Gm_E}{R_O}},$$

where m_p is the mass of the probe, v_p is its velocity, m_E is the mass of the Earth, and R_O is the radius of the Earth.

B. The condition for the probe to escape from the solar system is similar to that of escaping the Earth. In this case, however, the potential energy function is related to the mass of the Sun:

$$v_a = \sqrt{\frac{2Gm_S}{R_E}},$$

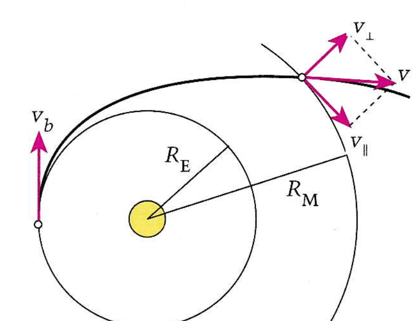


Figure 1

where m_S is the mass of the Sun, R_E is the Earth–Sun distance, and v_a is the probe's velocity relative to the Sun. The Earth's velocity about the Sun can be determined by recognizing that the gravitational attraction of the Sun holds the Earth in orbit:

$$\frac{m_E v_E^2}{R_E} = \frac{Gm_E m_S}{R_E^2},$$

$$v_E = \sqrt{\frac{Gm_S}{R_E}}.$$

The escape velocity can now be written

$$v_a = v_E\sqrt{2}.$$

Since the probe can be shot in the direction of the Earth's orbital velocity, the required velocity is diminished by the value of the Earth's velocity:

$$v_a' = v_E\left(\sqrt{2} - 1\right) = 12.3 \text{ km/s},$$

where v_a' is the probe's velocity relative to the Earth.

C. Let v_b and v_b' be the velocities of launching the probe in the Sun's and Earth's system of reference, respectively. Then $v_b = v_b' + v_E$ (see part B). From the conservation of angular momentum of the probe

$$m_p v_b R_E = m_p v_\parallel R_M$$

(R_M is the Mars–Sun distance) and the conservation of energy

$$\frac{m_p v_b^2}{2} - \frac{Gm_p m_S}{R_E} = \frac{m_p\left(v_\parallel^2 + v_\perp^2\right)}{2} - \frac{Gm_p m_S}{R_M},$$

we get, for the parallel component of the velocity (fig. 1),

$$v_\parallel = (v_b' + v_E)r$$

and, for the perpendicular component,

$$v_\perp = \sqrt{\left(v_b' + v_E\right)^2\left(1 - r^2\right) - 2v_E^2(1 - r)},$$

where $r = R_E/R_M$.

D. The minimum velocity of the probe in the Mars system of reference to escape from the solar system is $v_p'' = v_M\left(\sqrt{2} - 1\right)$ in the direction parallel to the orbit of Mars (v_M is the velocity of Mars around the Sun). The role of Mars is thus to change the velocity of the probe so that it leaves its gravitational field with this velocity.

In the Mars system, the energy of the probe is conserved. That is not true, however, in the Sun's system, in which this encounter can be considered an elastic collision between Mars and the probe. The velocity of the probe before it enters the gravitational field of Mars is therefore, in the Mars system, equal to the velocity with which the probe leaves its gravitational field. The components of the former velocity are $v_\perp'' = v_\perp$ and $v_\parallel'' = v_\parallel - v_M$, so

$$v'' = \sqrt{v_\parallel''^2 + v_\perp''^2} = \sqrt{v_\perp^2 + \left(v_\parallel - v_M\right)^2} = v_S''.$$

Using the expressions for $v_\parallel$ and $v_\perp$ from part C, we can now find the relation between the launching velocity from the Earth v_b' and the velocity $v_p'' = v_M\left(\sqrt{2} - 1\right)$:

$$\left(v_b' + v_E\right)^2\left(1 - r^2\right) - 2v_E^2(1 - r) + v_M^2 + \left(v_b' + v_E\right)^2 r^2 - 2v_M\left(v_b' + v_E\right)r = v_M^2\left(3 - 2\sqrt{2}\right).$$

The velocity of Mars around the Sun is $v_M = \sqrt{Gm_M/R_M} = \sqrt{r}v_E$, and the equation for v_b' takes the form

$$\left(v_b' + v_E\right)^2 - 2r\sqrt{r}v_E\left(v_b' + v_E\right) + \left(2\sqrt{2r} - 2\right)v_E^2 = 0.$$

The physically relevant solution is

$$v_b' = v_E\left(r\sqrt{r} - 1 + \sqrt{r^3 + 2 - 2\sqrt{2r}}\right) = 5.5 \text{ km/s}.$$

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M56

Splitting reciprocals. (a) Prove that for any integer $a > 1$ the equation

$$\frac{1}{x} + \frac{1}{y} = \frac{1}{a}$$

has at least three positive integer solutions (x, y).

(b) Find the number of positive integer solutions of this equation for $a = 1992$. (M. Slavinsky)

M57

One plus one exceeds two. Does there exist a figure F that can't cover a semicircle of radius 1, while two copies of F are enough to cover an entire unit circle, in the case of (a) an arbitrary plane set F, (b) a convex F? (N. Vasilyev, A. Samosvat)

M58

Decreasing squares. For any nonnegative numbers $a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$, prove the following inequalities:

(a) $a_1^2 - a_2^2 + a_3^2 \geq (a_1 - a_2 + a_3)^2$,
(b) $a_1^2 - a_2^2 + a_3^2 - a_4^2 \geq (a_1 - a_2 + a_3 - a_4)^2$,
(c) $a_1^2 - a_2^2 + \ldots - (-1)^n a_n^2 \geq (a_1 - a_2 + \ldots - (-1)^n a_n)^2$.

(L. Kurliandchik)

M59

All kinds of centers. Given three points O, I, and E in the plane, construct a triangle such that its circumcenter is at O, its incenter is at I, and one of its excenters is at E. (An excenter is the center of an *escribed circle* of a triangle—a circle tangent to one side of the triangle and to the extensions of the other two sides. Every triangle has three such circles.) (B. Martynov)

M60

Coloring a chessboard. A square of size $n \times n$ ruled with a grid of unit squares is colored by using n colors (each cell is colored in one of the colors or not colored at all). A coloring is called *regular* if there are no two squares of the same color in any row or any column. Is it possible to complete a regular coloring if k unit squares have already been regularly colored for (a) $k = n^2 - 1$, (b) $k = n^2 - 2$, (c) $k = n$? (D. Logachev)

Physics

P56

Running ant. An ant runs from an anthill in a straight line so that its speed is inversely proportional to the distance from the center of the anthill. At the moment the ant is at point A at a distance $l_1 = 1$ m from the center of the anthill, its speed is equal to $v_1 = 2$ cm/s. How long will it take for the ant to run from point A to point B at a distance $l_2 = 2$ m from the center of the anthill. (S. Krotov)

P57

Jumping wheel. A small weight of mass m is fitted on the rim of a massive wheel of mass M ($M/m = 15$). How fast must the wheel be rolling before it jumps up off the surface?

P58

Nitrogen and oxygen. A mixture of gases, consisting of $m_1 = 100$ g of nitrogen and an unknown amount of oxygen, is subjected to isothermal compression at a temperature $T = 74.4$ K. Figure 1 shows the dependence of the gas mixture's pressure on its volume (in relative units). Let's determine the mass of oxygen m_{O_2}. To do this we have to calculate the saturated vapor pressure of oxygen p_{O_2} at this temperature.

Note: $T = 74.4$ K is the boiling point of liquid nitrogen at normal pressure; oxygen boils at a higher temperature. (A. Buzdin)

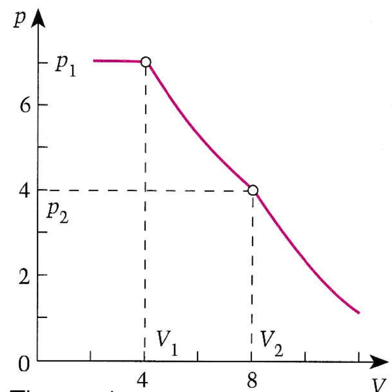


Figure 1

P59

Mercury watch. Two columns of mercury separated by a drop of electrolytic solution HgI are in a capillary tube. The inner diameter of the tube is $d = 0.3$ mm. The tube is connected by a resistor with resistance $R = 390$ kΩ to a battery with electromotive force $\mathcal{E} = 10$ V. How much time will it take for the drop to move one notch on the scale (see figure 2)? (E. Yunosov, I. Yaminsky)

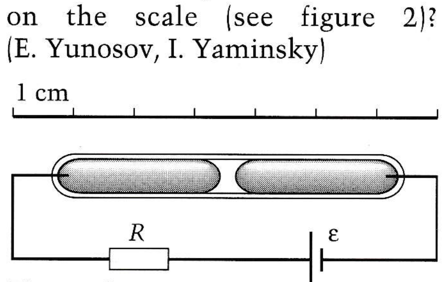


Figure 2

CONTINUED ON PAGE 30

AT THE BLACKBOARD I

Summertime, and the choosin' ain't easy

An ice cream counting problem

by Kurt Kreith

HARD ON THE HEELS OF McDonald's and Pizza Hut, an American company made famous by its trademark number—31—has opened several stores in Russia. Muscovites are said to relish ice cream, even in the chill of winter, and many people might frequent Baskin-Robbins primarily for the thirty-one different flavors of ice cream it pledges to keep on hand. As a thoughtful reader of *Quantum*, though, you'll realize that such a store offers far more than thirty-one different options.

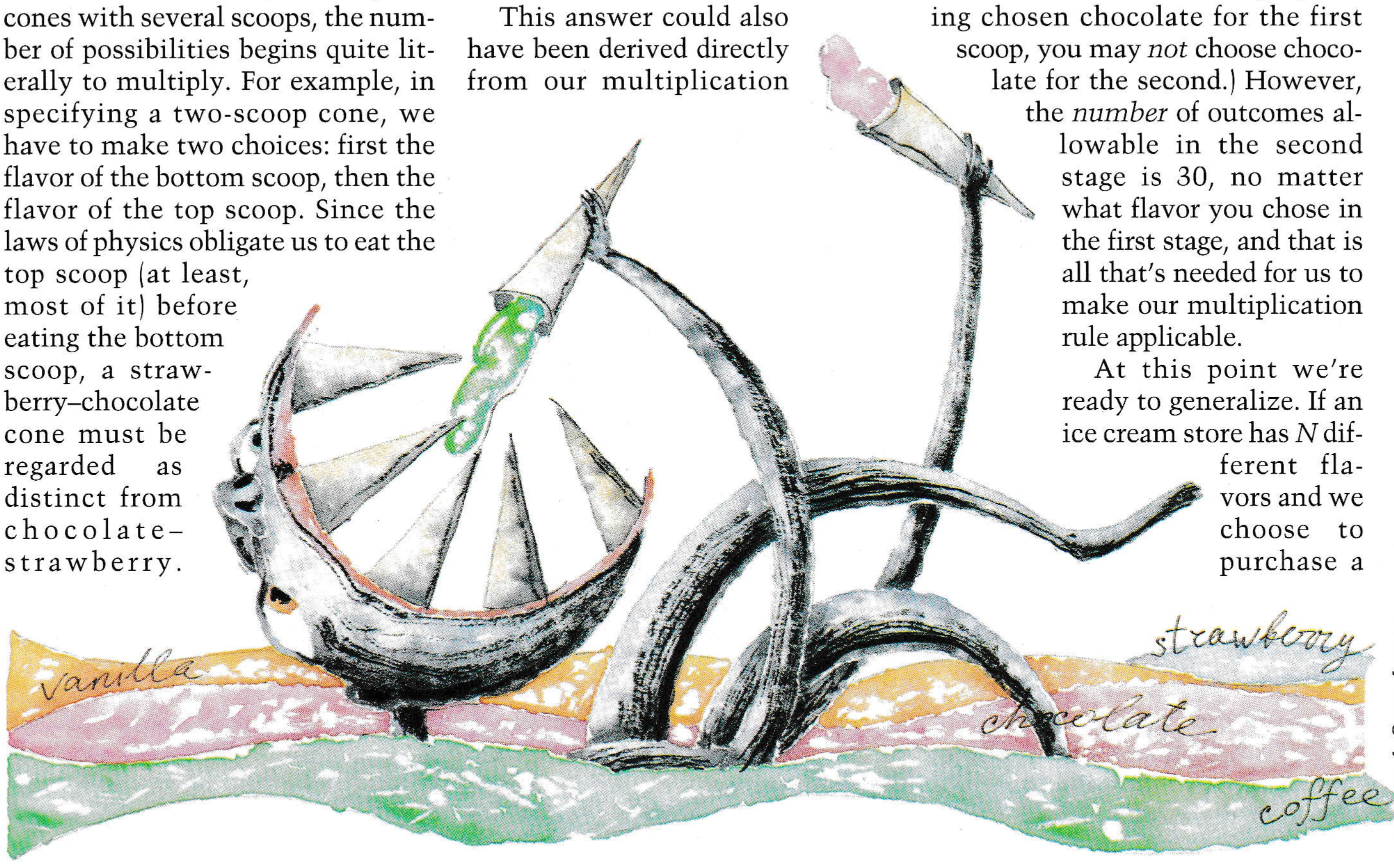
When we contemplate cups or cones with several scoops, the number of possibilities begins quite literally to multiply. For example, in specifying a two-scoop cone, we have to make two choices: first the flavor of the bottom scoop, then the flavor of the top scoop. Since the laws of physics obligate us to eat the top scoop (at least, most of it) before eating the bottom scoop, a strawberry–chocolate cone must be regarded as distinct from chocolate–strawberry. With this understanding, it's clear that Baskin-Robbins offers $31 \times 31 = 961$ different kinds of two-scoop cones.

An important combinatorial rule arises when we impose the following constraint on the construction of ice cream cones: *At most one scoop of any single flavor is allowed.* This rule reduces somewhat the number of two-scoop cones available, since thirty-one possibilities (cones with both scoops of a single flavor) are no longer allowed. In other words, $961 - 31 = 930$ different two-scoop cones are now available.

This answer could also have been derived directly from our multiplication rule. To see this we again regard the construction of two-scoop cones to be a two-stage process, in which the first stage again allows for 31 different outcomes, but the second stage allows for only 30. According to the multiplication rule, the composite process allows for $31 \times 30 = 930$ different outcomes. Fortunately, this is the same answer as before.

A subtle point you may have noted is that the two stages of constructing a cone aren't "independent"—that is, the outcome of the first stage of such a process affects the set of possible outcomes in the second stage. (Having chosen chocolate for the first scoop, you may *not* choose chocolate for the second.) However, the *number* of outcomes allowable in the second stage is 30, no matter what flavor you chose in the first stage, and that is all that's needed for us to make our multiplication rule applicable.

At this point we're ready to generalize. If an ice cream store has N different flavors and we choose to purchase a



Art by Sergey Ivanov

cone with K scoops, the number of possible outcomes depends on whether we're allowed to repeat flavors. If repetition is allowed, then the number of options is N^K. If, however, repetition is not allowed, then the number of outcomes is

$$N \times (N-1) \times \ldots \times (N-K+1).$$

This answer is often referred to as "the number of permutations of N things taken K at a time" and written as

$$_NP_K = \frac{N!}{(N-K)!}.$$

Our ice cream saga might end here except for the fact that many people prefer to eat their ice cream from cups rather than from cones. One reason for preferring cups is that they allow you to eat the various scoops in any order. That is, a strawberry–chocolate cup is the same as a chocolate–strawberry cup.

Let's return to our counting problem, keeping this new circumstance in mind. If we insist, as we did to arrive at permutations, that *only one scoop of any flavor is allowed*, then the answer is a familiar one: each three-scoop cup corresponds to $3! = 6$ different three-scoop cones. To arrive at the number of different three-scoop cups that are possible, we simply divide our previous answer of $31 \times 30 \times 29 = 26{,}970$ by $3!$ to get 4,495. Those of you who have studied combinatorics will recognize this problem as one of "choosing 3 things from 31," which corresponds to the combinatorial formula

$$_{31}C_3 = \frac{31!}{3! \cdot 28!}$$

and is read "thirty-one choose three." These ideas generalize readily to choosing K flavors from N and the combinatorial formula

$$_NC_K = \frac{N!}{K!(N-K)!}.$$

To complete this problem, we need to remove the rather arbitrary restriction on the choice of flavors. Why should one *not* be able to order two scoops of chocolate and one scoop of strawberry in a cup? Given this rather fundamental right, how many different K-scoop cups can one generate from N different flavors?

Surprisingly, the answer to this question appears not to correspond to any of the well-known counting rules of combinatorial analysis. We can, however, answer the question rather easily for two-scoop cups and 31 flavors. As we've already seen, there are 961 different cones, of which

31 have two scoops of the same flavor

and

930 have two scoops of different flavors.

Converting from cones to cups has the effect of halving the number of possibilities with two scoops of different flavors while leaving the number of possibilities with two scoops of the same flavor unchanged. That is, our answer is now

$$31 + \frac{930}{2} = 31 + 465 = 496.$$

This kind of analysis is also possible for three-scoop cups. I leave it to you to confirm that the answer is now

$$31 + 930 + 4{,}495 = 5{,}456.$$

But such a direct approach is rather awkward for dealing with the general problem of K scoops selected from N flavors when repeated flavors are allowed and the order doesn't matter—that is, when we're eating K scoops from a cup rather than a cone.

To address this general problem we conceive of an ice cream order form in which the N flavors are listed horizontally. Reducing the number of flavors from 31 to 6 makes such a form look like this:

vanilla	chocolate	strawberry	peach	mint	coffee

When we want to order a scoop of a certain flavor, we put "0" under its name. For example, an order form corresponding to a cup with two scoops of chocolate and one scoop of peach would be

vanilla	chocolate	strawberry	peach	mint	coffee
	00		0		

After a while we might get so familiar with this system that we could eliminate the top line. To place the order above, we'd simply write

	00		0		

Can you figure out the order corresponding to this?

	0	0			0

(It calls for one scoop of chocolate, strawberry, and coffee.)

And how about the very hungry customer who, not restricted to three scoops, orders the following?

00	0	00		00	

What these order forms show is that there is a one-to-one correspondence between three-scoop cups chosen from six different flavors and sequences of three zeros (corresponding to the scoops) and five ones (corresponding to the dividers between the six flavors on our order form). In other words, the abstract versions of such cups of ice cream are sequences of the form

$$(0, 1, 1, 0, 0, 1, 1, 1).$$

While less attractive than the real thing, this representation has some mathematical advantages. In particular, we can readily determine the number of three-scoop cups that six flavors generate by counting the number of sequences consisting of three zeros and five ones.

One way to do this is to revise our order form to one used only for choosing three scoops from among six flavors. Such a form consists simply of $3 + 5 = 8$ boxes arranged horizontally:

The way we use this form is to choose three of the boxes in which to enter zeros and then enter ones in the remaining five boxes—for example,

1	0	0	1	1	1	1	0

Each such choice corresponds to "a sequence of three zeros and five ones," and since we're "choosing three boxes from eight" in which to enter our zeros, the number of such choices is ${}_8C_3 = 8!/3! \cdot 5! = 56$.

The advantage of this approach is that it allows for a direct generalization to the problem of N flavors and K scoops. An order form for N flavors requires $N - 1$ dividers among the N flavors. So it corresponds to a sequence of K zeros and $(N - 1)$ ones. Since the number of such sequences is

$${}_{N-1+K}C_K = \frac{(N-1+K)!}{(N-1)! \cdot K!},$$

this is also the number of K-scoop cups that can be produced with N flavors when more than one scoop of a flavor is allowed. In the case $K = 3$ and $N = 31$, we get

$$C = \frac{33!}{30! \cdot 3!} = 5{,}456$$

as before.

After all this work we deserve to sit back and enjoy some ice cream!

Problems

One of the interesting features of the counting rules we've developed is that they arise in a great variety of situations. Recognizing situations that correspond to combinations, permutations, or related counting rules is an important mathematical skill. But one has to accept that there are counting problems that, though very similar to those we have solved, may not correspond to any of these or, in fact, to any (known) rule.

Which of the following problems can you solve and generalize with the skills you've developed in thinking about ice cream?

1. A poor father has three pennies and six children. In how many ways can he distribute the three pennies to six children?
2. Repeat problem 1 under the assumption that the father gives at most one coin to each child.
3. Repeat problems 1 and 2 under the assumption that the father has a penny, a nickel, and a dime.
4. A rich father has six pennies and three children. In how many ways can he distribute six pennies to three children?
5. Repeat problem 4 under the assumption that the father gives at least one coin to each child.
6. Repeat problems 4 and 5 under the assumption that the father has one penny, one nickel, one dime, one quarter, one half dollar, and one silver dollar.

ANSWERS ON PAGE 61

Kurt Kreith *teaches mathematics at the University of California—Davis.*

"FIBONACCI STRIKES AGAIN!" CONTINUED FROM PAGE 17

Mathematicians haven't yet unraveled this mystery. We may never fully understand why this series or the related golden ratio are so prevalent, but it would appear that the Fibonacci series is truly a mathematical phenomenon worthy of continued study and special recognition.

A few teachers we know have even given the Fibonacci series a place of honor on their classroom bulletin boards. One of them has gone so far as to hold a "Fibonacci Day," giving students the opportunity to research and present information on different applications of the Fibonacci series. While some math skeptics may secretly believe that number-series problems really don't relate much to real life, the Fibonacci series strikes again and again to show them otherwise!

Elliott Ostler, *M.S., teaches mathematics at Elk Horn–Kimballton High School in Elk Horn, Iowa, and is pursuing a doctorate in mathematics education at the University of South Dakota.* **Neal Grandgenett** *earned a Ph.D. from Iowa State University and is now an assistant professor of mathematics education at the University of Nebraska at Omaha.*

"CHALLENGES IN PHYSICS AND MATH" CONTINUED FROM PAGE 27

P60

Reflecting cone. The inner surface of a cone coated with a reflective layer forms a conical mirror. A thin incandescent filament is aligned along the axis of the cone. Determine the minimum angle α of the cone such that the rays emitted by the filament will be reflected from the conical surface no more than once. (D. V. Belov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 56

"MALTHUS AND VOLTERRA" CONTINUED FROM PAGE 23

So the man–virus system differs from the system described in the preceding paragraphs only in its periodic changes in the parameter k. However, that makes the analysis of this system much more complicated, taking it far beyond the limits of this article. Still, I think you'll take my word for it (perhaps out of fatigue) that the solutions of equations describing epidemic processes can be either regular oscillations or completely chaotic oscillations, depending on the values of the parameters.

As you've probably already guessed, chaotic solutions are characteristic of nonlinear differential equations—that is, those in which the function (in our example, the population N) is a power other than one. So chaos isn't something characteristic only of living nature. In the material world, too, a great many processes continuously change from regular (periodic) to chaotic. Some examples are water dripping from a faucet, turbulent flows of liquid and gas, and the circulation of our planet's atmosphere.

Long live the human being, an example of a nonlinear system and thus capable of the most unpredictable actions! And long live Nature, ever gladdening us with its unpredictable beauty!

Just for the fun of it!

B56

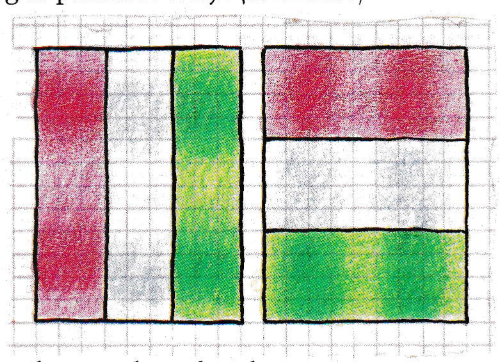
Square anniversary. In the year x^2 my nephew will be x years of age. In what year was he born? (L. Kurliandchik)

B57

In search of consensus. Is there a temperature expressed by the same number of degrees in both the Celsius and Fahrenheit scales? (We remind you that these two uniform scales can be defined by the melting point of ice, which is 0°C or 32°F, and the boiling point of water, which is 100°C or 212°F.) (A. Savin)

B58

Dirty windshield. Sometimes a car's windshield gets dirty when blobs of muddy water are tossed up by other vehicles, reducing visibility. Yet the experienced driver doesn't turn on the windshield wipers right away and avoids getting the glass wet for as long as possible. Why? (S. Krotov)



B59

Cheap remake. The figure at right shows two flags measuring 9 × 12. Cut the flag on the left into four pieces so that you could stitch them together to make the flag on the right. (A. Shvetsov)

Art by Edward Nazarov

B60

Greed punished. Koshchei the Immortal, a greedy and malicious tsar-sorcerer (and an indispensable character in Russian fairy tales), buried his ill-gotten treasure in a hole 1 meter deep. That didn't seem safe enough for him, so he dug his treasure up, deepened the hole to 2 meters, and buried his hoard again. He was still worried, so he dug up his hoard, made the hole 3 meters deep, and hid his treasure once more. But he just couldn't stop—he kept increasing the depth of the hole, to 4 m, 5 m, 6 m, and so on, each time extracting his property and burying it again, until on the 1,001st day he died of exhaustion. It's known that Koshchei digs a hole n meters deep in n^2 days. How deep was his treasure when he dropped dead? (Neglect the time needed to refill the hole each time.) (D. Fuchs)

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Wake up!

It's time for a pop quiz

by Anatoly Savin

YES, WE KNOW IT'S VACAtion time. But we also know that, as the bees drone into their final months and the air settles over you like a wool blanket, the blood stops circulating in the brain. We're into the lazy days of summer, and let's face it—you're bored. So here are a dozen recreational problems to be solved *without pencil and paper*. All you need is a nice shady place.

And we offer you three games for when the weather is cooler and physical exercise becomes an acceptable choice. (Rest assured, the exercise proposed is quite modest!) Each game has a winning strategy—it simply depends on who goes first. We think you'll enjoy trying to figure it out.

If you're going to be at a summer camp or school, maybe you'd like to share these problems and games with your acquaintances there. A fun competition might be a welcome diversion from more strenuous, unsummerlike activities, and it's an amusing way to get to know each other.

Problems

1. The number 606 is written on a sheet of paper. What operation must be performed to make the number 3/2 times greater?

2. What mathematical sign must be put between the digits 5 and 6 to obtain a number greater than 5 but smaller than 6?

3. It takes four seconds for a clock to strike three times. How long will it take this clock to strike nine times?

4. The sum of the three numbers 1, 2, and 3 equals their product. Are there any other such triples of integers?

5. I have two coins in my pocket, whose value totals 15 cents. Is it possible that one of them isn't a dime?

6. How many times must one shoot at the target in the illustration, and what circles must one hit, in order to score exactly 100 points?

7. The Moscow TV tower is 530 m high and weighs 30,000 metric tons. How many grams does a scale model, made of the same material, weigh if the model is 53 cm high?

8. I have a box of nails, a box of screws, and a box of nuts. Each box had a label telling what's inside. Then my little brother switched the labels so that none of them corresponds with the contents of its box. Is it possible to figure out what's in each box after opening only one of them?

Art by Edward Nazarov

9. Find the smallest positive integer that becomes the square of an integer when multiplied by 2 and the cube of an integer when multiplied by 3.

10. Six glasses are lined up on a shelf: three of them are filled with water and the other three are empty (see the illustration). Touching only one glass, how can one arrange them so that glasses with water alternate with empty ones?

11. Find two numbers whose sum, product, and ratio are equal to each other.

12. A glass of water is set on a stool covered with a napkin. How can one remove the napkin, leaving the glass on the stool, without touching the glass?

Games

1. You have two piles of rocks. Two players take turns (1) removing one of the piles from the playing area and (2) dividing the remaining pile into two piles. The player who can't divide a pile (since it consists of only one rock) loses.

2. Several pegs are driven into the ground, and the two players have some string. Each player takes turns tying together pairs of pegs that haven't been connected earlier. The player who creates a closed figure wins.

3. Two players pluck petals from a daisy, tearing off one or two neighboring petals at a time. The player who plucks the last petal wins. ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Meandering down to the sea

The wandering ways of rivers

by Lev Aslamazov

HAS ANYONE EVER SEEN A river that doesn't have any bends in it? A short portion of a river can certainly cut a relatively straight path, but there are no rivers that have absolutely no bends in them. Even if the river flows through a plain, it often loops around, and these bends are often repeated with a definite period. The shoreline at these bends is also curious: one bank tends to be steep, while the other slopes gently. How can we explain these peculiarities of river behavior?

Hydrodynamics is the branch of physics that deals with the motion of liquids. Although it's a well-developed science, rivers are such complicated natural features that even hydrodynamics can't explain all the peculiarities of their behavior. Nevertheless, it can answer many of our questions.

You may be surprised to learn that the problem of meanders was investigated by Albert Einstein. In a report delivered in 1926 at a meeting of the Prussian Academy of Sciences, Einstein compared the movement of water swirling in a glass and in a riverbed. This analogy allowed him to explain why rivers tend to acquire a meandering shape.

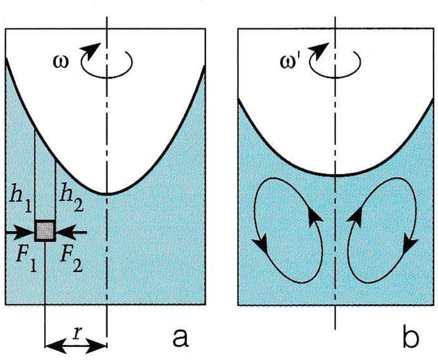


Figure 1

Let's try to understand this phenomenon, at least qualitatively. And let's start with a glass of tea.

Tea leaves in a glass

Make a glass of tea with loose tea leaves (not in a bag), stir it well, and remove the spoon. The water will gradually stop and the tea leaves will gather in the center at the bottom of the glass. What made them gather there? To answer this question, we need to determine the shape the water's surface takes as it swirls in a glass.

This experiment with the tea shows that the surface of a liquid—in this case, water—is curved. It's easy to understand why that is. In order to make the molecules of water move in circular motion, the net force acting on each molecule must provide the centripetal acceleration. Let's consider a cube of mass Δm situated in the liquid at a distance r from the axis of rotation (the gray square in figure 1). At an angular speed of rotation ω, the centripetal acceleration of the cube is $\omega^2 r$. This acceleration is created by the difference in the pressures acting on the faces of the cube (the left and right faces in figure 1). So

$$\Delta m \cdot \omega^2 r = F_1 - F_2 = (p_1 - p_2) \cdot \Delta s, \quad (1)$$

where Δs is the area of one face. The pressures p_1 and p_2 are determined by the heights h_1 and h_2 to the surface of the liquid:

$$p_1 = \rho g h_1, \; p_2 = \rho g h_2, \quad (2)$$

where ρ is the density of the liquid and g is the acceleration due to gravity. F_1 must exceed F_2, so h_1 must exceed h_2—that is, the surface of the rotating liquid must be curved, as shown in figure 1. The greater the speed of rotation, the greater the surface's curvature.

We can find the shape of the curved surface of a liquid. It turns out to be a paraboloid—that is, a surface with a parabolic cross section. (Try to prove it yourself. You can check your proof by looking on page 61.)

While we stir tea with the spoon, we keep the liquid swirling. But when we remove the spoon from the glass, the friction between layers of the liquid (its viscosity) and the friction of the liquid with the sides and bottom of the glass converts the liquid's kinetic energy into heat, and the liquid gradually stops swirling.

As the frequency of rotation decreases, the surface of the liquid flattens out. Vortex currents appear within the liquid (their directions are shown in figure 1b). The vortex currents form because of the difference in the friction of the liquid at the bottom of the glass and at the surface. The liquid decelerates more near the bottom, where the friction is greater, than near the surface. So, even though

Art by Elena Trofimova

the particles of liquid are equal distances from the axis of rotation, they have different speeds: particles that are closer to the bottom are slower than those near the surface. But the net force due to the pressure differences acting on all these particles is the same. This force can't cause the necessary centripetal acceleration for all the particles (as in the case of turning the entire liquid with the same angular speed). Near the surface the angular speed is too large, and particles of water are thrown to the sides of the glass; near the bottom the angular speed is too small, and the resultant force makes the water move to the center.

So tea leaves gather at the center of the bottom of the glass because they're drawn there by vortex currents that arise during deceleration. Of course, this is a simplified version of what occurs, but it accurately reflects the gist of the phenomenon.

How riverbeds change

Now let's look at how water moves at a bend in a river. A picture forms that is quite similar to that of our glass of tea. The surface of the water inclines toward the bend so that pressure differences create the necessary centripetal acceleration. (Figure 2 shows the cross section of the river at the bend.) And just like in the glass of tea, the velocity of the water near the bottom is less than that near the surface of the river (the velocity distribution by depth is shown in figure 2; the vertical plane normal to the cross section of the river is depicted in red). Near the surface the net force due to the pressure difference can't move the water particles along the circumference, so water is "thrown" to the far shore (the one that's farther from the center of the bend). Near the bottom, on the other hand, the velocity is small, so the water moves toward the near shore (the one nearer to the center of the bend). So this additional circulation of water appears in addition to the main flow, as demonstrated in the glass of tea. Figure 2 shows the direction of water circulation in the plane of the river's cross section.

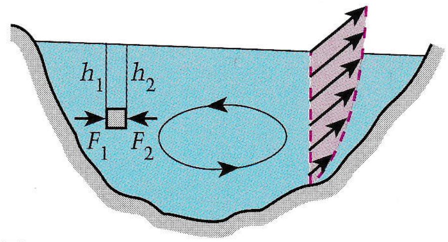


Figure 2

This type of water circulation causes soil erosion. As a result, the far shore disintegrates as the ground there is undermined, while soil gradually settles along the near shore, forming an ever thicker layer (recall the tea leaves in the glass!). The shape of the riverbed changes, acquiring the cross section shown in figure 3.

It's also interesting to observe how the velocity of the water flow varies across the breadth of the river (from bank to bank). In straight stretches, the water flows most quickly in the middle of the river. At a bend, the line of fastest flow shifts to the far shore. This happens because it's more difficult to turn fast-moving water particles than slow-moving ones. A larger centripetal acceleration is needed. But where the velocity of the flow is greater, the circulation of the water, and consequently the soil erosion, is also greater. That's why the fastest place in a riverbed is usually the deepest—a fact well known to river pilots.

Soil erosion along the far shore and sedimentation along the near shore causes a gradual shift of the entire riverbed away from the center of the bend, thereby increasing the river's meander. Figure 3 shows the cross section of the same place in a real riverbed in different years. You can see clearly the gradual shift of the riverbed and the increase in its meander.

So even a small bend created by chance—by a landslide or a fallen tree, for example—will increase. The straight flow of a river across a plain is unstable.

How meanders are formed

The shape of a riverbed is largely determined by the relief of the terrain it flows through. A river flowing through an uneven landscape meanders in order to avoid high places and fill low places. It "chooses" the path with the maximum slope.

But how do rivers flow along a plain? How does the instability of a straight flow described above influence the shape of a river? The instability increases as the river grows in length and the river begins to meander. It's natural to think that in the ideal case (an absolutely flat, homogeneous surface), a periodic curve must appear. What will it look like?

Geologists have put forth the proposition that at their meanders, rivers flowing through plains should look like a bent ruler.

Take a steel ruler and bring its ends together. The ruler will bend as in figure 4. This elastic bending is called Euler bending after the great mathematician Leonhard Euler (1707–1783), who described this phenomenon theoretically. The shape of the bent ruler is described by a special curve. This Euler curve has a wonderful feature: of all the possible curves of a given length connecting given points, it has the smallest curvature on average. If we measure the angular deflection θ at equal intervals along the curve (fig. 4) and add up the squares of the angular deflections, we'll obtain the minimum sum for the Euler curve. This "economical"

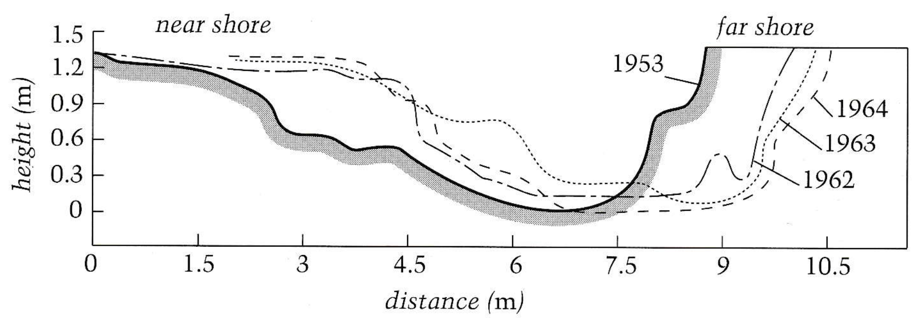


Figure 3

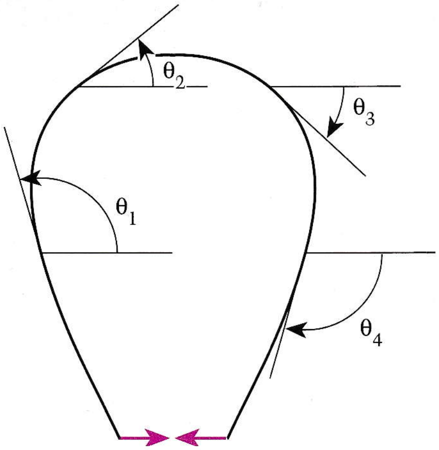


Figure 4

bending of the Euler curve was the basis for the hypothesis about the shape of riverbeds.

To test this hypothesis, geologists modeled a changing riverbed. They used an artificial channel of water in a homogeneous medium made of small particles held together so that it was easily eroded. Soon the straight channel began to meander, and the shape of the bend was described by the Euler curve (fig. 5). Of course, under actual conditions such perfection in the shape of a riverbed isn't attainable (because of the heterogeneity of the soil, for instance). But rivers flowing through a plain usually meander and form a periodic structure. In figure 6 you can see a real riverbed and the Euler curve (the dashed line) approximating the shape of the riverbed.

You may be interested to know that the term "meander" originated in ancient times and comes from the Meander, a river in Turkey famous for its twists and turns (now called the Menderes). The periodic deflections of ocean currents and of the brooks that form on the surfaces of glaciers are also called meanders. In each of these cases, random processes in a homogeneous medium cause a periodic structure to form; and though the reasons for the creation of meanders can differ, the shape of the resulting periodic curves is always the same. ◘

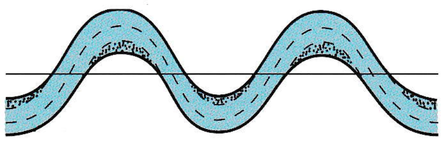


Figure 5

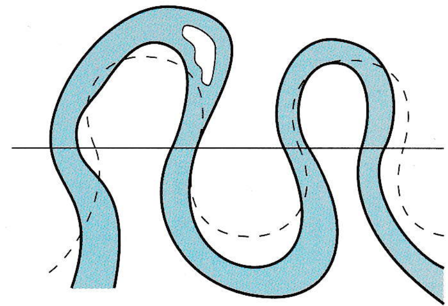


Figure 6

What did the conductor say?

The trains of thought of lazy wise guys with dirty faces

by Mikhail Gerver

1. Mathematical induction

"You know what nights in Ukraine are like . . ." Oska began with great feeling.

"No! No, we don't!" came several voices from the audience. "Tell us! Please, tell us!"

"No—you don't know what they're like!" Oska continued, a bit taken aback.

"Of course we don't," the mothers agreed. "How could we? We never had time for book-learning."

—L. Kassil, *The Black Book and Schwambrania*

Do you know what mathematical induction is?

"We know, we know!" I seem to hear my readers answer. "Let the first in a line be a woman, and let a woman follow every woman in the line. Then all the people in the line are women!" This is a humorous rendering of the principle of mathematical induction. And here is a serious one: "Let there be a sequence of statements $Y_1, Y_2, Y_3, \ldots$. Let the first statement Y_1 be true, and let every true statement Y_k be followed by a true statement Y_{k+1}. Then all statements Y_n are true."

Well, then. This is indeed so. Since true statement Y_{k+1} follows each true statement Y_k, and since statement Y_1 is true, statement Y_2 is true; and because Y_2 is true, Y_3 is true; but then Y_4, following Y_3, is true; and so forth.

And yet I say (you'll forgive me, won't you?): "No, you don't! You don't know what mathematical induction is!" Let's just check it out. A number of trials await you: in sections 2 and 3 they're simpler, in sections 4 and 5 more complex.

2. Hocus-pocus

If it says "buffalo" on an elephant's cage, don't believe your own eyes.

—Kozma Prutkov, *Aphorisms*

Using mathematical induction, one can provide a theoretical basis for the following amazing trick.

Cut out 999 identical cards. Write 1 on 111 cards, write 2 on 111 cards, and so forth; on the last 111 cards write 9. Turn over all the cards, figures down, and shuffle them thoroughly. Then choose at random n cards, where n is a whole number from 1 to 100, and put them on the table, figures up. The figures written on all n chosen cards will unfailingly be the same! No matter how many times you repeat this trick, and whatever n you choose ($1 \le n \le 100$), the result will invariably be the same: only one figure will appear on n cards. You can check it experimentally at your leisure. Now I'll prove this remarkable fact by induction.

We must prove the statement Y_n: *The same figure shows up on all n chosen cards ($1 \le n \le 100$).* Actually, there are 100 statements here: $Y_1, \ldots, Y_{100}$ (depending on what value n takes).

Obviously, our statement is valid for $n = 1$. This is no surprise: if only one card is turned up, then one figure will of course show up.

Now let's prove that if statement Y_n is true for $n = k$ (where k is any integer from 1 to 99), then it's true for $n = k + 1$.

Put $k + 1$ cards on the table, figures up. Take away one card for the time being (let's call it A). Then k cards remain on the table. By the induction hypothesis, the same figure (some figure X) is written on all of them.

Thus, the figure X is written on all the cards except, maybe, card A.

Now, replace one of the k cards on the table with card A. Again, k cards are on the table, so the same figure (which we've denoted by X) is written on all of them by the inductive hypothesis. In particular, the figure X is written on card A, which was removed the first time and has now returned to the table.

Thus, the figure X is written on card A, too.

So the figure X is written on all n cards ($n = k + 1$), and the proof is complete.

In the terms introduced in section 1, statement Y_1 is true, and every true statement Y_k ($1 \le k \le 99$) is followed by the true statement Y_{k+1}. Consequently, all statements Y_n are true ($1 \le n \le 100$). And as amazing as our trick with cards is, now it's always bound to succeed—it's been proven!

Art by Pavel Chernusky

3. Using our fingers

> Out of the sack tumbled a large, very angry two-toed sloth.
>
> —Gerald Durrell, *Three Tickets to Adventure*

By an argument similar to that in section 2, one can prove quite a number of surprising theorems. For instance, it isn't difficult to establish that *all numbers are equal to each other* or that *all girls have the same color eyes.*

I think this last statement is worth dwelling on. So *let's bring together n absolutely random girls. Then* (we'll prove it in a minute) *each and every one of them will have the same color eyes.*

For $n = 1$ this statement is obvious (though insipid). I'll explain the transition from $n = k$ to $n = k + 1$ literally "on our fingers." To do this, let's take $k = 4$, $k + 1 = 5$ (since I—and probably you—have five fingers per hand.) Now we can "play" the proof, following the illustration below and replacing the girls with our fingers.

According to the induction hypothesis the eyes of any four girls are the same color. Bring together five completely random girls A, B, C, D, and E (hand number 1 in the illustration). Then any four of them have the same color eyes. In particular, the eyes of all the girls except A are the same color as those of C (hand 2), and the eyes of all the girls except E are the same color as those of C (hand 3). So the eyes of any five girls are the same color.

For greater values of k the proof proceeds unchanged (except that if you want to play it "on your fingers," you might have to ask some friends to help out).

Exercises

1. If you doubt the validity of the statement proved above, "you can try the experimental approach by looking into the eyes of some girls" (George Polya, *Mathematics and Plausible Reasoning*).

2. If the name of the biologist and writer Gerald Durrell wasn't familiar to you before, then the epigraph to section 3 has already played its primary role: I strongly advise you to get acquainted as soon as possible with Durrell's books, which are interesting and full of his own brand of gentle humor. In choosing the epigraph I had another end in view, though, besides pure propaganda. What do you think it was?

(Hint: the epigraph is related to the title of section 3, to its contents, and to the contents of section 2. Any associations with the title of section 4, however, are coincidental.)

4. Lazy sages

> Three ladies A, B, and C with dirty faces are sitting in a compartment of a railway car, and all three are laughing. Suddenly, A thinks: "Why doesn't B understand that C is laughing at her?—Oh, my Lord! They're laughing at me."
>
> —J. E. Littlewood, *A Mathematician's Miscellany*

N sages are traveling in a railroad car. A pretty landscape lies beyond the windows. From time to time the train dives into a tunnel, taking one's breath away. All the sages have gathered in the corridor, looking out the window—just looking, looking . . .

Suddenly in one of the tunnels there's a roar, and smoke, and dust! Dirt is pouring through the windows. After they leave the tunnel, the conductor comes in. "Somebody here has gotten dirty," he says. "Unfortunately, we don't have any water on the train. But we'll be having some long stops now, so it'll be possible to get off the train and wash up."

Now, I should tell you there was little to choose among our sages: all were as wise as they were lazy. None of them would go get cleaned up for no good reason (if he didn't know for sure he's dirty). Nor would he ask anyone if his face is clean or dirty—why should he bother other people and trouble himself? It's easier to figure it out.[1]

So what will the sages do after the conductor's announcement?

I claim that *if n of them have dirty faces, then all these n sages will get off the train to wash at exactly the nth stop.*

We'll prove this statement by induction.

The case n = 1. For $n = 1$ the statement is obvious. Sage S_1 with a dirty face learns from the conductor's announcement that there are passengers with dirty faces in the car. Having

[1]Perhaps one more feature of our decent sages should be mentioned: it's true they're lazy, but not so lazy as to stay dirty in the train when there's an opportunity to go out and wash up!—*Ed.*

taken a look at those around him, he discovers that their faces are clean. So the dirty face must be his. He gets off to wash at the first stop.

Transition from $n = k$ to $n = k + 1$. Let's show that if the statement is true for $n = k$, then it's true for $n = k + 1$.

Let sages S_1, ..., S_{k+1} have dirty faces. Then S_{k+1} sees k dirty faces (S_1, ..., S_k) around and thinks: "There are two possibilities:

1. My face is clean.
2. My face is dirty.

In the first case all the k sages with dirty faces that I see will get off to wash at the kth stop, since by the induction hypothesis the statement is true for $n = k$.

"Since the first case is possible, I must get off neither at the kth stop nor earlier: if I am clean, it would be an unpardonable waste of energy. *Any intelligent person not given to pointless bustling*[2] *would come to the same conclusion if they were in my place.*

"But of course the second case is possible as well: my face might be dirty too. But then each of the soiled sages S_1, ..., S_k sees k dirty faces around him. In this situation none of these wise and unhurried sages will get off to wash at the kth stop (see the previous paragraph).

"So, if sages S_1, ..., S_k don't get off to wash at the kth stop, then I am dirty and I must go wash. I'll wait until the kth stop. If nobody gets off to wash, I'll have to get off at the next—$(k + 1)$th—stop and wash my face!"

All the sages with dirty faces reason identically. And so all of them will get off to wash at the $(k + 1)$th stop, completing the proof.

Problems

1. Let's change our story slightly.

Since he knows that the people traveling in the train are *sages*, and since he's seen that *many* of them are soiled, the conductor decides to shorten his announcement.

"Why should I say that somebody is soiled," he thinks, "when they can see for themselves?" So he skips the first phrase of the announcement.

Can we maintain as before that if exactly n people have dirty faces, then they'll get out to wash at the nth stop?

2. Let's change the story another way.

Suppose that just when the train was passing through this nasty tunnel, some of the sages were in their compartments—looking out the window, taking a nap . . .

By a lucky chance, everybody heard the loud voice of the conductor as he made his announcement in full, but no one could guarantee that the others heard it too. Some time later (even before the first stop) all the sages gathered in the corridor . . .

The question is the same as in problem 1.

Warning. If you want to solve these problems yourself (without hints), don't read section 5 for now.

5. What did the conductor say?

"... He thought I thought he thought I slept."

—Coventry Patmore, "The Kiss" (in *The Angel in the House*)

Imagine now that our sages on their own—without the conductor's telling them—knew that there was no water on the train and that after the tunnels there'd be a series of long stops where one could wash up. Suppose also that more than one person got soiled in the dusty tunnel. Then the sages were apparently none the wiser after the conductor's announcement!

So what then? If the conductor didn't come by, would the n soiled sages go to wash at the nth stop?

The temptation is strong to answer "yes": since nothing has changed in the problem's condition, the answer must not change either—or so it seems. But common sense tells us that without the conductor's information, none of the lazy sages is likely to go out to wash! And then again, what does that mean: "nothing has changed in the condition"? After all, in one of the variants the conductor didn't show up at all, while in the other he came and said something!

So, what did he say?!

Apparently this question can't be answered by exclamations alone. So let's coolly and scrupulously analyze the simplest possible case of $n = 2$ (we can't set $n = 1$, since it's given that more than one person got dirty).

Case $n = 2$. The faces of sages S_1 and S_2 are dirty. They see each other, and therefore *each of them indeed knows that someone is dirty.*

Let's take the point of view of S_2 and repeat the reasoning of section 4.

"There are two possibilities:

1. My face is clean.
2. My face is dirty.

In the first case S_1 sees only clean faces, but he *knows that someone got dirty*. Therefore . . ."

Hold on! Yes, in fact, S_1 knows that someone got dirty, but how does S_2 know about it? *How does S_2 happen to know S_1 knows someone got dirty?*

It seems at last we've stumbled onto a clue.

If the conductor had come by, his information wasn't news to S_1 and S_2—that's true. But S_2 saw S_1 listening to the announcement. And that's why, if the conductor had come by, then S_2 *knows S_1 knows someone got dirty.*

But if the conductor didn't come by, *there's no way for S_2 to learn about it.* Indeed, *we* know that S_2's face is dirty, and so we know S_1 knows someone (maybe just S_2, but maybe he too) has gotten soiled. But S_2 has to allow that his face might be clean, which means that S_1 sees only clean faces. And if S_1 doesn't see any dirty faces, and the conductor didn't come by, then *there's no way for S_1 to learn that someone is dirty.* Therefore, if the conductor didn't come by, then *S_2 can't learn that S_1 knows someone is dirty.*

Exercise. In the case $n = 2$ suppose sages S_1 and S_2 got dirty, sage S didn't, and the conductor didn't come by. Which of the following statements is true?

(a) S_2 knows S knows someone

[2]Evidently this is how he thinks of his own laziness.

got dirty.

(b) S knows S_1 knows someone got dirty.

(c) S_2 knows S knows S_1 knows someone got dirty.

(d) S_2 knows S_1 knows someone didn't get dirty.

Further analysis. When $n > 2$, each sage, even without the conductor's announcement, learns not only that someone got dirty but also that all the rest know about it. For instance, let there be three dirty sages S_1, S_2, S_3. Then S_2 knows S_1 sees S_3's dirty face, which means that S_2 *knows* S_1 *knows someone* (say, S_3) *got dirty.*

The difference between the two possibilities—did or didn't the conductor come—becomes even more subtle for $n > 2$. Let's give its exact formulation for $n = 3$:

Case n = 3. If the conductor came by, then S_3 *knows* S_2 *knows* S_1 *knows someone got dirty*; if he didn't, then S_3 *can't find out about it.*

Exercise. Verify the last statement.

General case. A famous philosopher once noted, "If a person says the phrase 'I'm thinking about how I'm thinking about . . . ,' then at the third or fourth iteration the speaker loses the sense of what's being said."

In the general case we'll have to deal with not just three or four but with n iterations. Denote by U_n the following statement: S_n *knows* S_{n-1} *knows . . .* S_2 *knows* S_1 *knows someone got dirty.*

Then the difference between the two variants being compared can be formulated as follows:

Suppose the conductor had come by and sages S_1, ... S_n *(and maybe somebody else) have dirty faces. Then statement* U_n *is true.*

Suppose the conductor didn't come by and the faces of sages S_1, ..., S_n *(and no one else's) are dirty. Then statement* U_n *is false.*

We'll prove this theorem by induction.

For $n = 1$ it's obvious: if the conductor had come by, S_1 knows from his announcement that someone got dirty; if the conductor didn't come by and only S_1 has a dirty face, then there's no way for him to know about it. So in the first case U_n is true, and in the second case it isn't.

The transition from $n = k$ to $n = k + 1$ is performed as follows.

First case (the conductor came by). The faces of S_1, ..., S_{k+1} are dirty. This means, in particular, that the faces of S_1, ..., S_k are dirty, and we can apply the induction hypothesis, according to which statement U_k is true. Sage S_{k+1} know this, of course—that is, S_{k+1} *knows* S_k *knows . . .* S_2 *knows* S_1 *knows someone got dirty.* Which is to say, the statement U_{k+1} is true.

Second case (the conductor didn't come by). Sage S_{k+1} may assume that his face is clean. If so, then *from the point of view of* S_{k+1}, only k sages S_1, ..., S_k have dirty faces. Therefore, by the induction hypothesis, statement U_k is false. Thus, having assumed that his face is clean, S_{k+1} is compelled to assume that S_k *doesn't know whether* S_{k-1} *knows . . .* S_2 *knows* S_1 *knows someone got dirty.*[3] In other words, S_{k+1} *doesn't know whether* S_k *knows* S_{k-1} *knows . . .* S_2 *knows* S_1 *knows someone got dirty.* And this means that statement U_{k+1} is false.

The proof is complete.

At the same time we've finally figured out what the conductor said. The larger n is, the more the sages learned from his announcement (that is, "those that have ears to hear").

6. Dotting the i's

The reader who has managed to struggle through all these word mazes to the end of section 5 will hardly need the following explanations. Still, for the sake of completeness, here they are.

1. In section 3, the transition from $k = 4$ to $k + 1 = 5$ is absolutely correct. It's also true that for greater k's the proof holds without alterations. And it's also easy to pass from $n = 2$ to $n = 3$ and from $n = 3$ to $n = 4$. "Only" one transition doesn't work: from $n = 1$ to $n = 2$. It can't work because for $n = 2$ the statement is wrong: any two girls don't have eyes of the same color.

But if, nevertheless, you try to repeat the proof "on your fingers" for this case, nothing will come of it (in section 3 we joined A and E by a "bridge" C: both A and E happened to have eyes the same color as C's; if there are only two "fingers," there's no such bridge).

The *two-toed* sloth from the epigraph was supposed to draw your attention to this exclusive case.

2. In section 2, statement Y_1 is true, whereas all the statements Y_n starting from the second are false. The transition from $n = k$ to $n = k + 1$ was correctly established only for $k \geq 2$. It fails for $k = 1$: when card A is removed from the table, only one card remains; and only one figure is written on this card, but it's not necessarily X.

3. Problem 1 in section 4 is virtually equivalent to the problem examined in section 5, if we put aside the question of whether S_2 knew that S_1 knew that one can wash up at a stop and not in the train (and if he did, then the question is whether S_3 knew about it, and so on). If the conductor didn't announce that "someone has gotten dirty," none of the sages will go out to wash (neither at the nth nor at any other stop).

Nobody will go out to wash up in the setting of problem 2 from section 4 either—that is, in the case when $n > 1$ and *not all of the n soiled sages are together* when the conductor makes his announcement. The proof is similar to that from section 5 except that induction must begin from $n = 2$. (For $n = 1$ the only soiled sage will go out to wash up at the first stop no matter where he was when the announcement was made.)

4. In the first exercise from section 5, statements (a), (b), and (d) are true; statement (c) is false.

5. Along with statement U_n (which is true if the conductor had come by and false if he didn't) one can consider a whole set of $n!$ $(1 \cdot 2 \cdot \ldots \cdot n)$ statements obtained from U_n by all possible permutations of the letters S_1, ..., S_n. Any of these statements can also be used to tell one of the two variants compared in section 5 from the other. For instance, we could work with the statement "S_1 knows S_2 knows . . . S_n knows someone got dirty."

[3]In fact, S_k knows about it; that is, in actual fact—and not from the point of view of S_{k+1}, who supposes his face to be clean—statement U_k is true. Check it yourself.

Each of these $n!$ statements consists of n iterations. Shorter statements (of less than n iterations) aren't enough to distinguish between the variants in which the conductor did or didn't come by. For instance, U_{n-1} is true even without the conductor's announcement, since S_{n-1} knows S_{n-2} knows . . . S_2 knows S_1 knows S_n got dirty.

6. The amused lady from the epigraph to section 4 figured out her face was dirty even though there was no conductor to make any announcements. The part of the conductor was played by the unrestrained laughter of her companions. Staid sages, of course, would never take such unseemly liberties!

SMILES

Extra! Extra! Read all about it!

New result in ancient theory!

TWO ARTICLES IN THIS ISSUE OF *QUANTUM* are devoted to the method of mathematical induction. Recently I came across a miraculous application of this method to elementary number theory: I proved that *if the product of several integers is divisible by some integer, then at least one of the factors in the product must be divisible by this integer.*

Here's the proof of this theorem by the method of induction in its "strong" variant (see examples 2 and 5 in the article "Jewels in the Crown" in this issue).

Let N be the number of factors in the product. For $N = 1$, there's nothing to prove. Suppose the theorem is true for all values of N less than k. For $N = k$, consider the product $a_1 a_2 \ldots a_{k-1} a_k$ of integers, which is divisible by some number b. Let $a = a_1 \ldots a_{k-1}$; then aa_k is divisible by b, and by the induction hypothesis (for $N = 2$) either a_k is divisible by b (and we're done) or a is divisible by b; then, by the hypothesis, for $N = k - 1$ one of the number $a_1, \ldots, a_{k-1}$ is divisible by b, completing the proof.

Everything would be wonderful if not for a "minor disagreement with experiment"—the number 6 divides the product $3 \cdot 4 = 12$ but doesn't divide any of the factors 3 and 4. What went wrong: the inductive method? the proof? Or, maybe, we should revise all our notions of divisibility . . .

Think about it. Then go ahead and look for the answer somewhere in this issue. (Can you imagine where?)

—V. Dubrovsky

Art by Sergey Ivanov

A wrinkle in reality

"Space is the universal form of the existence of matter and, consequently, when the problem of the properties of space is raised, no domain of facts can be separated artificially."
—A. D. Alexandrov, "Non-Euclidean Geometry," *in* Mathematics: Its Contents, Methods and Meaning

by Yuli Danilov

FOR MORE THAN TWO MILLENIA ONLY ONE geometry was known: the geometry so nicely laid out in Euclid's *Elements.* Euclidean geometry was considered the geometry of real space because there simply wasn't any other geometry, and the *Elements* were taken as a model of scientific exposition worthy of imitation. For example, Spinoza wrote his *Ethics* in a "geometrical manner"—that is, he imitated the style and structure of Euclid's famous treatise.

Later a lot of weak spots were discovered in the *Elements,* and this is quite understandable: what is considered a strict result depends on the historical epoch. Great controversy arose in connection with famous "parallel postulate." Many mathematicians tried to deduce this "stubborn" postulate from the other axioms and postulates, but in vain. Step by step a set of geometrical theorems was found that didn't depend on the parallel postulate and therefore must be valid whether or not the dubious postulate is correct. Such theorems constitute so-called *absolute geometry*—that is, geometry that doesn't depend on the postulate on parallels.

At last two noncontradictory versions of non-Euclidean geometry (geometry without Euclid's postulate on parallels) were created. One was created by the Hungarian officer János Bolyai (1802–1860), who was the son of a friend of the "prince of mathematicians" Karl Friedrich Gauss. The other was the brainchild of the Russian mathematician Nicolay Lobachevsky (1792–1856).

The following fragment is taken from one of Lobachevsky's papers on his new geometry, which he called "imaginary." Its publication can be considered our modest tribute to the memory of a great scientist as his 200th birthday approaches (December 1).

The world was slow to acknowledge the creators of non-Euclidean geometry. The idea of the existence of "other" geometries was too bizarre, almost unbearable. (By the way, Gauss confessed in a letter to Bolyai's father that he also developed his own approach to non-Euclidean geometry but never published it because he was afraid of "rousing the Boeotians" (the inhabitants of a province in ancient Greece whose mental endowments were thought to be rather meager).

The discovery of non-Euclidean geometry put forward some very deep and important questions. What is geometry? How many different geometries are possible? Which geometry corresponds to the real world? The answers will be given in later installment of this department.

Lobachevsky called his geometry "imaginary" in order to distinguish it from the usual Euclidean geometry. Using astronomical observations, he showed that Euclidean geometry is valid with very great accuracy. But Lobachevsky's "imaginary" geometry governs some surfaces—for example, the *pseudosphere,* which resembles a funnel (see the illustration on page 46)—and is the "real" geometry for high-speed elementary particles. But that's another story.

Translation and notes by Yuli Danilov

New elements of geometry with a complete theory of parallels

(Excerpt)

It is generally known that the theory of parallels in Geometry has up to now remained incomplete. The fruitless efforts since Euclid's time, over the course of two thousand years, has led me to suspect that the notions themselves do not contain the truth that people have tried to prove and that, like other physical laws, can be verified only by experiment—for example, by Astronomical observations. Convinced at last of the validity of my guess, and considering this difficult problem to be completely solved, I wrote a dissertation on this topic in 1826.[1] The application of this new theory to analytics can be found also in papers entitled "On the Elements of Geometry" published in the *Proceedings of Kazan University* for 1829 and 1830. The main conclusion, at which I arrived by supposing that lines are dependent on angles, allows the existence of Geometry in a broader sense than was first presented by Euclid. I named the science in this extended form *Imaginary Geometry*, which includes, as a particular case, *Practical Geometry* with its restriction in the usual assumptions demanded by actual measurement. I undertook to prove the sufficiency of the new elements[2] in a work not long ago published in the *Proceedings of Kazan University*. Wishing to achieve this goal if not by a direct path then at least by the shortest backward path, I preferred at that time to proceed from hypothetical foundations to equations for all the relations and to expressions for any Geometrical quantity. If my discovery brought no other benefit than repairing a deficiency in the original doctrine, then at least the attention continuously paid to this subject obligates me to present a detailed account. I will begin by examining previous theories.

It is easy to prove that two straight lines, inclined at the same angle to a third line, never meet, as they are therefore perpendicular to one line. Euclid supposed, on the contrary, that two lines inclined unequally to a third line must always intersect. In order to convince themselves of the validity of this last sentence, people tried different means, either attempting to find in advance the sum of the angles in the triangle, or comparing infinite planes in the apertures of the angles and between the perpendicules,[3] or allowing the dependence of the angles only on the content of the sides,[4] or, finally, giving new properties to the straight line in addition to its definition. Some of these proofs might be called ingenious, but all of them are false—insufficient in their foundations and lacking the required rigor; there is not one among them that, combining simplicity with conclusiveness, could be recommended to beginners.

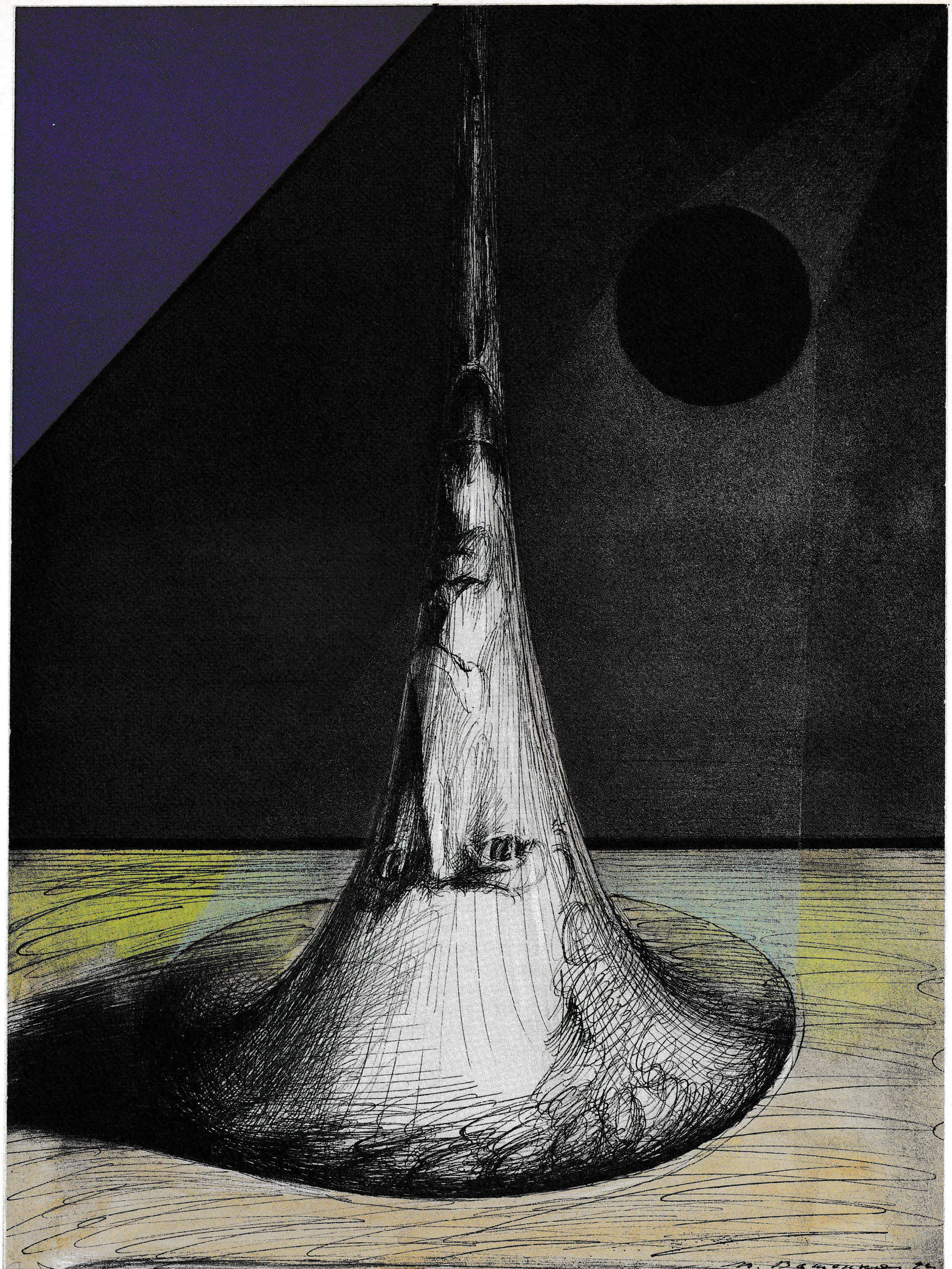
In 1800 Legendre published the third edition of his Geometry, in which he included a theorem that the sum of the angles in any triangle could not be more than π—that is, two right angles. There he proved also that the sum of the angles could not be less than π, but he lost sight of the fact that the lines could pass each other without making a triangle and at the same time the value of the sum of the angles, derived by the other method, would be some sort of absurdity. That is why I do not consider it necessary to expand here upon this error, which Legendre himself admitted later, saying that, while the chosen foundations are beyond question, he met some obstacles that he could not overcome. In the *Compte rendus* of the French Academy in 1833, he added a theorem that the sum of the angles must be π in all

[1]The first version of Lobachevsky's *Geometry* was completed in 1823, but the authorities of Kazan University did not allow its publication. On February 23, 1826, Lobachevsky delivered a report entitled "A Brief Statement of the Elements of Geometry with a Rigorous Proof of the Theorem on Parallels" at a session of the physics and mathematics department, but this report, too, was denied publication. It wasn't until 1829 that the work excerpted here was printed in the *Proceedings of Kazan University*.

[2]Here "beginnings" is translated as "elements," just as the term was translated into English from Euclid's Greek; elsewhere, as context demands, it is translated as "foundations."

[3]"Perpendicule" is an older version of the term "perpendicular."

[4]The "content of a side" is here used to mean the length of a side.



triangles if it is π in some single triangle. It had been necessary for me to prove the same thing in my theory, which I wrote about in 1826. I even think that Legendre had several times found himself on the very path I had chosen with such success; but prejudices in favor of the generally accepted assumptions no doubt made him continually force a conclusion or fill the gaps in a way that was inadmissible even with the new assumption. . . .

There cannot be any contradiction in our minds when we suppose that some forces in nature follow one Geometry and others follow their own special Geometry.

There arose the idea of accepting as a foundation in the theory of parallels the notion that the angles in triangle must depend on the content of the sides. At first glance such a hypothesis seems as simple as it is necessary; but when we delve into our notions and learn what they are based on, then we are forced to call them arbitrary as all the others that were embraced earlier. In nature we directly comprehend only motion, without which we could comprehend nothing through our senses. So all other notions—for example, geometrical notions—are created by our mind artificially and are taken to be aspects of motion; therefore, space by itself, separately, does not exist for us. So there cannot be any contradiction in our minds when we suppose that some forces in nature follow one Geometry and others follow their own special Geometry.[5] In order to make this idea clearer, let us suppose that, as many people believe, attractive forces weaken as they propagate along a sphere. In practical Geometry the area of the sphere is taken to be $4\pi r^2$ for a semidiameter r, so the force must diminish in content as the inverse square of distance. In imaginary Geometry I found the surface of the sphere to be

$$\pi(e^r - e^{-r})^2,$$

and it may be that molecular forces, whose diversity will depend on the number e (always extremely large), obeys such a Geometry.[6] . . .

If the difficult problem of parallelism must be solved experimentally, the method proposed by Legendre—laying a semidiameter six times around a circle—without any doubt must be considered insufficient. In my *Elements of Geometry* I proved, using Astronomical observations, that in a triangle with sides as large as the distance from the Earth to the Sun, the sum of the angles cannot differ from two right angles by more than 0.000003 second of a degree. This difference varies geometrically with the sides of the triangle, and therefore the practical Geometry used previously, as I mentioned before, is more than adequate for actual measurements. One can come to such a conclusion by means of propositions that are simple enough and consistent with the foundations of the science, although a complete theory demands that the sequence of education be changed, and that Trigonometry be added here.

Among the shortcomings of the theory of parallels is the definition of parallelism itself. But, contrary to Legendre's suspicions, this shortcoming does not depend in any way on any defect in the definition of a line, nor on those defects, I would add, that were hidden in the original notions and that I intend to point out here and try, to the extent I can, to correct.

One usually begins Geometry by giving three extensions to bodies, two extensions to surfaces, one extension to lines, and no extension to a point. Calling the three extensions length, width, and height, and taking these names to mean three coordinates, people hurry to communicate this premature notion with words that everyday language gives a certain meaning, but one that is indefinite for exact science. Indeed, how can one clearly imagine measuring a length without knowing what a straight line is?[7] How can one say anything about width, or height, without saying a word in advance about perpendicules, or planes, or about perpendicules in one plane and in different planes? Finally, if there is no extension at all in a point, then what remains in it such that a point could be the subject of consideration? Let us say that every person clearly imagines a straight line without giving any account of what a line is; but how, using a straight line, is one now to designate one extension in a curved line and two extensions in a curved surface?

It is true that there is no need to require that length, width, height be mutually perpendicular: it is sufficient to take them as being lines in different directions. But this case presents difficulties it its own. Keeping to the rule of not borrowing prematurely from those notions that must be developed later, the question arises: how are we to express the requirement that the three dimensions of bodies belong to three straight lines in different planes? In addition, the different directions of two segments from the point where the line breaks must not be confused with a double extension in a plane. And, finally, how are we to define adequately what we mean by "direction" or "angle"? In sum: space, extension, place, body, surface, line, point, direction, and angle are the words with which

[5]These words can be considered prophetic: as is well known now, the space of relativistic particles—that is, particles moving at velocities close to the speed of light—is governed by Lobachevsky's geometry.

[6]Here Lobachevsky is not using the letter e to denote the base of the system of natural logarithms; rather, e is a number related to a circle's radius r by the formula $r = 1/\ln e$.

[7]The shortcomings of the "naive" Euclidean definition ("a line is length without width," and so on) were known long ago. Constructive criticism at the beginning of the 20th century brought to light such unusual objects as the Peano curve, which passes through all the points of a square, and curves that are nowhere differentiable, among many others.

Art by Yuri Vashchenko

Geometry begins, but no clear understanding is ever attached to them

It is possible, however, to view all these objects from another direction. One must keep in mind that the obscurity of the notions here is caused by an abstractness that is superfluous when they are applied in actual measurements and, therefore, was introduced into the theory for no good reason. Surfaces, lines, and points, as they are defined in Geometry, exist only in our imagination; but when we perform actual measurements of surfaces and lines, we use bodies. That is why we must speak about surfaces, lines, and points only as they are understood in actual measurement, and then we can hold those notions that are directly linked in our minds with the notion of bodies, to which our imagination has become accustomed, and which we can directly verify in nature, without embracing other notions that are artificial and extraneous. But with these new notions the science acquires a new direction from the very beginning, which it follows until it turns into analytics. So the manner of teaching takes on a quite different aspect. I will try to explain what sort of transformation this is.

The obscurity of the notions here is caused by an abstractness that is superfluous when they are applied in actual measurements.

There are two approaches in Mathematics: analysis and synthesis. The distinguishing feature of analysis consists of the equations that serve as the first foundation for any assertion and lead to all conclusions. Synthesis, or method of construction, requires that very representation that is directly connected in our minds with the first (that is, fundamental) notions. The main benefit of analysis is that, starting from equations, one always moves directly to the proposed goal. Synthesis is not subjected to any general rules, but one necessarily has to start from synthesis in order to reach, having found an equation, that borderline after which everything turns into the science of numbers. For example, one proves in Geometry that two perpendicules do not intersect; that if some parts of triangles are equal, then the triangles are identical. But it would be vain to try to consider such cases, or the entire theory of parallels, analytically. Such an approach will never be successful, just as one cannot avoid synthesis in measuring planes delimited by straight lines, or in measuring bodies delimited by planes. It stands to reason that in synthesis one must turn to analysis for help; nevertheless, it is indisputable that analysis can never be the only tool in the foundations of Geometry and Mechanics. To a certain extent Geometry will always contain something properly geometrical that cannot be separated from it. One can restrict the scope of synthesis, but it is impossible to eliminate it completely. But even in this attempt to replace synthesis with analysis, one must not be so hasty as to introduce functions any time when it is only possible to foresee a dependence, not knowing what it consists of, let alone how it will be expressed. With this restriction on analysis we designate the true goal and proper place for another method that alone will found the science on such notions, from which reasoning draws all the rest, deducing from new data from the initial data and enlarging the limits of our knowledge infinitely in all directions. The initial data will undoubtedly be those notions we acquire in nature through our senses. Our mind can and must reduce them to the smallest number so that they could serve as a solid foundation for science. But usually no one follows the synthetic approach in this form, obeying all the rules mentioned here; people prefer to bring in analysis, even if prematurely, and to assume the development, albeit incomplete, of the notions that constitute our natural mind and that only need be given names, without going into broad explanations and not bothering with precision in defining them. But if ease and simplicity induce us to choose such a method of instruction, then strict truth will always have its own advantage, which it is sometimes necessary to use.

Vavilov's paradox

On paper, it's a violation of the energy conservation law

by V. A. Fabrikant

THE NOTION OF PARALLEL light beams with finite cross sections is frequently and fruitfully applied in geometrical optics. Not only that, even in the theory of such a wave phenomenon as interference it's permissible in many cases to make use of this notion.

In many cases, yes, but not in every case. A very instructive optical paradox is discussed in *The Microstructure of Light* by the brilliant Russian physicist S. I. Vavilov (1891–1951).

Let me just remind you briefly of what an interference pattern looks like and the energetics involved. Interference, as we know, results from the superimposition of waves. As far as electromagnetic waves are concerned, both the electric and magnetic fields oscillate. At every point in space and at every moment in time these quantities determine the energy of the electromagnetic field. An electromagnetic wave transfers energy, and we can introduce the idea of the density of the energy flux. This is the name we give to the energy of a field passing through a unit area per unit time.

Every wave is also characterized by its phase. If two light beams have a constant phase difference, we say that the beams are coherent.

In the interference pattern that arises when two coherent light beams are superimposed, the light energy is spatially redistributed. In the bright bands the energy is greater than the sum of the energies of the component

Art by Dmitry Krymov

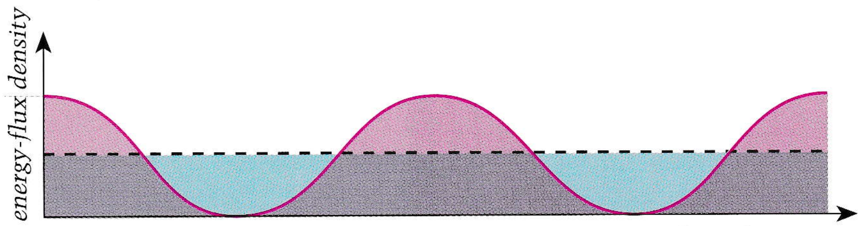


Figure 1

beams; in the dark bands, the opposite is true. The excess energy in the bright bands is compensated by the lower energy in the dark bands. The total energy distributed throughout the entire interference pattern is equal to the sum of the energies of the two interfering beams.

Figure 1 shows the dependence of the energy-flux density in the interference pattern on the displacement along the screen on which it is observed. This pattern was obtained with two coherent light beams of equal energy. The dashed line indicates the sum of the energy-flux densities of the two beams. The portions of the curve above this dashed line correspond to the bright interference bands, and the portions of the curve lying below the line correspond to the dark bands. The total energy distributed throughout the interference pattern is given by the area under the curve. This area is equal to the area under the dashed line. The requirement of nature's strict "bookkeeper"—the law of conservation of energy—is unswervingly fulfilled.

Now let's move on to Vavilov's paradox. Imagine two absolutely coherent, narrow light beams of width a that intersect at a small angle α (fig. 2). The area $ABCD$ is where the interference occurs.

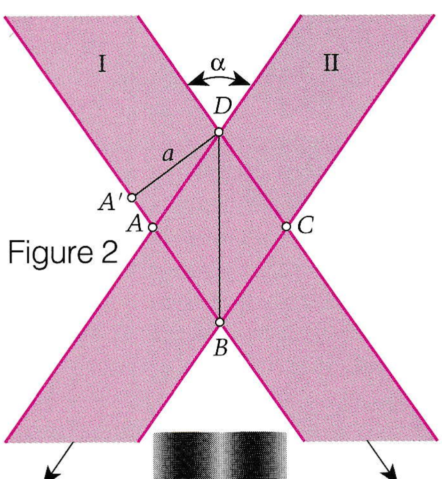


Figure 2

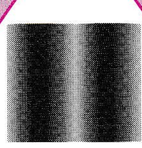


Figure 3

In order to observe the interference pattern, we place a screen perpendicular to the plane of the drawing and passing through points A and C. The interference pattern will consist of alternating straight bright and dark bands filling the screen from point A to point C (fig. 3).

The distribution of the energy-flux density corresponds to the diagram in figure 1. If at point D (fig. 2) both beams have the same phase, the phase difference at points lying along the line BD will be equal to zero. So line BD corresponds to the middle of the central bright band. In the middle of the neighboring dark band, the phase difference must be equal to π—that is, the light oscillations in both beams must be completely out of phase. The phase difference $\Delta\phi$ is equal to the path difference ΔL for both waves to the given point divided by the wavelength λ and multiplied by 2π:

$$\Delta\phi = 2\pi\frac{\Delta L}{\lambda}. \qquad (1)$$

A phase delay of 2π corresponds to a path difference of λ. From formula (1) it follows that, in the middle of the interference pattern of the dark band nearest the center, ΔL must be equal to $\lambda/2$.

Let's calculate the path difference at point A. The parallel beam can be considered to consist of plane waves perpendicular to the direction of the light beam. Now let's draw one of the wave surfaces (whose phases are equal) of the first beam through point D (fig. 2). Points D and A' lie on this wave surface.

The paths taken by both beams to point D are equal. In order to get to point A, the wave surface of the first beam must travel the extra distance $A'A$ after passing through point D. The wave surface of the second beam must travel the extra distance DA after passing through point D. So we obtain the path difference

$$\Delta L = DA - A'A = \frac{a}{\sin\alpha} - \frac{a}{\tan\alpha}$$

$$= \frac{2a\sin^2\frac{\alpha}{2}}{\sin\alpha} = a\tan\frac{\alpha}{2}. \qquad (2)$$

It's clear that the same path difference, but with the opposite sign, will hold at point C. Now let's begin to decrease angle α. As we can see from formula (2), ΔL will decrease as well. For small α the path difference ΔL can become equal to $\lambda/4$. Then the entire region from point A to point C will be filled with a single bright interference band. Consequently, the energy will everywhere exceed the sum of the energies of the two intersecting beams. There is no compensation from the formation of dark bands, since there aren't any!

We can also obtain a negative result, so to speak, if we take intersecting beams with an initial phase difference of π. Then the region from A to C will be filled with a dark interference band. In the first case, it's hard to understand where energy comes from; in the second case, where it disappears to.

Both cases clearly contradict the law of conservation of energy. There must be some defect in our reasoning that causes this apparent violation of one of the basic laws of nature. To understand what's going on here, let's write formula (2) for the case when $\Delta L = \lambda/4$ and use the fact that α is small ($\tan\alpha \cong \alpha$ for small α):

$$\frac{\lambda}{4} = \frac{a\alpha}{2}. \qquad (3)$$

We'll show that in order to fulfill the condition $\Delta L = \lambda/4$, the angle α must indeed be very small.

CONTINUED ON PAGE 62

STUDENT CORNER

Summing up

The possibilities of one-time digital use

by Mark Lucianovic

SOME TIME AGO I WAS reading an old computer magazine and came across the following contest problem: *Write a computer program that will find all of the valid addition expressions of the form $a + b = c$ such that each digit from 1 to 9 is used exactly once.* The expression $124 + 659 = 783$ is an example. Actually, the author was looking for the *fastest* program that would print all valid expressions. However, I was struck by the mathematics of the problem. Writing a brute-force algorithm to find all such expressions didn't interest me as much. I worked on the contest question as a mathematical, not programming, challenge.

To find all of these expressions, we must begin by considering three forms:

$$\begin{array}{r} ABCD \\ E \\ \hline FGHI \end{array} \qquad \begin{array}{r} ABC \\ DE \\ \hline FGHI \end{array} \qquad \begin{array}{r} ABC \\ DEF \\ \hline GHI \end{array} \qquad (1)$$

where A through I represent different digits. It isn't difficult to see that the first two of these digits don't have a solution. So we're left with finding all solutions to $ABC + DEF = GHI$. Let's assume the first addend is always less than the second addend. This way, no equations will be double-counted because of commutativity.

I spent some time on this puzzle and came up with a surprisingly large number of expressions of the desired form. I then yielded to curiosity and wrote a program to confirm what I had found on paper. Indeed, I had discovered them all—there were 168 of them. (Actually, my program found twice as many, because it considered $a + b = c$ and $b + a = c$ to be different expressions.)

I thought the problem was solved. But I looked again and noticed some very peculiar properties. First of all, in each of the 168 expressions, the sum was divisible by 9. Can you prove this? (Hint: use the fact that a number and the sum of its digits leave the same remainder when divided by 9.)

Notice also that the 168 equations can be grouped in sets of four based on the sums. For the original example (1), the expressions

$$\begin{array}{r} 124 \\ 659 \\ \hline 783 \end{array} \qquad \begin{array}{r} 129 \\ 654 \\ \hline 783 \end{array} \qquad \begin{array}{r} 154 \\ 629 \\ \hline 783 \end{array} \qquad \begin{array}{r} 159 \\ 624 \\ \hline 783 \end{array} \qquad (2)$$

make a group. So once we have found one valid expression, three more can be generated.

I found another very interesting property, for which I haven't discovered a satisfactory explanation. In every single expression, carrying takes place *exactly once*. That is, it's always necessary to carry from either the ones to the tens column or the tens to the hundreds. Can *you* explain why?

With this fact in mind, we can actually place the 168 into groups of eight, not four. Here's how: for each expression, there is one column that doesn't involve carrying. (It's either the ones or the hundreds.) We can move this column to the other end and create four new expressions.

The eight-member group for the original example includes the four in (2) plus

$$\begin{array}{r} 241 \\ 596 \\ \hline 837 \end{array} \qquad \begin{array}{r} 246 \\ 591 \\ \hline 837 \end{array} \qquad \begin{array}{r} 291 \\ 546 \\ \hline 837 \end{array} \qquad \begin{array}{r} 296 \\ 541 \\ \hline 837 \end{array} \qquad (3)$$

Finally, why 168? I mean, what makes this number so special? Clearly, based on my other findings, the number of valid expressions must be divisible by 8. But can you find a method of predicting mathematically why there are this number of expressions? This seems to be a pretty simple problem, but there are so many intricacies involved. Please let me know if you can find a way to explain this.

Here are some other questions to consider.

- Each sum is, of course, divisible by 9. This means that the sum of its digits can be either 9 or 18. But in actuality, all valid sums have digits that add up to 18. Why?
- Some valid sums, like 783, have eight valid expressions. Four of these are listed in (2) above. Others, like 639, have only four. What causes this distinction?

At the time he wrote this piece, **Mark Lucianovic** *was a senior at Thomas Jefferson High School for Science and Technology in Fairfax, Virginia. He is now a student at Princeton University and also uses his digits to manipulate the valves of the French horn.*

MATH INVESTIGATIONS

Triangles of sums

And onward to pyramids

by George Berzsenyi

IN MY PREVIOUS COLUMN ("Triangles of Differences" on page 30 of the May/June issue of *Quantum*), I presented a problem as it originally occurred to me, allowing my readers to start from scratch. The purpose of this column is to restate that problem in a slightly different form, obtained by rotating the triangles of numbers to obtain "triangles of sums" rather than differences. Thus, for example,

```
6       10      15      23      43
    4       5       8       20
        1       3       12
            2       9
                7
```

becomes

```
7       2       1       4       6
    9       3       5       10
        12      8       15
            20      23
                43
```

One of the advantages of this new form is that the numbers in its first row are smaller and their influence on the number at the apex can be analyzed more efficiently. In fact, one can prove that if the first row consists of the sequence $n_1, n_2, \ldots, n_k$, then the number at the apex is given by the formula

$$N_k = \sum_{i=1}^{k} \binom{k}{i} n_i .$$

Verification of this fact is left to my readers as a warm-up exercise. Our original problem can then be restated as follows: **For each $k = 1, 2, 3, \ldots$, find the smallest value of N_k so that all numbers in the resulting triangle of sums are distinct positive integers.**

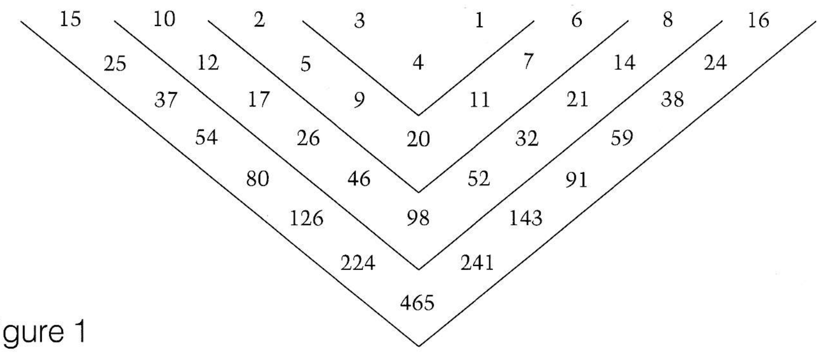


Figure 1

In view of the formula above, it is clear that the middle entries of the sequence $n_1, n_2, \ldots, n_k$ have a greater influence on the size of N_k than those on either end. Consequently, in search of such triangles one may wish to start with smaller positive integers and then add larger ones on each end. This leads to the idea of embedding smaller triangles (of numbers) in larger ones, and hence forming them recursively. Figure 1 exemplifies this development for even k; unfortunately, it is slightly flawed since N_2 is definitely 3 rather than 4.

For odd k, the recursion is flawless for the presently known values of N_k; they are $N_1 = 1$, $N_3 = 8$, $N_5 = 43$, $N_7 = 212$, and $N_9 = 1{,}000$, and it is conjectured that $N_{11} = 4{,}562$. **The second challenge is to verify this claim by using the witnesses given in "Triangles of Differences."**

Yet another advantage of switching to triangles of sums is the possibility of extending our inquiries to higher dimensions. One such extension is exemplified by figure 2, where each number is the sum of the four numbers at the corners of the square directly above it (for example, 17 = 9 + 2 + 1 + 5). **Your final challenge is to determine whether it is always possible to arrange the numbers 1, 2, 3, …, n^2 in the top layer so that the resulting "pyramid of sums" consists of distinct integers and the number at the apex is as small as possible.**

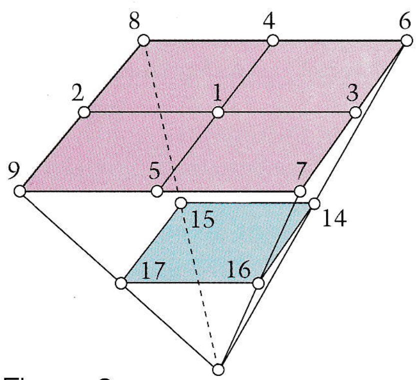


Figure 2

In closing, I wish to express my appreciation to my mathematical friends Basil Rennie, Stanley Rabinowitz, and Stan's former colleagues at Digital Equipment Corporation, for recasting my original problem in its present, much more trackable form.

INDEX

Volume 2 (1991—92)

Correction²

The Toy Store correction in the January/February issue should have referred to the *September/October* issue.

Thanks!

We'd like to thank the following friends of *Quantum* who reviewed manuscripts during the past year: William A. Hiscock, Physics Department, Montana State University; Meg Hall, a graduate student in physics at the same institution; and Al Rosenberg, Physics Department, Mississippi State University.

ANSWERS, HINTS & SOLUTIONS

Math

M56

(a) For any positive integer $a > 1$, the three pairs $(2a, 2a)$, $(a + 1, a(a + 1))$, and $(a(a + 1), a + 1)$ are all different and satisfy the given equation.

(b) We'll prove that, in general, the number of solutions for the equation $1/x + 1/y = 1/a$ is equal to the number of divisors of a^2. In particular, for $a = 1992 = 2^3 \cdot 3 \cdot 83$ this number is equal to $7 \cdot 3 \cdot 3 = 63$, because any divisor of 1992^2 can be represented as $2^k \cdot 3^n \cdot 83^m$, where k can take seven values 0, 1, 2, ..., 6; n and m can each take 3 values 0, 1, 2; and the value of each exponent is chosen independently.

To prove this statement, rewrite the equation in the form $ax + ay = xy$, or

$$xy - ax - ay = 0,$$
$$xy - ax - ay + a^2 = a^2,$$
$$a^2 = (x - a)(y - a).$$

It's not hard to see now that distinct ordered pairs of factors of a^2 give distinct pairs of solutions for x and y, which proves our assertion. The solution to part (a) of this problem can be obtained as a special case of part (b), if we think of a as a prime number.

Notice that the number of factorizations of a^2 into two factors equals the number of divisors of a^2. (M. Slavinsky, A. Vaintrob)

M57

The answer to both questions (a) and (b) is yes. Examples are shown in figure 1 for (a) and figure 2 for (b). In addition to the "propeller" and the ancient Chinese symbol yin-yang (fig. 1), whose areas are half the areas of the whole

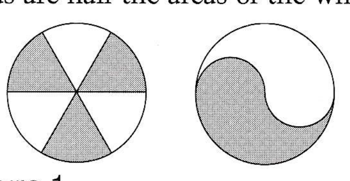


Figure 1

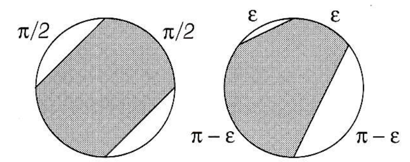


Figure 2

circles, the following degenerate example satisfies the problem's condition too: a circle of radius 1 and center O from which three vertices of an equilateral triangle with center O are deleted. In figure 2 the unit circle is divided into four arcs of angular measure ε, ε, $\pi - \varepsilon$, $\pi - \varepsilon$ (for the figure on the left, $\varepsilon = \pi/2$), and the "convex hull" of two nonadjacent arcs is constructed. Taking a small enough ε, we can make the area of this figure arbitrarily close to that of the semicircle.

Obviously, two copies of each of these figures can cover a unit circle. Suppose that some figure F covers the semicircle. Then it must cover the diameter of the semicircle. This means that the unit circle C from which F is cut (as shown in figures 1 and 2) must match exactly the unit circle from which the semicircle is cut. So F must contain a semicircle of the circle C, which is not the case. (N. Vasilyev)

M58

We begin with an algebraic proof.

(a) Substitute x for a_3 and consider the difference between the left and right parts of the inequality as a function of x:

$$f_3(x) = a_1{}^2 - a_2{}^2 + x^2 - (a_1 - a_2 + x)^2.$$

The quadratic terms x^2 cancel out here, so this function is actually lin-

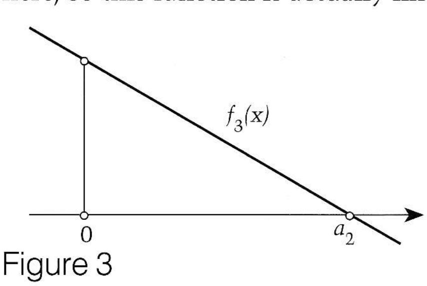


Figure 3

ear (its graph is sketched in figure 3). To prove $f_3(x) \geq 0$ for $0 \leq x \leq a_2$, it suffices to verify $f_3(0) \geq 0$ and $f_3(a_2) \geq 0$, which is easy:

$$\begin{aligned} f_3(0) &= a_1{}^2 - a_2{}^2 - (a_1 - a_2)^2 \\ &= (a_1 - a_2) \cdot 2a_2 \geq 0, \\ f_3(a_2) &= a_1{}^2 - a_1{}^2 = 0. \end{aligned} \qquad (1)$$

(b) Likewise, consider the function

$$f_4(x) = a_1{}^2 - a_2{}^2 + a_3{}^2 - x^2 - (a_1 - a_2 + a_3 - x)^2.$$

This is a quadratic function of the form $f_4(x) = -2x^2 + px + q$ (see the graph in figure 4). Again, the inequality $f_4(x) \geq 0$ for $0 \leq x \leq a_3$ will be proved if we show that it's valid at the endpoints of this interval. Now, $f_4(x) \geq 0$ is exactly inequality (a), and $f_4(a_3) \geq 0$ coincides with inequality (1) above.

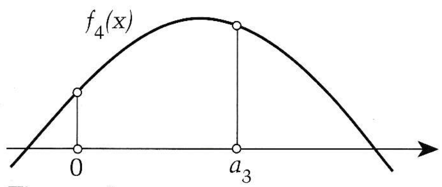


Figure 4

(c) Proceeding in the same way, consider the function

$$f_n(x) = a_1{}^2 - a_2{}^2 + \ldots + (-1)^n a_{n-1}^2 - (-1)^n x^2 - (a_1 - a_2 + \ldots + (-1)^n a_{n-1} - (-1)^n x)^2.$$

This function is linear for odd n and has the form $-2x^2 + p_n x + q_n$ for even n. So to prove that $f_n(x)$ is nonnegative on the interval $0 \leq x \leq a_{n-1}$, it suffices to verify conditions $f_n(0) \geq 0$ and $f_n(a_{n-1}) \geq 0$ at its endpoints. But these two inequalities coincide with the statement of the problem for the first $n - 1$ and $n - 2$ numbers, respectively, of the sequence $a_1 \geq a_2 \geq \ldots \geq a_{n-1} \geq a_n$.

From here, the general statement emerges by the principle of mathematical induction, in the following version: if a statement depending on n is valid for $n = n_0$ and $n = n_0 + 1$, and its validity for $n-1$ and $n-2$ implies its validity for n, then it's true for all $n \geq n_0$. In our case, $n_0 = 2$ [inequality (1) above], $n_0 + 1 = 3$ [part (a) of the problem; part (b) is, in fact, the case of $n = 4$], and the "inductive step" was proved above.

Figures 5 and 6 illustrate, for the case $n = 7$, an elegant geometric solution.

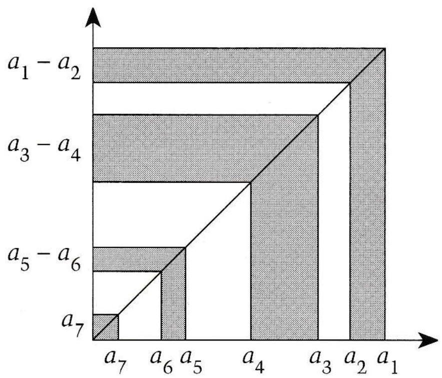


Figure 5

The left part $(a_1^2 - a_2^2) + (a_3^2 - a_4^2) + \ldots$ of our inequality is the total area of the shaded trapezoids in figure 5 (two adjacent trapezoids fill the gap between the square $a_k \times a_k$ and the square $a_{k-1} \times a_{k-1}$). The right part $(a_1 - a_2 + a_3 - a_4 + \ldots)^2$ is the area of a square whose side length equals the sum of the heights of every other trapezoid along the bottom of the original square. This square can be cut into trapezoids with the same heights (fig. 6) but with shorter (or not longer) bases.

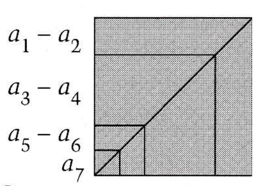


Figure 6

So the right part does not exceed the left part.

A slight adjustment to this argument is necessary for cases where n is an even integer. (N. Vasilyev)

M59

Suppose ABC is the required triangle and point E lies inside angle BAC (fig. 7). Since both points I and E are equidistant from lines BA and BC, lines BI and BE are the bisectors of the interior and exterior angles of

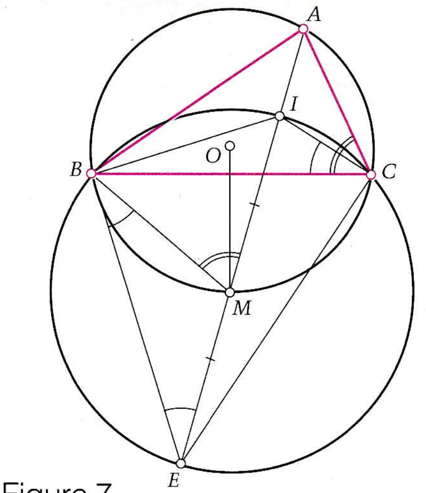


Figure 7

the triangle at vertex B, so $\angle IBE = 90°$. Similarly, $\angle ICE = 90°$. It follows that the quadrilateral $BICE$ is inscribed in the circle with diameter IE, and the midpoint M of IE is the center of the circle. Angle IMB is an exterior angle of isosceles triangle MBE, so

$$\begin{aligned}\angle IMB &= \angle MBE + \angle MEB \\ &= 2\angle MBE = 2\angle IEB = 2\angle ICB\end{aligned}$$

(the last equality follows from the Inscribed Angle Theorem: angles IEB and ICB intercept the same arc IB). Since CI bisects angle ACB, $2\angle ICB = \angle ACB$, and so $\angle AMB = \angle IMB = \angle ACB$, which means that points A, C, M, and B lie on the same circle—the circumcircle of triangle ABC.

Now, given points O, I, and E, we find the midpoint M of segment IE. Then two of the vertices of the unknown triangle are the points of intersection of two circles: with center O and radius OM, and with center M and radius $MI = ME$; the third vertex is the second intersection point of the first circle and line EI. This argument is easily reversed to show that this construction yields the required triangle if point I lies inside the circle (O, OM). The reader is invited to investigate the situation in which point I is on or outside the circle with center O and radius OM. The original problem has no solution in this situation. (V. Dubrovsky)

M60

(a) Yes, it's always possible. We'll even prove a more general statement: *if all the cells of the $n \times n$ square, except for some cells of one row (or column), are regularly colored, then the coloring of the entire square can be completed in the regular way.*

Let the uncolored cells lie, for instance, in the first row. Paint each of them with the color that doesn't yet occur in its column. We must prove that there are no two cells of the same color in the first row—say, no two red cells. To this effect, consider the $(n-1) \times n$ rectangle obtained from the square by erasing the first row. Since the initial coloring was regular, each of $n-1$ rows of the rectangle has exactly one red cell, so the total number of red cells in the rectangle is $n-1$. At the same time, each of n columns in the rectangle has no more than one red cell, and so *only one* of the columns doesn't contain a red cell. It follows that only one cell of the first row will be painted red.

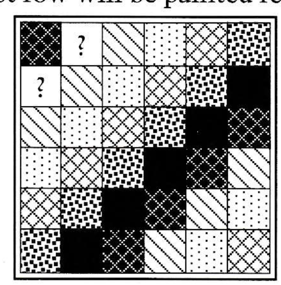


Figure 8

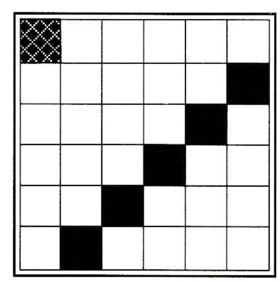


Figure 9

(b), (c) In these two cases the answer, in general, is no, as the examples in figures 8 and 9 show (the figures show $n = 6$, but it's clear, of course, what should be done for an arbitrary n). Evidently the answer is negative for any number of initially colored cells between n and $n^2 - 2$ (as a counterexample, one can take a coloring "intermediate" between those shown in the figures). On the other hand, it seems quite evident that a regular coloring of fewer than n cells can be regularly completed, though it's not so easy to prove it accurately. It's an even more interesting problem to find some simple necessary and sufficient conditions for a coloring to be completed. (N. Vasilyev)

Physics

P56

The ant's speed doesn't change linearly. Therefore, its average speed is different on the different sections of its path, and so in our solution we can't use the usual formulas for average velocity.

Let's break the ant's path from point A to point B into small parts, each of which the ant covers in the same interval of time Δt. Then $\Delta t = \Delta l / v_{av}(\Delta l)$, where $v_{av}(\Delta l)$ is the average speed over the given distance Δl. This formula suggests an approach to the solution: let's draw the curve of the dependence of $1/v_{av}(\Delta l)$ on l for the path from point A to point B. This graph is a straight line segment (fig. 10). The shaded area in figure 10 is numerically equal to the required time. It's not difficult to find it:

$$s = \frac{1/v_1 + 1/v2}{2}(l_2 - l_1)$$
$$= \left(\frac{1}{2v_1} + \frac{1}{2v_1}\frac{l_2}{l_1}\right)(l_2 - l_1)$$
$$= \frac{l_2^2 - l_1^2}{2v_1 l_1} \quad \left(\text{as } \frac{1}{v^2} = \frac{1}{v_1}\frac{l_2}{l_1}\right).$$

This means that the ant will run from point A to point B in the time

$$T = \frac{4-1}{2\cdot 2\cdot 10^{-2}}\,\text{s} = 75\ \text{s}.$$

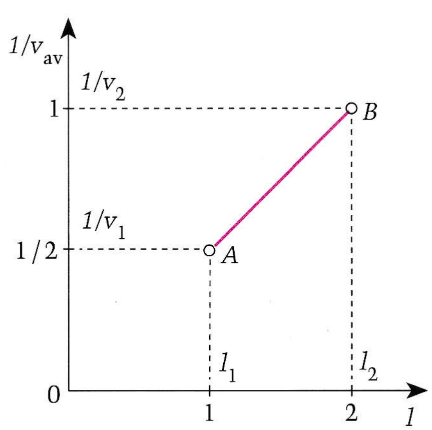


Figure 10

P57

Since the weight's mass is much less than the wheel's mass, we can ignore the weight's gravitational force on the wheel's motion and assume that the wheel rolls with a constant velocity. In this case the acceleration of the weight in the coordinate system centered on the Earth is equal to the centripetal acceleration of the weight in the coordinate system centered on the rolling wheel—that is,

$$a = \frac{v^2}{R}$$

(v is the wheel's velocity, R is its radius). The acceleration is supplied by the force of gravity mg and the reaction of the wheel N (fig. 11a).

The weight acts on the wheel with a force Q (fig. 11b) numerically equal to the force N but oriented in the opposite direction. It's clear that the wheel jumps up if the vertical component of force Q is larger than the gravitational force Mg on the wheel. The vertical component of force Q is maximum and equal to Q at the moment when the weight is at the top of the wheel. So the velocity at which the wheel will jump up is determined from the equation

$$Q \geq Mg. \tag{1}$$

Applying Newton's second law to the weight at the highest point, we can write (see figure 11c)

$$mg + N = m\frac{v^2}{R},$$

which yields

$$N = m\left(\frac{v^2}{R} - g\right).$$

Since $Q = N$, equation (1) can be expressed in the form

$$m\left(\frac{v^2}{R} - g\right) \geq Mg,$$

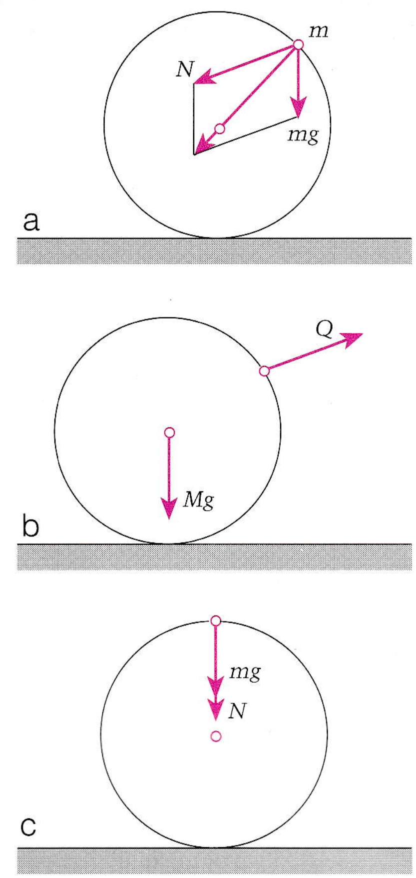


Figure 11

from which we get

$$v \geq \sqrt{Rg\left(\frac{M}{m} + 1\right)} = 4\sqrt{Rg}.$$

P58

Let's refer to the characteristic points on the curve as illustrated in figure 12. At $V < V_1$ the pressure of the gas mixture doesn't change, which means that both oxygen and nitrogen condense, and the pressure is equal to the sum of the saturated vapor pressure of the oxygen $p_{s,\,O_2}$ and nitrogen $p_{s,\,N_2}$ at $T = 74.4$ K. Since the given temperature is the boiling point of liquid nitrogen, $p_{s,\,N_2} = p_0$, which is the atmospheric pressure (100 kPa). The breaks in the graph at the points (V_1, p_1) and (V_2, p_2) are evidence of the phase transitions and gas condensations: if for $V < V_1$ both gases condensed, then for $V_1 < V < V_2$ one gas condensed, and for $V > V_2$ condensation does not occur. Suppose that at the point (p_1, V_1) nitrogen liquefies; then oxygen liquified at the point

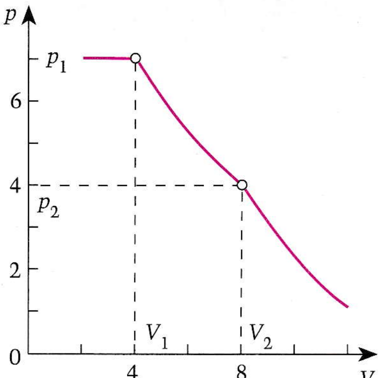


Figure 12

(p_2, V_2), and we can write down the equations

$$p_1 = p_{s,O_2} + p_0,$$
$$p_2 = p_{s,O_2} + p_{N_2}, \qquad (1)$$

where p_{N_2} is the partial pressure of the nitrogen at the point (V_2, p_2). Insofar as nitrogen is only in the gaseous state on the section V_1V_2, then according to Boyle's law we can write

$$p_{N_2} = p_0\frac{V_1}{V_2}.$$

Substituting this value in equations (1) and dividing those equations by one another (using the expressions $p_1/p_2 = 7/4$ and $V_1/V_2 = 1/2$), we find

$$p_{s,O_2} = \frac{1}{6}p_0 \cong 17 \text{ kPa}.$$

If we had assumed that at the point (V_1, p_1) oxygen condenses, we would get $p_{s,O_2} = 6p_0$, which is easy to check. This contradicts the fact that oxygen boils at a higher temperature—the saturated vapor pressure of oxygen at $T = 74.4$ K must be less than p_0.

Let's find the mass of oxygen m_{O_2}. The point (p_2, V_2) corresponds to the onset of oxygen condensation—that is, its pressure is equal to p_{s,O_2}, and in this case all oxygen is in the gaseous state. According to the Clausius–Clapeyron–Mendeleyev law,

$$p_{s,O_2}V_2 = \frac{m_{O_2}}{\mu_{O_2}}RT, \qquad (2)$$

where μ_{O_2} is the molar mass of oxygen. For nitrogen the condensation begins at the point (p_1, V_1), which means that

$$p_0V_1 = \frac{m_{N_2}}{\mu_{N_2}}RT, \qquad (3)$$

where μ_{N_2} is the molar mass of nitrogen. Dividing equation (2) by equation (3) and taking into consideration that $\mu_{N_2}/\mu_{O_2} = 7/8$ and $p_{s,O_2} = p_0/6$, we determine the mass of the oxygen to be

$$m_{O_2} = \frac{8}{21}m_{N_2} \cong 38 \text{ g}.$$

P59

The electrical current in the electrolyte is produced by mass transfer. In this case the metallic mercury from the electrolytic solution is reduced on the cathode, and oxidation occurs on the anode. In accordance with Faraday's law of electrolysis, the mass m of mercury produced on the cathode in time t is equal to

$$m = \frac{1}{F}\frac{M}{n}It, \qquad (1)$$

where $F = 9.65 \cdot 10^4$ C/mol is the Faraday number, $M = 0.201$ kg/mol is the molar mass of mercury, $n = 2$ is the valency of mercury, and I is the current. Since the resistance of the metallic mercury and electrolyte and the internal resistance of the source are negligibly small compared to R, the current I is equal to

$$I = \frac{\mathcal{E}}{R}. \qquad (2)$$

The production of mercury on the cathode and its dissolution on the anode causes the drop of electrolyte to move to the anode side. The distance l the drop shifts is related to the mass of the mercury produced on the cathode by the relation

$$m = \rho\frac{\pi d^2}{4}l, \qquad (3)$$

where $\rho = 13.6 \cdot 10^3$ kg/m^3 is the density of mercury.

From equations (1) through (3) we find the time it takes the drop of electrolyte to move the distance l:

$$t = \frac{\pi d^2 Fn\rho Rl}{4M\mathcal{E}}.$$

Substituting the values given in the original statement of the problem, and taking $l = 1$ cm, we get

$$t \cong 100 \text{ hr}.$$

Such "mercury watches" are used in electronics as miniature timers. They're also used to measure the charge in a circuit over an extended period of time.

P60

Let's consider a certain glowing point A of the filament and an arbitrary ray AB emerging from it. We draw a plane through the ray and the filament. It follows from geometrical considerations that with all possible reflections, the given ray will remain in the constructed plane (fig. 13). After the first reflection at the conical surface, the ray AB will propagate as if it emerged from point A'—that is, the virtual image of point A. The necessary condition for preventing any of the rays emerging from A from landing on the mirror is that point A' must not be higher than the straight line OC—that is, the second generator of the cone, lying in the plane of the ray (point O is the vertex of the conical surface). This requires

$$\angle A'OD + \angle AOD + \angle AOC$$
$$= 3\frac{\alpha}{2} \geq 180°.$$

Consequently, $\alpha_{min} \geq 120°$.

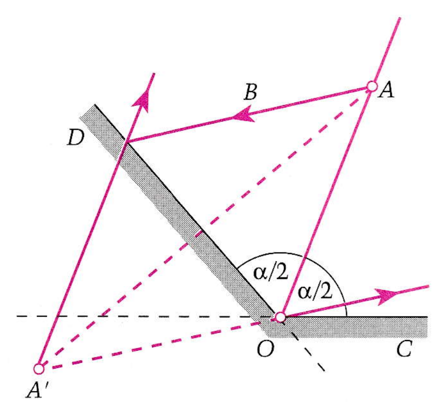


Figure 13

Brainteasers

B56

In 1980. The nephew was born in the year $x^2 - x$. So $x^2 - x < 1992 \leq x^2$. A quick trial-and-error shows that $x = 45$.

B57

The temperature t_F in Fahrenheit is a linear function of the temperature t_C in Celsius: $t_F = at_C + b$. Coefficients a and b can be found by equating the temperatures of the freezing point and boiling point: $t_F = (9/5)t_C + 32$. Equation $t_F = t_C$ yields –40°F = –40°C.

B58

At first the small drops of dirty water hitting the windshield don't spread over it because the dry glass isn't wet enough. The windshield wipers moisten all the glass they touch, and the drops of dirty water hitting the wet glass disintegrate due to capillary action. Because of the drastic reduction in visibility, experienced drivers hold off turning on the wipers.

B59

See figure 14.

B60

When Koshchei died, he left the treasure on the surface.

The first hole was dug in one day, the second in $2^2 = 4$ days, the third in $3^2 = 9$ days, and so on. So in $1 + 2^2 + 3^2 + \ldots + 13^2 = 819$ days, Koshchei buried his hoard at a depth of 13 m. Then he dug it up ($819 + 13^2 = 988 < 1{,}001$), but died before he could bury it 14 m down, because $819 + 14^2 = 1{,}015 > 1{,}001$.

Friction

1. The fall factor can be found from the length of the rope $l = 10$ m and the height of the fall $h = 12$ m. The fall factor $\phi = h/l = 1.2$. This is a pretty serious fall. Generally climbers don't like taking falls with fall factors greater than 1. Apart from the dangers involved, it means replacing expensive equipment more frequently because ropes that have been stressed by a serious fall must be viewed with suspicion.

2. Substituting numerical values into the equation for the generated force derived in the article gives F = 6,543 N, a figure well within the safety limits.

3. The deceleration ratio is

$$\frac{a}{g} = \frac{F - mg}{mg} = 7.3.$$

Bungee jumpers (who jump head-first off bridges and towers with a tether attached to their ankles) use very elastic cords and don't approach these kinds of accelerations.

4. The extension is $\Delta l = Fl/M$ = 1.64 m, or about 16%.

5. A $\phi = 0$ fall means the height of the fall is zero. A climber secured by a tight rope and level with her protection slips, and the catching shock is just her own weight mg.

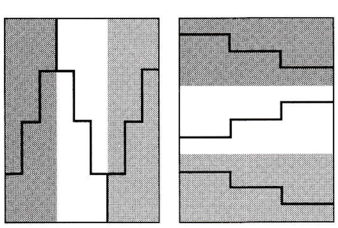


Figure 14

Kaleidoscope

Problems

1. Turn the number upside down.
2. The decimal point.
3. 16 seconds. The interval between two strikes is 2 seconds. The number of intervals is one less than the number of strikes.
4. Yes, there are—for instance, (–1, –2, –3), or (a, 0, $-a$) for any integer a.
5. Of course it is: the coin that isn't a dime is a nickel, but the other one is in fact a dime!
6. One must shoot 6 times and score 17 four times and 16 twice.
7. 30 grams. We use the fact that the ratio of the masses (that is, volumes) of similar figures is the cube of the ratio of the sides of the figures.
8. Yes. If, say, in the box labeled "nails" you find screws, then the "nuts" box must have nails in it, and the "screws" box must contain nuts.
9. 72. If the number is N, then the number of factors of 2 in N must be an odd multiple of 3, and the number of factors of 3 must be both even and one less than a multiple of 3.
10. Pour water from the second glass into the fifth.
11. 1/2 and –1.
12. Roll up the napkin, gently pushing the glass with the part that's rolled up.

Winning strategies for games

1. If initially both piles are "odd" (that is, consist of odd numbers of rocks), the first player loses: no matter what his move is, he'll leave two piles, at least one of which is "even"; then the second player can divide it into two odd piles and take away another pile, restoring the initial situation. A little reflection will show that this implies that the second player will always be able to make a move.

If initially at least one of the piles is even, then the first player divides it into two odd piles, creating a losing position for the second player (by the above argument).

2. If there's a peg tied to two other pegs, the player who plays next wins by binding together these pegs. So the game will continue as long as there are at least two "free" pegs (that is, pegs that are tied to only one other peg). It will take $n/2$ moves for an even number n of pegs, and $(n - 1)/2$ moves for odd n. So the first player wins if n is divisible by 4 (that is, $n/2$ is even) or if n has a remainder of 1 when divided by 4 (that is, $(n - 1)/2$ is even). If the remainder is 2 or 3, the second player wins.

3. The second player always wins. After the first player makes his move, the second one can tear off one or two petals in such a way that the remaining petals are divided into two equal parts. Then the second player can always repeat each move of the first player on the other half of the flower.

Ice cream

1. The father can include exactly one, exactly two, or exactly three children in his distribution. If he chooses one child to give all three pennies to, he can make his distribution in six ways. If he chooses two children, he must give one penny to a first and two pennies to a second. That is, he is choosing an *ordered pair* of children. There are six ways of making the first choice, and, by the multiplication principle, five ways of making the second; so he has 30 ways to choose altogether. If he gives a single penny to each of three children, he is choosing a subset of three from a set of six children. This allows for ${}_3C_6 = 20$ possible choices. Altogether, he has 6 + 30 + 20 = 56 possible choices.

2. This problem reduces to the third case above, so there are 20 possible ways to make the distribution.

3. We can distinguish the same cases as in problem 1. There are six ways in which the father can give all his coins to one child. If he chooses exactly two children to favor, one child gets a single coin, and the other gets two coins. So we can again consider the children as an ordered pair, and there are 30 such ordered pairs possible. We then choose one of the three coins to give to the first of the ordered pair, leaving the other coins for the second child. There are three ways of doing this for each ordered pair, making 60 possibilities altogether. Finally, he can choose three children in 20 ways, then match one of the three coins with each of the three chosen children. There are six ways to do this, so there are 120 ways to include three children in the distribution. This makes 6 + 60 + 120 = 186 ways to distribute the coins.

4. We can diagram the coins as follows:

0 0 0 0 0 0

We want to separate the coins into three sets, some of which may be empty. We can then match each set with one of the children. Since there will be six ways to make this match, the total number of possibilities will be six times the number of "partitions" of the coins.

To count these partitions, we can use the "order form" device by putting a "stick" between subsets. There are seven places to put a stick, and each stick can go anywhere (two adjacent sticks will include the null set between them, so one child gets no pennies). So we must choose two positions out of seven for the sticks. For choices in which these positions are distinct, the number is ${}_7C_2 = 21$. But there are seven more positions where the sticks are adjacent, making 21 + 7 = 28 possible partitions of the coins. So there are 6 × 28 = 168 ways to distribute the coins.

5. We can use the same device for this problem if we agree that the sticks cannot be adjacent and that neither can be at the beginning or the end of the row of coins. We must choose two out of five positions for the sticks, to give a total of ten possible ways to distribute the coins.

6. For each coin, there are three choices of a child to give it to. This will make $3^6 = 729$ possibilities for distribution of the coins, including cases where one or two children get no coin at all.

Let's count these cases, which are prohibited in problem 5. There are three cases in which one greedy child gets all the coins. If one child is excluded, then we can repeat the previous argument, except that now there are only two choices for a recipient for each coin. This makes $2^6 = 64$ possibilities if any one child is left out. But this count includes the two cases in which another child is also left out, so there are really only 62 possibilities in this case. Now there are three ways to leave out a child, so there are in all 3 × 62 = 186 ways to distribute the coins to exactly two children, and 189 "forbidden" cases (in which some children are left out). So there are 729 – 189 = 540 ways to distribute the coins, if each child must get at least one coin.

Meandering

Let's find the shape of the rotating liquid in the glass with angular velocity ω (see figure 1 in the article). If the linear dimension of the chosen region is Δr, then its mass is equal to

$$\Delta m = \rho \cdot \Delta s \cdot \Delta r.$$

Using this expression, from equations (1) and (2) (see page 34) we get

$$\Delta h = \frac{\omega^2}{g} r \cdot \Delta r.$$

So the local slope of the surface profile increases proportionally with the distance r from the axis of rotation:

$$\tan\alpha = \frac{\Delta h}{\Delta r} = \frac{\omega^2}{g} r.$$

If we take the height of the liquid level at the axis to be h_0, the height at a distance r from the axis can be obtained by adding the product of r and the average value of tan α. To calculate the average value of a linear function, it's sufficient to calculate the value at the midpoint of its range. And so we get

$$h(r) = h_0 + \frac{\omega^2 r}{2g} r.$$

It's clear that the dependence of $h(r)$ is quadratic on r and is graphically represented by a parabola. (Those who are familiar with integral calculus can integrate the expression for $h(r)$ from $r = 0$ to $r = r$ to get the same result.)

It's curious that the parabolic shape of a rotating liquid was put to practical use in creating a special kind of telescope. The renowned American physicist Robert Wood constructed a telescope with a parabolic mirror by putting a rotating container of mercury at the bottom of a well.

Smiles

The second part of the inductive argument presented implies that $k > 2$, because it refers to the validity of the induction hypothesis for $N = 2$. So there's a gap between the initial value $N = 1$ and the values ($N > 2$) involved in the second part of the proof. This is one more example of a bad inductive proof, similar to those considered in the article "What Did the Conductor Say?" on page 38.

Dimensional thinking

(See the May/June issue of *Quantum*)

1. The minimum period of rotation corresponds to the maximum angular frequency: $\omega_0 = 2\pi/T_0$. The relation between ω_0, M, R, and the gravitational constant G is simply found from considerations of dimensionality and takes the form $\omega_0^2 \propto GM/R^3$ (the dimensions of G are easily found from the law of gravity.) The numerical estimates for a planet with the same parameters as Earth are obtained if we consider that $G \cong 6.7 \cdot 10^{-11}$ N $\cdot$ m^2/kg^2, $M \cong 6 \cdot 10^{24}$ kg, $R \cong 6.4 \cdot 10^3$ km. We suggest that you calculate how many times slower the Earth would rotate as a result and consider the question of what would happen if we could manage to rotate the Earth with $\omega > \omega_0$.

2. In addition to the geometrical similarity of the tuning forks, we have to take into account that they are made of the same material and that they have the same elasticity characteristics and density (Young's modulus E is the natural elastic property). If L is the length of the tuning fork's leg, then from dimensional considerations the formula for the frequency gives

$$\omega \propto \frac{1}{L}\left(\frac{E}{\rho}\right)^{1/2}.$$

So the ratio of the frequencies of geometrically similar tuning forks is inverse to the ratio of the corresponding lengths and is equal to 1 : 3. We might also point out that the ratio E/ρ is the square of the speed of sound in that solid. Then the previous formula becomes quite clear. The oscillation period of the tuning fork $T \propto \omega^{-1}$—that is, it's simply determined by the time it takes for the sound wave to travel the length of the tuning fork. Now think of why the geometrical similarity of the tuning forks was emphasized in stating the problem. Might something change if the cross sections of the legs of the tuning forks aren't similar?

3. We'll assume that the period T is determined by the pressure p, the energy E released by the explosion, and the water density ρ. Here the density plays the role of a mass characteristic. Writing $T \propto p^x\rho^y E^z$ and comparing the dimensions of the left and right sides of this formula, we get $T \propto p^{-5/6}\rho^{1/2}E^{1/3}$. If we take into account that the water pressure p is related to the depth H by the relation $p = \rho gH$, we arrive at the formula

$$T \propto (gH)^{-5/6}\left(\frac{E}{\rho}\right)^{1/3}.$$

The gas bubble created by the explosion will expand until the internal pressure becomes equal to the pressure of the surrounding water. Using dimensional considerations, we again get a rough but reasonable estimate for the internal pressure by dividing the energy E by the bubble's volume: $p_{\text{int}} \propto E/r^3$. Since the order of magnitude of p_{int} is equal to the external pressure $p = \rho gH$: $(E/r^3) \propto \rho gH$. From this we get $r^3 \propto E/\rho gH$. So the size of the bubble is proportional to $H^{-1/3}$. In our estimates we neglected the change in water density ρ with depth.

4. The pressure at the center of a star of mass M and radius R is expressed by the formula $p \propto GM^2/R^4$ obtained directly by the dimensional method. We'll let you do the numerical estimates for this problem and problem 5 yourself.

"VAVILOV'S PARADOX" CONTINUED FROM PAGE 50

As is evident from formula (3),

$$\alpha = \frac{1}{2}\frac{\lambda}{a}. \qquad (4)$$

Suppose $a = 1$ cm, $\lambda = 5 \cdot 10^{-5}$ cm; then $\alpha = 2.5 \cdot 10^{-5}$ rad.

So problems with the law of conservation of energy arise when the angle between beams is of the order of the ratio of the wavelength to the beam width.

The solution to the paradox is that you can't use the notion of an ideal parallel beam of finite cross section with such small angles. We'll fail in any attempt to actually produce such beams. Because of diffraction, the limitation on the size of the beam necessarily causes it to diverge. The diffraction angle ϕ is determined by formula (4). Therefore, the angle α at which the beams intersect can be determined with a precision only of the order of angle ϕ. As long as $\alpha >> \phi$, the divergence of the intersecting beams can be neglected. But when α becomes comparable to ϕ, the idea of a narrow parallel beam becomes meaningless.

If diffraction hadn't been discovered, we could, on the basis of the law of conservation of energy, not only surmise that it exists but obtain the fundamental rule governing the size of the diffraction angle (the angle of divergence). This is a good example of how the law of conservation of energy can serve as a reliable guiding star in physics.

If we take actual light beams, we'll certainly never contradict the law of conservation of energy. In experiments of this type, we'll always end up with a spatial redistribution of the energy flux.

Index of Advertisers

Out of Flexland

The third and last of Mr. Flexman's twisty tours

by Vladimir Dubrovsky

BEFORE WE END OUR magical Flexland tour, Mr. Flexman wants to take us into the deepest recesses of this realm of surprises. This area is still buzzing from the sensation caused when it was discovered some 15 years ago. The region is inhabited by *flexors*, or flexible polyhedrons, whose existence was much doubted, even denied, by generations of mathematicians, starting with the great authority Leonhard Euler.

So we begin our last trip with an investigation of a "pseudo-flexor" (so to speak).

A ring of tetrahedrons

This amusing toy, described in *Mathematical Recreations and Essays* by W. W. Rouse Ball and H. S. M. Coxeter, is seen in figure 1. Its most remarkable feature is its flexibility—it can be twisted as the arrows in the figure show, intriguingly changing shape. The ring of tetrahedrons is easily made out of a paper strip with two rows of triangles drawn on it (see figure 2): just bend the paper up along the dotted lines and down along the solid lines, fold it along the creases, and paste the flaps to the corresponding edges according to the color scheme. This model helps us understand what flexibility in a polyhedron means. The term means that the polyhedron can be bent so that its faces stay rigid—retain their shape and size—while the angles between them change. So the ring of tetrahedrons is a flexible polyhedron. Yet it's not a genuine flexor, because it's not even a polyhedron in the strict mathematical sense. Some edges of our ring are sides of four faces at a time, whereas the definition of a polyhedron (too complicated to be cited here in full) requires that every edge be a side to two faces, if it's an interior edge, or to one face, if it's a border edge. It's clear, though, that a polyhedron with a border can be flexible—take, for instance, two squares with a common side. So the problem of rigidity or flexibility has to do with closed polyhedrons—those having no border. The history of this problem began with a conjecture.

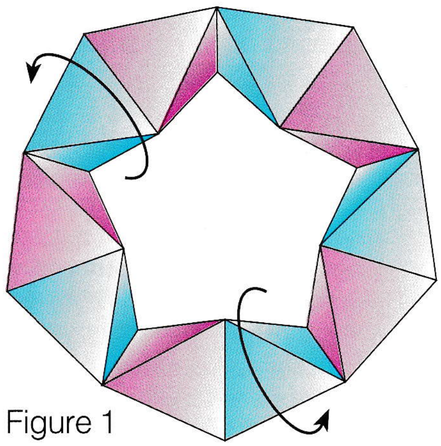


Figure 1

Euler's rigidity conjecture

This conjecture, proposed by Euler in 1766, reads: "A closed spatial figure does not admit any changes until it tears." By a "closed spatial figure" Euler meant what is now called a closed surface. But the assumption seemed especially sound for closed polyhedral surfaces—in particular, for *simple* polyhedrons. A polyhedron is called simple if it is "topologically equivalent" to a sphere—that is, if we made it out of thin rubber and inflate it, it would swell into a sphere. If we designate vertices as V, edges as E, and faces as F, Euler's formula

$$V - E + F = 2$$

is always true for a simple polyhe-

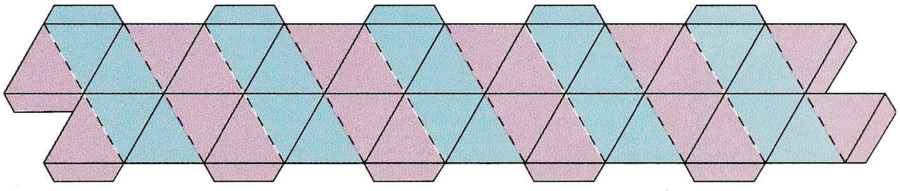


Figure 2

Articles by V. Zalgaller, N. Dolbilin, and A. Panov, published in *Kvant* magazine, and research articles on the problem of the rigidity of polyhedrons were used in preparing this installment of the Toy Store.

dron.[1] Conversely, this equation implies the simplicity of a polyhedron.

In 1813, Euler's conjecture received a powerful confirmation when Augustin Louis Cauchy proved that any *convex* polyhedron is rigid. Actually, Cauchy proved even more than that: he showed, in a very ingenious way, that any two convex polyhedrons whose respective faces are congruent and are joined to each other in the same order are congruent. So you can't bend a convex polyhedron, because such a bending would yield new convex polyhedrons with the same shapes, sizes, and order of the faces. Convexity is indispensable to Cauchy's theorem, as the following example shows.

Imagine an open cube box covered with a lid in the shape of a low quadrilateral pyramid. Now take off the lid, turn it over, and cover the box again. We get a pair of different polyhedrons that satisfy all the conditions of Cauchy's theorem but are not congruent. This is possible only because one of them is not convex. However, both of these "covered boxes" are rigid, and neither can be transformed into the other. A more interesting example, ostensibly refuting the rigidity conjecture, was given in 1962 by the prominent Soviet geometers A. D. Alexandrov and S. M. Vladimirova.

A gasping sea star

This polyhedron has 20 congruent triangular faces combined to form an "angular sea star" (fig. 3). A paper model of this creature can exist in two basic states (shown in the figure) that can be "snapped" into one another by squeezing the thicker one in the vertical direction and the thinner one horizontally. If the paper isn't too stiff, the transformation seems very smooth, creating the impression of flexibility. But this model is misleading.

Suppose we fix the side lengths a, b, c of a face of our sea star (fig. 3).

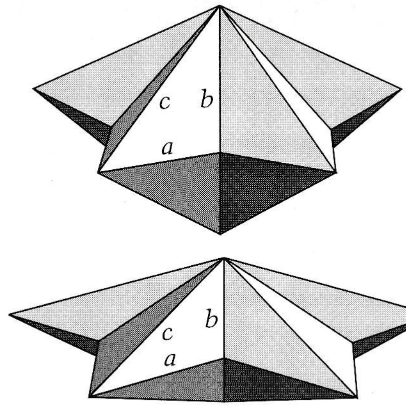


Figure 3

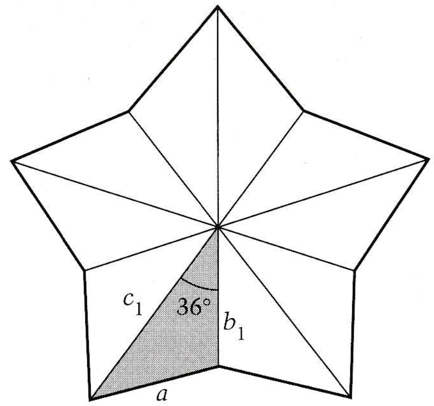


Figure 4

What can we say now about its shape as a whole? When we look at it from the top, we see a star divided into 10 congruent triangles (fig. 4). Consider one of them (shaded in the figure). One of its sides, a, is the side of the face whose projection forms this triangle. We also know the opposite angle, which obviously measures $360°/10 = 36°$. As for the other two sides, b_1 and c_1 (the projections of the sides b and c of the face), it's easy to figure out by the Pythagorean Theorem that the difference of their squares is the same as for sides b and c: $c_1^2 - b_1^2 = c^2 - b^2$.

Let's try to construct a triangle from these data. Draw a line segment BC of length a. Then the opposite vertex A of the triangle lies on the intersection of two loci: the locus of points A such that angle BAC measures 36°, and the locus of points A such that $AB^2 - AC^2 = c^2 - b^2$. The first locus is an arc (more exactly, two arcs symmetric with respect to line BC); the second locus is a line perpendicular to BC; and their intersection depends on the given parameters a, b, and c.

Without going into more detail (which would be a good exercise in geometry for the reader), I'll present the result: if

$$1 < \frac{c^2 - b^2}{a^2} < \frac{1}{\sin 36°} \cong 1.70,$$

then our construction yields exactly two noncongruent triangles. So in this case exactly two starlike polyhedrons can be assembled from the same set of 20 triangles—just as with the cube and its pyramidal lid. This means that theoretically both states of the sea star are perfectly rigid. The model can "breathe," transforming from one state to the other, merely because of the elasticity of the material and the relatively small range of changes it undergoes during the transformation. (Mr. Flexman says the sea star is always hanging around the door of the Flexor Club but is never admitted.)

Here are good parameters for making a model of the sea star: $a = 5.4$, $b = 5.9$, $c = 9.1$.

Flexors

Crucial new advances in the study of the rigidity problem were made not too long ago. In 1975, Herman Gluck proved that "almost all" simple polyhedrons are rigid. That didn't mean "absolutely all," but, in a sense, it "almost proved" Euler's conjecture. Yet, unexpectedly, just two years later the conjecture was refuted when Robert Connelly from Cornell University constructed an example of a flexible polyhedron. The first flexors had a

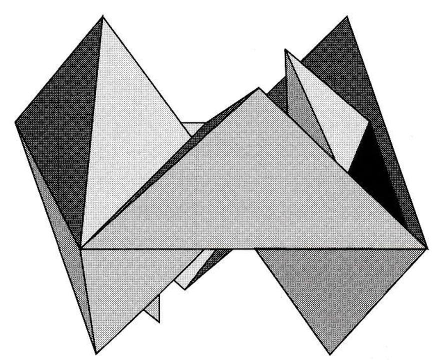


Figure 5

[1]This will be proved in the article "Topology and the Lay of the Land" in the next issue (in the corollary to Theorem 3 there).—*Ed.*

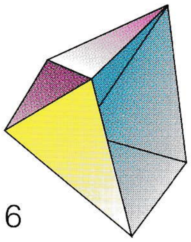


Figure 6

very intricate structure, like the one shown in figure 5. But in 1978 Klaus Steffen created a simpler flexor with only 9 vertices (fig. 6). You can make a model of it using figures 7 and 8. Cut out the three pieces in figure 7—their appropriate dimensions are $a = 12$, $b = 10$, $c = 5$, $d = 11$, $e = 17$ (flaps for gluing the pieces together aren't shown and won't be needed if you use cellophane tape). Tape together the pairs of adjacent edges marked c on the red piece so that a valley appears at the white circle on the colored side and a peak appears at the black circle. Do the same with the blue piece. You'll end up with two pieces that fit together exactly. Nest the pieces so that the colored sides are visible (fig. 8). Tape the edges b of one piece to the corresponding edges b of the other one. Pull apart the free vertices, using the flexibility of the pieces, and attach the yellow piece in between to edges a. That completes the assembly. I must warn you, though, that the "range of flexibility" of this (and other) models is rather limited and can be less than you expect.

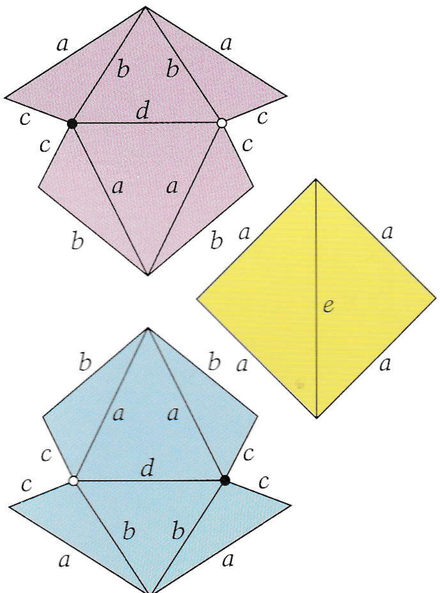


Figure 7

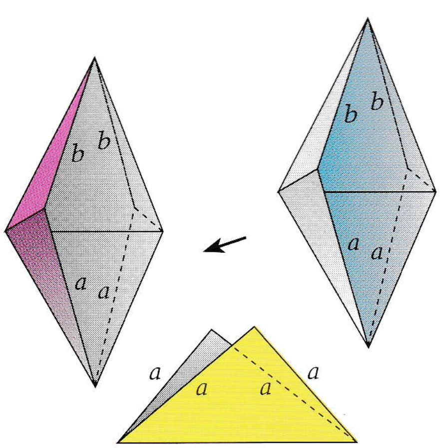


Figure 8

An interesting feature of all known flexors is that their volumes remain the same after any deformation. It has been conjectured that this property is valid for all flexors.

At this point we'll say good-bye to Mr. Flexman and leave Flexland, perhaps to come back some time and make still other discoveries.