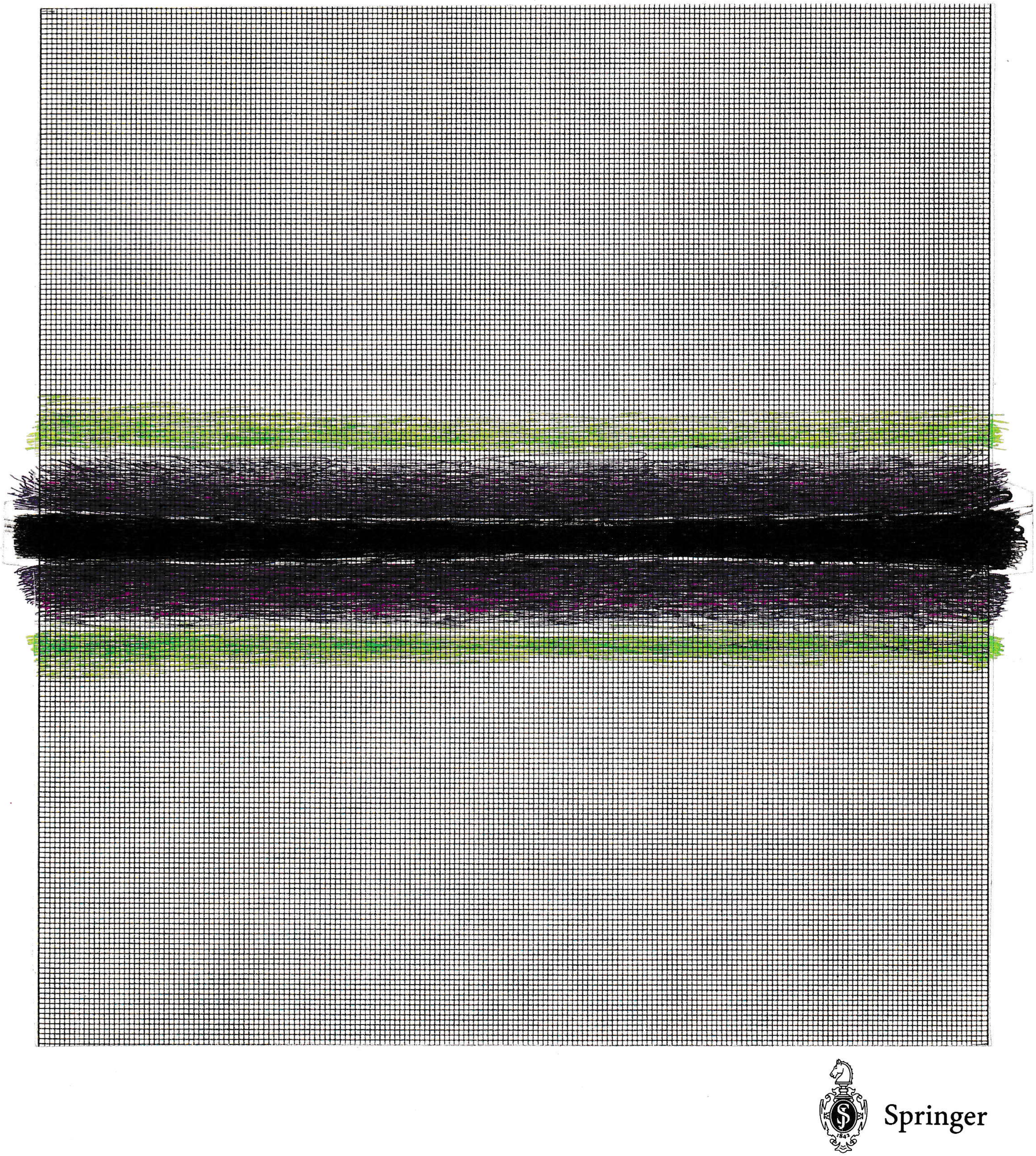
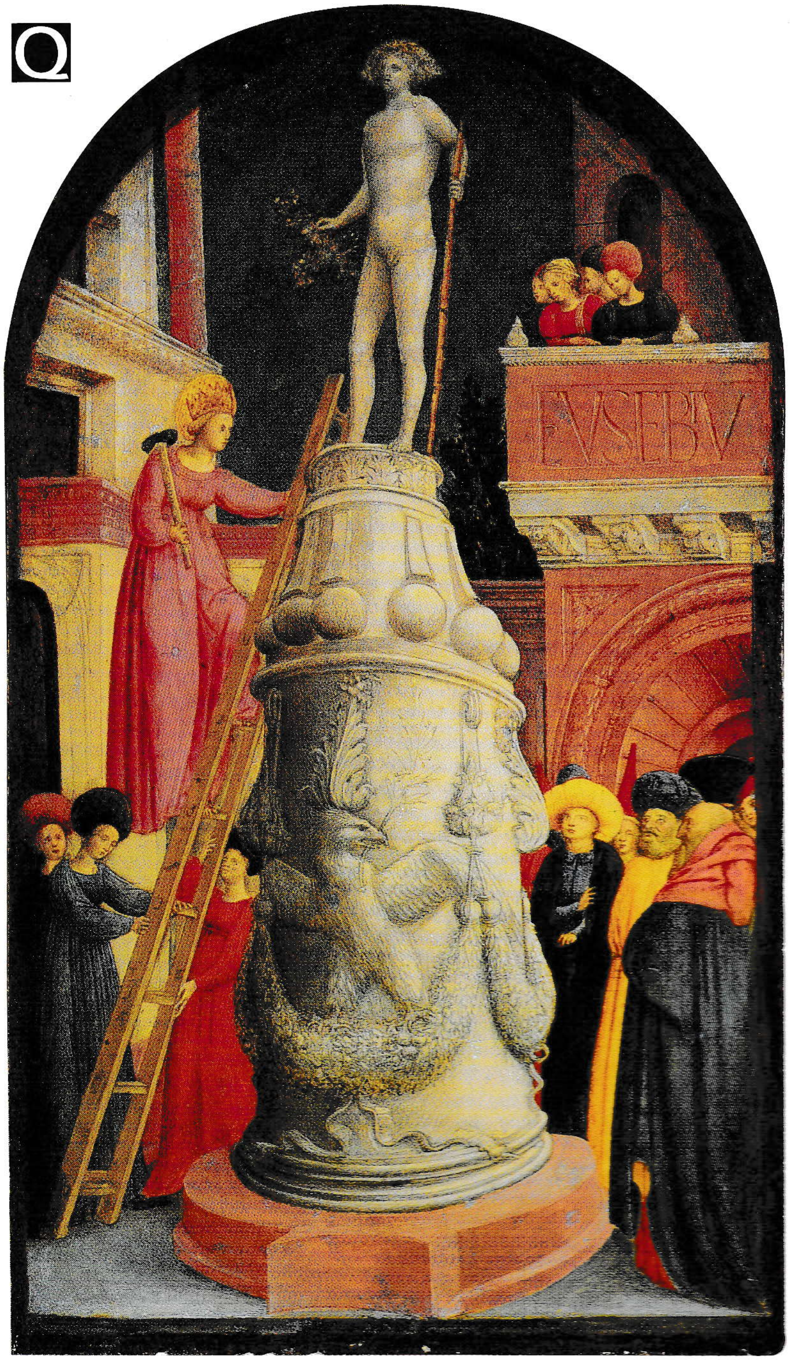
QUANTUM

MAY/JUNE 1995 $5.95



Springer

GALLERY Q



Saint Apollonia Destroys a Pagan Idol (ca. 1450) by Antonio Vivarini

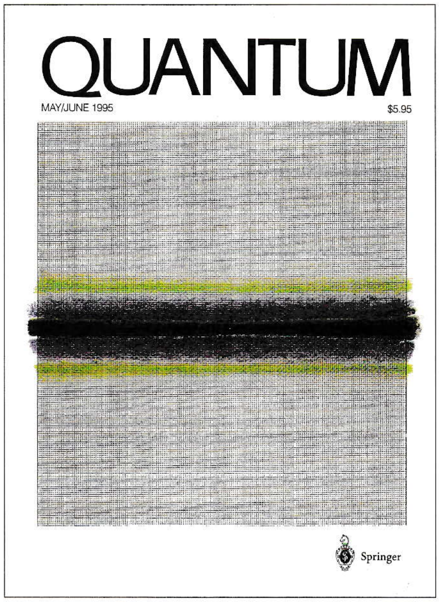
IT'S CURIOUS THAT, PAINTING IN THE MID-15th century, Antonio Vivarini depicts the destruction of a work of art from antiquity. The transition to the Renaissance was well under way by this time, and one would expect to see the greater tolerance of that period reflected here. More than tolerating Greek and Roman culture, the Renaissance actively embraced it. How to explain this untimely assault?

Perhaps Vivarini's motives (or more likely, his patron's) are more subtle. After all, Vivarini is showing us the statue, intact, as if to say: "Back then, when St. Apollonia was alive, we had reason to fear the ideas represented by this object, but no longer. Now that our Faith is fully established, we can appreciate the beauty of this sculpture, as well as the beauty of St. Apollonia's action." Was the irony of the saint's name—taken from one of the greatest gods in the Greek pantheon—lost on Vivarini and his patron? Or the irony of destroying the representation of something that the saint, and Vivarini, and his patron all agree does not even exist?

Nonexistent things clearly wield a fair amount of power. Without ascribing sainthood to our authors, we direct your attention to the article on page 4, where John Wiley investigates (nonexistent?) magnetic monopoles.

QUANTUM

MAY/JUNE 1995 VOLUME 5, NUMBER 5



Cover art by Yury Vashchenko

The so-called "ozone hole" over Antarctica has been the subject of much research and debate since it was first observed in the early 1980s. It engendered fears that levels of atmospheric ozone might be falling elsewhere, posing a threat to life forms on Earth, since ozone absorbs harmful ultraviolet radiation from the Sun.

One question yet to be resolved is whether a new generation of high-flying supersonic aircraft will damage the ozone layer. Aircraft exhaust contains nitrogen oxides (NO_x), which *increase* ozone levels in the troposphere (adding to the photochemical smog that besets big cities), but may *decrease* them in the stratosphere. Albert Stasenko investigates the effect such "NO_xious" chemicals might have on atmospheric ozone in the article that begins on page 20.

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Raising the boats or lowering the water

Will the National Science Education Standards be used to sell our students short?

THE NATIONAL RESEARCH Council of the National Academy of Sciences (NAS) released its draft standards for science education in grades K–12 in late November. No doubt you read about them in the newspapers. Since then I have been working with more than 30 biology, chemistry, physics, and earth and space science teachers to interpret these standards. We have also had access to science educators and university scientists as we work through our interpretations.

You may be wondering: what are these "standards," and where did they come from? And why do they need "interpreting"?

The National Science Teachers Association had asked the NAS to take on the task of preparing those standards so that we could gain wide acceptance for a document that would say what science young people should learn before they graduate from high school. I wanted to use those standards in a project designed to provide a solid six years of science for all students. This multiyear course of study would include physics, chemistry, biology, and the earth and space sciences, each of which would be taught every year in grades 7 through 12.

Now, we have encountered a surprising amount of criticism from many of our science education colleagues, who feel that we have misinterpreted the National Science Education Standards. So I thought it might be useful to get the input from some of our best science teachers and from some of our best science and math students—readers of *Quantum* magazine.

Most science educators, and many scientists, take the position that real science or math—something of significant depth—is beyond the reach of most American young people. They suggest instead that we try to achieve something they call scientific "literacy" in our students. Science-literate students would be able to use science words in the right context and even be able to explain in words some of the laws and theories from science. But the greater emphasis would be placed on the relevance of the science to immediate personal or societal problems. The underlying assumption is that scientific methodology would enable a person to solve such problems even though he or she lacks a quantitative understanding of science.

I have taken the position that most young people *can* learn real science, and learn it at significant levels, if it is sequenced properly over long periods of time. It is essential that abstractions come only after the students have experienced what they're learning about. For example, various kinds of motion should be observed, and only then should the words distance, time, and speed be used descriptively. Next, symbols would be used for these quantities. Finally, numbers with units would be used. Only when concepts like speed, velocity, acceleration, displacement, and time have been well developed should students begin to learn relationships among these concepts, or their relationship to force or work. And, of course, before one talks of force and work, *these* concepts must first be grounded in experience.

So what do *you* think? If you are a student, do you believe that most of your fellow students are capable of learning physics and chemistry, as well as the quantitative and abstract aspects of biology and the earth and space sciences? Or do you think that you have some special, unique, inherent ability? I pose the same question to the teachers among our readers. You know your friends well. You can answer this question.

Now, I'm not asking whether they are inclined to learn such material, or feel disposed to study it as things stand now. It's obvious that many students have voted with their feet, avoiding science and math like the plague. I'm merely suggesting that we educators should be providing experiences that give rise to a sense of awe and excitement—the kind of feeling about science that you have already experienced. That positive feeling, along with striking examples of how science is useful, should provide sufficient motivation for any student.

Am I right? Or am I way off base? Send me your views by letter or by e-mail (bgaldridge@nsta.org).

—Bill G. Aldridge

Be a factor in the QUANTUM equation!

Have you written an article that you think belongs in *Quantum?* Do you have an unusual topic that students would find fun and challenging? Do you know of anyone who would make a great *Quantum* author? Write to us and we'll send you the editorial guidelines for prospective *Quantum* contributors. Scientists and teachers in any country are invited to submit material, but it must be written in colloquial English and at a level appropriate for *Quantum*'s predominantly student readership.

Send your inquiries to

Managing Editor
Quantum
1840 Wilson Boulevard
Arlington VA 22201-3000

Talk back to us

drop us a line at
Quantum,
1840 Wilson Blvd.,
Arlington VA 22201

or

zip us an electron or two at
quantum@nsta.org.

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

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

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

Advertising: MPA

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

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

Subscription Information:

North America: Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $34; Single Issue Price: $5.95. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus NJ 07096-2485; or call 1-800-SPRINGER (777-4643) (in New York, call [201] 348-4033).

All Countries Outside North America: Subscription rates in U.S. currency as above (all rates calculated in DM at the exchange rate current at the time of purchase) plus postage and handling. SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Air mail delivery to all other countries is available upon request. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Magnetic monopoly

A real monograph on something that may not exist

by John Wylie

THE REALLY FUN THING about writing an article for *Quantum* is spending time investigating something trivial in the name of science and education. This article represents some digging I did into the subject of magnetic monopoles. As a physicist, I don't feel that monopoles are particularly trivial. But my wife Holly, a very talented artist, wonders why I would spend so much time researching and writing about invisible little things that may not even exist. This is, of course, precisely why I find magnetic monopoles so fascinating—they don't exist. I can't tell you why they don't exist, but I might be able to tell you a little about how they might behave if they did. This is one of the things that makes being a physicist so much fun—imagining things that could be. In any case, if I can pass on my interest in monopoles to the *Quantum* reader, then maybe my wife will learn to understand me just a little better.

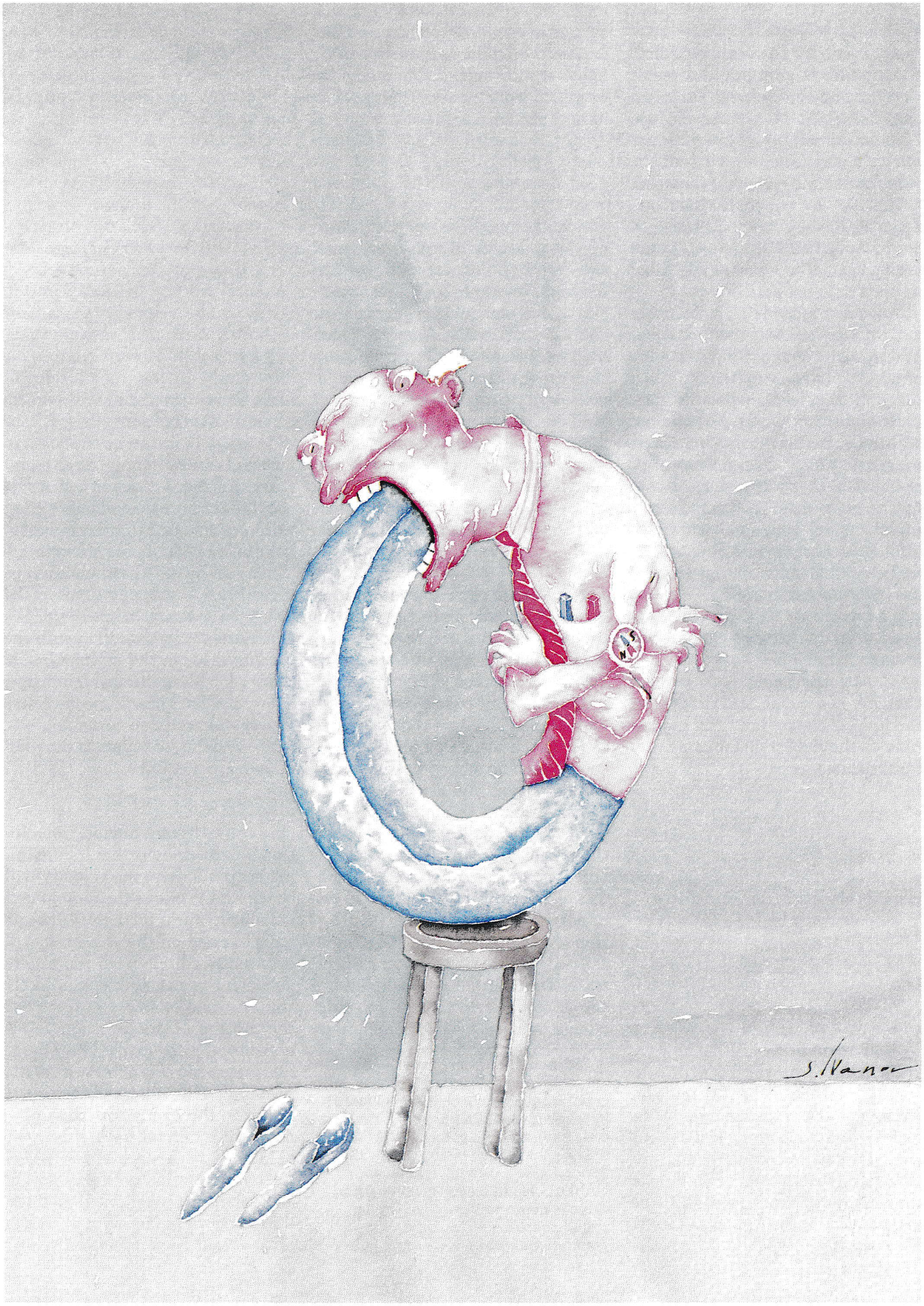
North Pole, South Pole

Before going into the subject of magnetic monopoles, it's worth pointing out that you actually know quite a bit about poles in general. The first ones you ever learned about were likely the Earth's geographic poles—the North and South poles, as they are called. The next ones you learned about were the magnetic north and south poles. These were first noted in 1269 by Petrus Peregrinus de Maricourt, a French military engineer, who noticed that the lines of force around a lodestone seemed to originate from two distinct locations. The simplest example of these today are the painted ends of the magnetized needle in a compass. Usually the end of the compass that points north is painted red and is called the north-seeking (or more simply, just the north) magnetic pole. The other end of the needle, which is often painted blue, is called the south magnetic pole.

Of course you know that as far as magnetism is concerned, north poles are attracted to south poles, and so we can conclude that near the Earth's north geographic pole (where the polar bears live) lies a south magnetic pole and near the Earth's south geographic pole (where the penguins live) lies a north magnetic pole. In fact these two distinct kinds of poles are not coincidental. The south magnetic pole lies a little to the south of the north geographic pole, so that in my location of Toronto, Canada, my compass actually points about 10° west of true north. The fact that the Earth's geographic and magnetic poles don't coincide makes for an interesting study in itself. It turns out that the Earth's magnetic poles wander about in time and that a good topographic map will tell you just how much you can expect the north magnetic pole to vary in position over the years. Moreover, there is geologic evidence that the Earth's magnetic poles have been reversed in the past and that these reversals happened relatively suddenly.

But what is always true about magnetic poles, whether in the Earth, a compass needle, or any magnet or magnetic device, is that they always occur in north–south pairs. These pairs are called magnetic dipoles. A magnetic monopole would be the occurrence of an isolated north or south pole, unpaired. There are pretty good reasons to think that such a beast should exist, and most of these reasons have to do with the symmetry of nature. You may hear a physicist say that the Maxwell equations that govern electromagnetism become perfectly symmetrical if magnetic monopoles exist. This in itself is not a bad reason to believe in monopoles, but we shall see that Dirac, working in 1931, found a way to explain one of the great mysteries of physics, and the matter rested on there being at least one monopole somewhere in the universe. We'll look into this and see how physicists are looking for monopoles. But first we had better fill in some basics.

Art by Sergey Ivanov


S. Ivanov

Electric and magnetic monopoles

Let's go back to some basic electricity and magnetism and recall that the force on a charge due to an electric field is $F = qE$. An electric "dipole" is composed of two separated charges $+q$ and $-q$. We could be very formal and call each charge an "electric pole," one positive and one negative. The electric pole strength could be defined as $q = F/E$. One such pole on its own would be called an electric monopole.

If we do the same thing for magnetism, we would say that two equal but opposite magnetic poles with pole strengths $q^* = F/B$ constitute a magnetic dipole. Here B is the magnetic field. A magnetic monopole would be the occurrence of a lone magnetic pole, north or south, of pole strength q^*. Since we know how to investigate problems involving electric monopoles, otherwise known as charges, we know how to make calculations involving magnetic monopoles. We can, for instance, write down what the magnetic field due to a single isolated monopole must be!

The electric field due to a single charge q is a radial field that obeys the inverse square law. The magnitude of the field a distance r away from the charge is

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

where $\varepsilon_0 = 8.85 \cdot 10^{-12}\ \mathrm{C^2/N \cdot m^2}$. Hence the magnetic field due to a monopole must be

$$B = \frac{\mu_0}{4\pi}\frac{q^*}{r^2},$$

where $\mu_0 = 4\pi \cdot 10^{-7}\ \mathrm{N/A^2}$.

You might be wondering how we knew exactly what the magnetic constant $\mu_0/4\pi$ had to be. I'll tell you without proof or explanation (you can look forward to learning more about this as your physics education progresses) that the root of the ratio of the electric constant $k_e = 1/4\pi\varepsilon_0$ to the magnetic constant $k_m = \mu_0/4\pi$ must equal the speed of light c. (In fact, the magnetic constant μ_0 in S.I. units is defined in terms of our choice for the electric constant and the speed of light—it's not an experimentally determined value in itself.) From this follows the amazing relationship

$$\sqrt{\frac{1}{\mu_0\varepsilon_0}} = c,$$

which ties together electricity, magnetism, and light as parts of the same physics. You already know, for instance, that light (or, more generally, electromagnetic radiation) is composed of oscillating electric and magnetic fields. For our purposes, this means that we have a good working formula for the magnetic field due to a magnetic monopole.

Sadly, and perhaps surprisingly, a lone magnetic monopole has never been found, but this does not prevent us from wondering what its properties might be.

Properties of a magnetic monopole

In 1904 J. J. Thomson studied the theoretical motion of an electron in the vicinity of a magnetic monopole. Of course, he wanted to know how he would recognize a magnetic monopole if he was ever to be so lucky as to come across one. We are experienced in testing for the presence of electric monopoles (charges), and I'll remind you how we do this. We fire a charge at it and look for its characteristic deflection.[1] To detect the presence of a magnetic monopole, we will also imagine firing a test charge toward it, but we must understand what kind of a deflection we are looking for. The force exerted by a charge on another electric charge is in the radial direction; the force exerted on a moving charge by a magnetic monopole is more complicated and will require a careful explanation.

[1]Although I haven't mentioned it yet, we can also think of a gravitational field as being generated by a gravitational monopole (usually called a mass). We detect the presence of such a pole also by looking at the deflection of a moving test mass. Unlike poles in electromagnetism, there is no negative pole in gravity, and so there is no repulsive gravitational force.

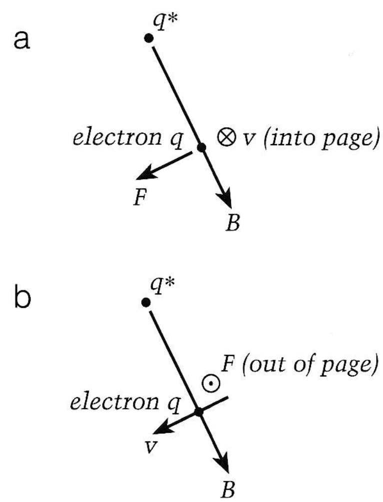


Figure 1

In figure 1a, we set up the configuration of our imaginary experiment. Our test charge of mass m and charge q will have an initial velocity v toward our fixed and stationary magnetic monopole of pole strength q^*. We can set up our experiment with a variable aiming error b. This is the distance between two lines, both parallel to the charge's velocity, with one passing through the monopole and the other passing through the charge. All we need to know is the force on a moving charge due to a magnetic field—and this, of course, is the well-known Lorentz force:

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}.$$

One immediately obvious conclusion we can draw is that if our aiming error b is zero, the velocity will be parallel to one of the monopole's radial field lines and there will be no force exerted on the charge. It will eventually collide with the monopole. This is an extremely unlikely scenario, and so we will sketch out the motion of our test charge for nonzero aiming errors. Our charge finds itself moving across magnetic field lines. If the magnetic field were uniform, the charge would simply move in a helical path about the field lines. In the special case where the charge's velocity makes a right angle with the magnetic field lines, the charge would move in a circular orbit. Most high school students study these situations and even learn to calculate the radius of the charge's circular (or helical) motion. Our situation is different for two very important reasons. The field is not uniform—it is diverging. So the charge finds itself entering regions of stronger and stronger magnetic field.

We can understand the charge's trajectory if we break its behavior down into two stages. In figure 1, the two essential characteristics of the charge's motion are summarized. At all times, one can imagine breaking the charge's velocity into two components: the radial component, parallel to the magnetic field lines; and the tangential component, which will always be perpendicular to the field lines. Initially the charge's motion is largely, but not entirely, radial as it has been directed toward the monopole with only a small aiming error. In figure 1a, the initial, small, tangential component of the velocity of the charge is shown as being "into the page." There is no force on the charge due to its radial component of the velocity. The force on the charge due to its tangential component will also be tangential as shown (try out your right-hand rule here to make sure you're with me on this).

The first general characteristic of our test charge is now apparent. As it approaches the monopole, it will exchange translational motion for circular motion. We know that an exchange of translational kinetic energy for rotational kinetic energy must take place, because the magnetic field is conservative. Since the force on a moving charge is always perpendicular to the charge's velocity, the force cannot change the speed of the charge, only its direction. The charge that initially was directed toward the monopole will slow its direct approach toward the monopole while picking up a circular motion. Since the field grows in strength as the charge nears the monopole, this effect becomes stronger and stronger until the charge has a purely circular motion and is no longer approaching the monopole. Although the charge had a largely radial velocity to begin with, it will eventually have a purely tangential motion.

In figure 1b, we consider the force on the tangential component of the charge's velocity due to the magnetic field. In this case, the force is "out of the page." So here's the second general characteristic of our charge's motion: as the charge gains rotational motion, a repulsion between the charge and the monopole grows that must reconvert the rotational motion back to translational motion but now in the opposite direction from whence it came. Putting this all together, we would expect our charge to approach the monopole but begin spiraling in along one of the field lines and being repulsed as it does so. At some point, the charge will have only circular motion and, still being repulsed, will begin to spiral out again along the same field line. Finally, we would expect the charge to come firing back at us with the same speed that we initially gave it. This motion is not unique to a purely radial field such as that from a monopole. Any strongly diverging magnetic field will reflect a moving charge in this way.

The motion of the charge is plotted in figure 2 in three dimensions. The initial velocity of the charge was in the z direction and the initial position of the charge was at the point $x = 1$, $y = -5$ relative to a monopole at the origin. The variables plotted are dimensionless quantities

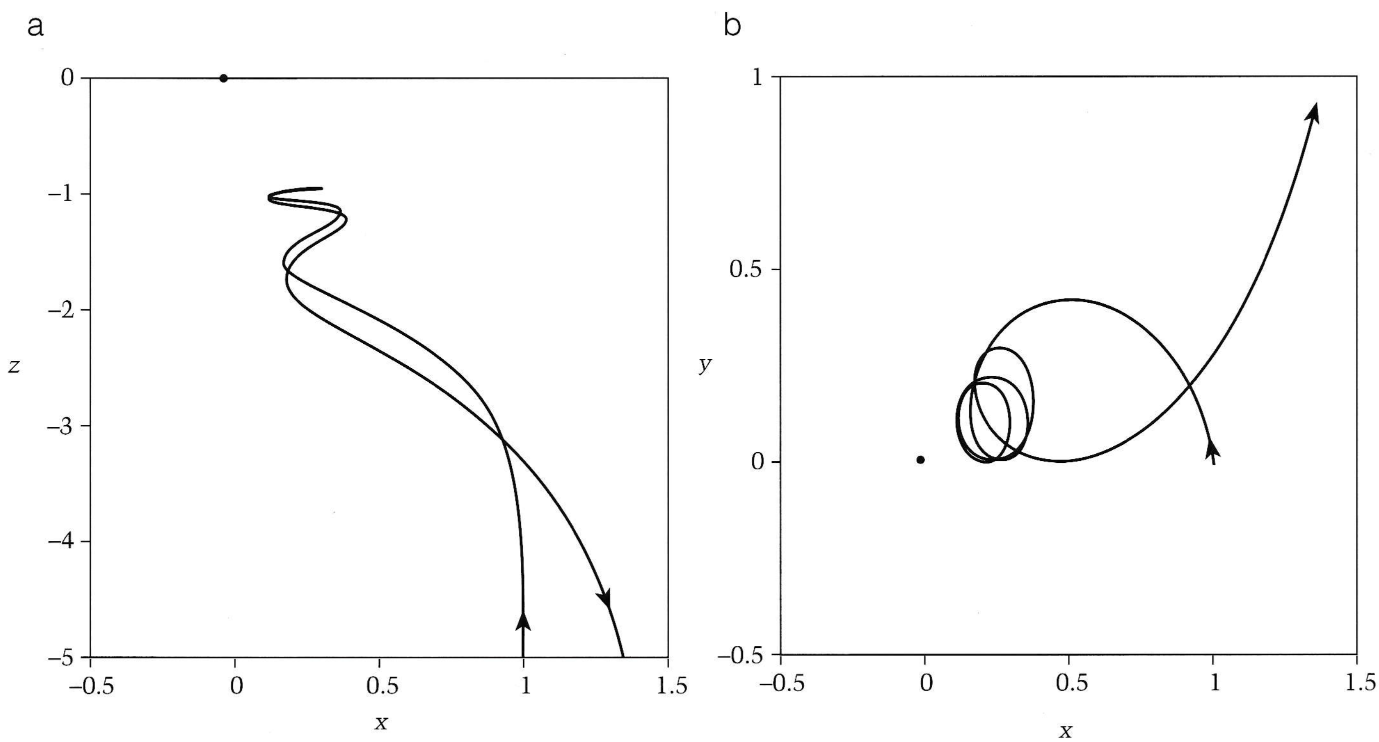


Figure 2
Computer calculations of an electric charge's trajectory near a magnetic monopole.

used to describe the charge's motion. Figure 2a represents a "top" view of the charge's x, z motion. Notice how the charge exchanges its translational motion for rotational motion, at first slowly and then more and more quickly as the magnetic field strength increases. Figure 2b shows an end-on view of the charge's motion. The charge is initially coming toward us (out of the page) and reaches a point where its motion is purely circular. The repulsive forces then send it back in the $-z$ direction, re-exchanging its circular motion for translational motion. The spiral motion of the charge along a radial field line is clearly seen in the two views shown.[2]

The problem with Thomson's question of 1904 is that it assumes that we might somehow be in a position to fire electrons at a stationary monopole. Solving this problem certainly gives us insight into the properties of a monopole, but we are far more likely to have one go zipping by us than to trip across one just sitting around. We'll have to get a bit more sophisticated in our ideas for detecting monopoles, but first it's useful to realize that there is much in nature that we can better understand in terms of our monopole physics.

Magnetic monopoles in nature

The title of this section is a bit of a cheat, since magnetic monopoles don't exist (or at least, one has never been seen). Nevertheless, our studies of monopole physics will allow us to understand nature better. To see this, let's first investigate the field produced by a magnetic dipole. In figure 3 we have a magnetic dipole and we wish to calculate the magnitude of the field at a point x

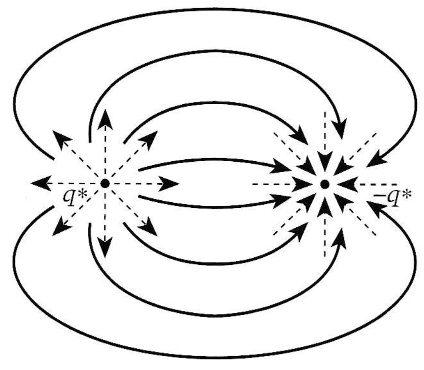


Figure 3

Magnetic field (solid lines) of a dipole, which is the superposition of the two radial fields (dotted lines) surrounding each separate pole.

from the dipole's midpoint and on a line along the dipole's axis. The classic field lines due to a dipole (think of the Earth or a simple bar magnet) have been sketched in. The magnitude of the field is found by superposing the individual monopole fields:

$$B(x) = \frac{\mu_0}{4\pi}\frac{q^*}{r_1^2} - \frac{\mu_0}{4\pi}\frac{q^*}{r_2^2},$$

where $r_1 = x - l/2$ and $r_2 = x + l/2$. Far from the dipole, for $x \gg l$, the expression simplifies to give

$$B(x) = \frac{\mu_0}{4\pi}\frac{2m}{x^3},$$

where $m = q^*l$ is the magnetic dipole moment. This is the correct expression for the on-axis dipole field (which is not usually calculated in this way). You might have a little fun and write down the expression for the magnetic field far from the dipole but along a line perpendicular to the line of the dipole. This is also an inverse-cube field, and so one could argue that the magnetic field due to a magnetic dipole, far from the dipole, goes as $1/x^3$.

The point of all of this is that the field due to a magnetic dipole is the superposition of two radial monopole fields. The above calculation is just one way of confirming this. Remember that the separation between the poles of a dipole is l. Suppose we examine the dipolar field at a distance $r \ll l$ from one of the poles. Essentially we would see only the monopolar radial field so that, while monopoles do not exist in nature, the behavior of a charged particle close to one of the poles in a dipolar field (or more generally, any multipolar field) is essentially that discussed in the previous section. To find a stunning example of this, let's visit the Van Allen Radiation Belts.

In May of 1958, an American physicist named J. A. Van Allen announced that there was an intense belt of high-energy particles surrounding the Earth. The announcement was based on the readings from a Geiger counter he had placed on the rocket that launched the first US satellite (Explorer I). Subsequent studies showed two belts: an inner belt within two Earth radii from the Earth's center and an outer belt between two and eight Earth radii. The particles are largely protons originating from the high-energy cosmic rays that impinge upon the Earth's atmosphere. The density of the atmosphere at the lower altitudes of the belts (between 400 and 1,000 km above sea level) is so low that the protons may travel hundreds of Earth radii between collisions with atmospheric molecules. The protons in the inner belt may have energies as high as $3.0 \cdot 10^7$ eV, and the intensity is such that as many as 20,000 particles may cross a 1-cm^2 area each second. It's estimated that, to produce such figures, an average proton must remain trapped within the

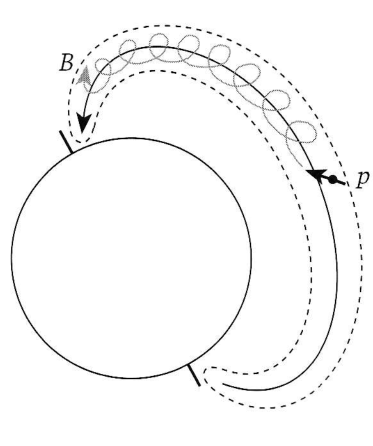


Figure 4

[2]I can't take total credit for the idea of making these computer plots. In 1988 I gave a talk on monopoles at a training session for the Canadian International Physics Olympiad Team. One member, David Hogg, sent me some similar computer plots that he did in the physics computer lab in his first year at MIT. It's always satisfying when a former student becomes my teacher.

belt for 10 years! What could account for this containment of high-energy charged particles?

In figure 4, the Earth's magnetic field is shown with the inner Van Allen Belt sketched in. The trajectory of a proton initially headed out of the belt is shown. As in our earlier discussion, the proton will spiral in along a field line toward a magnetic pole, but will eventually be reflected back upon its initial line. Reflecting back from the other pole in a similar fashion, the proton could be trapped for years.

It's time now to return to the main thrust of this paper—investigating the physics of monopoles. I promised we'd explore a great mystery in physics, and so—here we go.

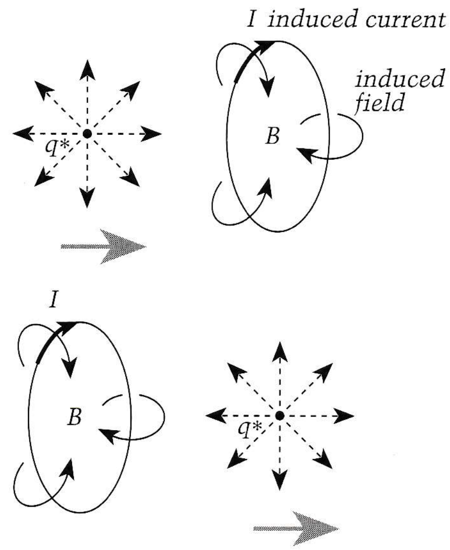


Figure 5

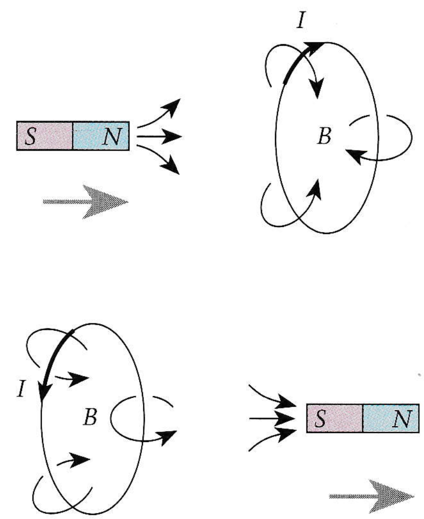


Figure 6

Dirac monopoles

In 1931 Paul Dirac used quantum mechanics to study the properties of a magnetic monopole and found a possible answer to the question of electric charge quantization. It remains to this day a mystery why electric charge exists only in multiples of the fundamental charge $e = 1.6 \cdot 10^{-19}$ C. Dirac found an expression for the magnetic pole strength that seemed to indicate that if at least one magnetic monopole exists in nature, electric charge would necessarily be quantized. Dirac used some pretty advanced concepts from quantum mechanics to do this, but using our modern understanding of superconductivity, we can avoid the advanced math and get to the heart of the matter.

A superconductor allows current to flow with absolutely no resistance. Many metals become superconducting below a very low temperature (often comparable to the temperature of liquid helium). Recently there has been a great deal of interest in the field of high-T_c superconductivity, where some special compounds have been made to superconduct at high critical temperatures—in the neighborhood of $T_c \cong 100$ K. Imagine if you will a loop made of superconducting wire. An emf induced in the wire for however short a period of time will cause a permanent current to flow. There is no resistance to damp out the induced current. We could induce this current by passing a magnetic flux through the loop. A magnetic flux is the product of the magnetic field passing through a current loop and the loop's enclosed area. Faraday's law states that

$$\text{emf} = -\frac{d\phi}{dt},$$

where ϕ is the magnetic flux through the loop of wire. If at any time the flux through the loop is changed, an emf will be induced in the loop and hence a current. The negative sign in Faraday's law is a nod toward Lenz's law, which states that the induced emf will oppose the change that caused it.

Imagine as well that a monopole with pole strength q^* (as in figure 5) is directed toward our superconducting loop. An initial current I would be induced, producing a magnetic field B through the loop, which would, in turn, oppose the direction of the field due to the monopole that caused the induction. After the monopole has passed through the loop, the induced current still flows in the same direction so as to produce a field that will replace the diminishing monopole field. So, as a monopole passes through the loop, a current will flow in only one direction.

This is very different from the current induced in a current loop as a magnetic dipole is passed through (figure 6). In this case, the current induced as the magnetic dipole approaches the loop is in the same direction as that for the monopole, but the induced current for the receding dipole is in the opposite direction.

The current induced in a superconducting loop by a passing monopole will remain long after the monopole has gone. This signature event would allow one to detect the passage of a magnetic monopole.

Many experiments have been designed along these lines. One problem that arises is to manufacture a large enough superconducting detector that can be shielded from external variations in the ambient magnetic field. On February 14, 1982 (St. Valentine's Day), a physicist named Blas Cabrera, working on an experiment at Stanford University, recorded the signal of a single large candidate event that gave exactly the right signature for the induced current. Cabrera used a four-turn, superconducting loop with an area of 20 cm^2. The "Valentine's event" was stated to have an uncertainty of only ±5%. Even though Cabrera could not attribute the event to any cause other than the passage of a monopole, it is not generally accepted as proof that

CONTINUED ON PAGE 46

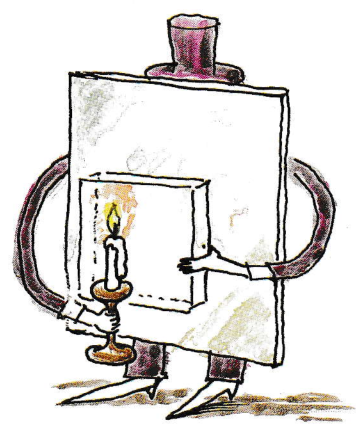
Just for the fun of it!

B141

River traffic. A raft and a motorboat set out downstream from a point A on the riverbank. At the same moment a second motorboat of the same type sets out from point B to meet them. When the first motorboat arrives at B, will the raft (floating with the current) be closer to point A or to the second motorboat? (G. Galperin)

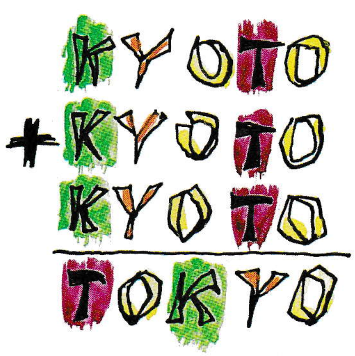
B142

Double-edged ruler. Construct the center of a circle drawn on the plane using only a ruler with two parallel edges whose width is smaller than the diameter of the circle. (A. Demidov)



B143

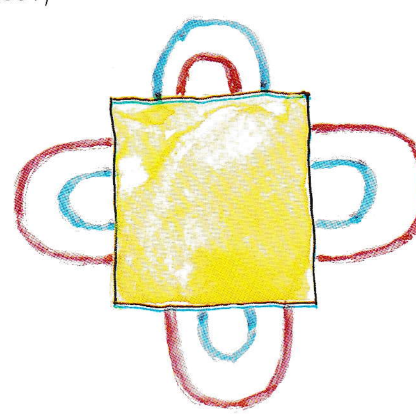
Candle in front of a mirror. The image of a candle is seen in a mirror. What will happen to the image if a sheet of glass is placed between the mirror and the candle? (The orientation of the glass is the same as that of the mirror.)



B144

Number system unknown. Find the number n such that the alphanumeric equation KYOTO + KYOTO + KYOTO = TOKYO has a solution in the base-n number system. (As usual, each letter in the equation denotes a digit in this system, and different letters denote different digits.) (V. Dubrovsky, A. Shvetsov)

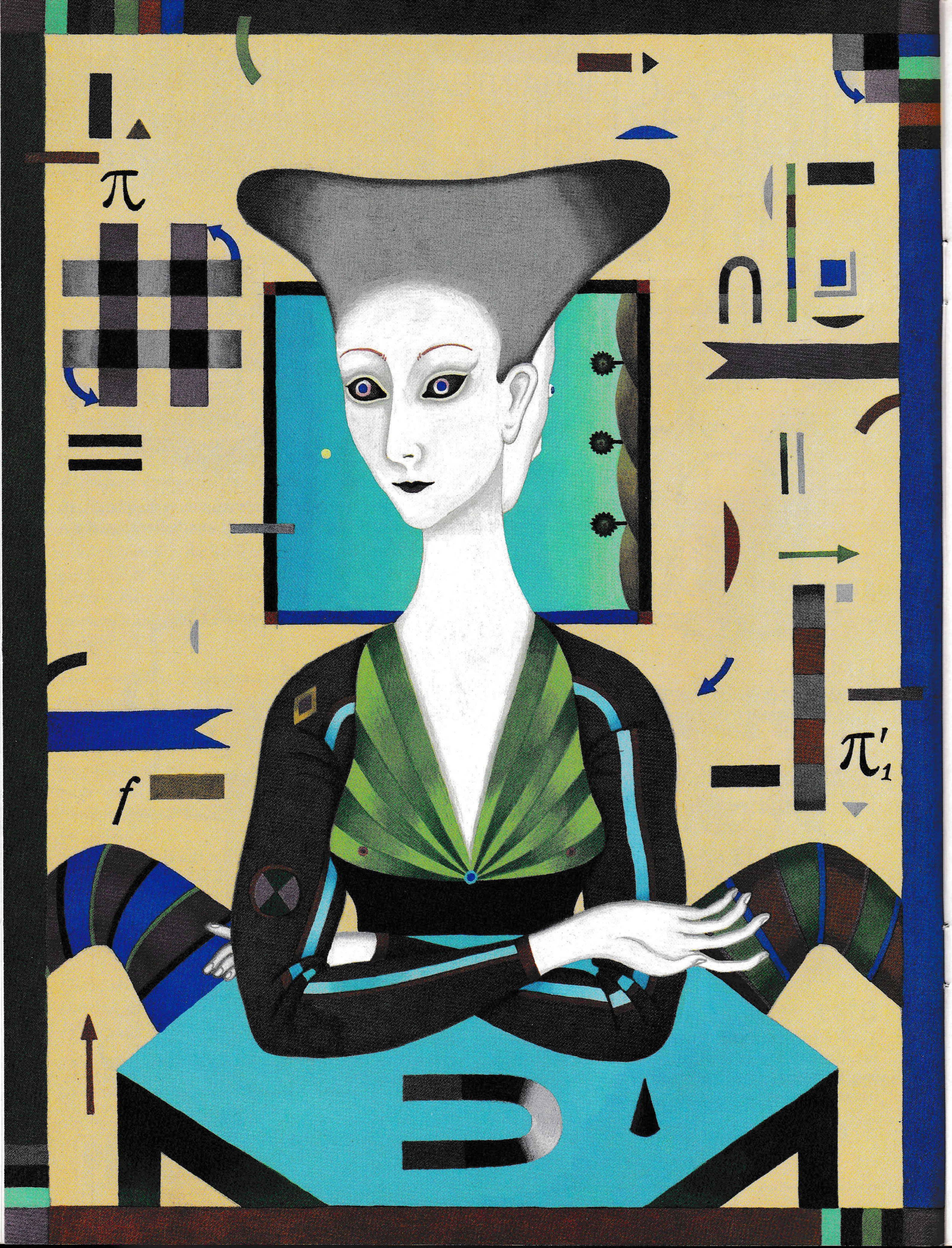
B145



X-ray vision required. Three red and three blue disjoint loops are drawn on the plane. A part of the figure is covered with a sheet of paper so that one loop is covered completely and all the others are partially visible (see the figure). What color is the covered loop? (V. Proizvolov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 59

Art by Pavel Chernusky


π
f
π'₁

Smale's horseshoe

Mathematical footwear that left an imprint

by Yuly Ilyashenko and Anna Kotova

WHEN THE AMERICAN mathematician and Fields Prize winner (1966) Stephen Smale offered his "horseshoe" to the world some 30 years ago, it caused a sensation in the theory of differential equations. And yet this construction is simple enough to be presented within the framework of the high school math curriculum.

Symbolic dynamics

To get our feet wet, we begin with a problem introducing symbolic dynamics that was used in the 56th Moscow Mathematical Olympiad:

For any two real numbers a and b consider the sequence

$$p_n = [2\{an + b\}],\ n = 0, 1, 2, \ldots$$

(where $\{x\}$ and $[x]$ denote the fractional and integer parts of a number x, respectively).[1] Any set of successive terms of this sequence will be called a word. Is it true that any ordered set of k zeros and ones is a word in the sequence p_n for a certain pair of numbers a and b if (a) $k = 4$, (b) $k = 5$?

That is, we form the arithmetic progression $na + b$, then take the fractional part of each term. We double each fractional part, then form a new sequence by taking the integral part of each result. Experimentation for different values of a and b will show that the sequence p_n consists of ones and zeros. Indeed, the fractional part varies in the interval $[0,1)$, so the integer part of twice the fractional part is either 0 or 1.

This problem has a nice geometric interpretation. Consider a circle whose circumference has length 1, touching a number line at its origin. Now imagine that both halves of the axis are wound about the circle. Then any point x on the axis fits onto the point on the circle obtained by rotating the origin through the angle $360°\{x\}$ about the circle's center. Clearly, all points of the axis that differ by an integer fit the same point on the circle, because the corresponding rotations differ by an integer number of full turns, and so are actually the same. So, the points x and $\{x\}$ are represented by the same point of the circle. Now the problem about the sequence p_n turns out to be a problem about *symbolic dynamics* of rotations of a circle. What do we mean by this?

Consider the points on the circle corresponding to b, $a + b$, $2a + b$, ... (see figure 1). As we noted above, they can be viewed as representing the sequence $\{na + b\}$, $n = 0, 1, 2, \ldots$. We also see that all these points are obtained from the first one, b, under successive rotations through the angle a (here and below all angles are expressed in fractions of a full turn). If point $na + b$ lies on the upper semicircle from 0 (inclusive) to 1/2 (exclusive), the corresponding value of p_n is 0. For points on the lower semicircle, $p_n = 1$.

Now we can explain the phrase "symbolic dynamics." *Dynamics* refers to the motion of a point around the circle under repeated (iterated) rotations. *Symbolic* refers to the character of the information about the positions of the point: we're interested only in what half of the circle the point belongs to, rather than its exact location. Symbolic dynamics allows us to solve the problem easily.

If there are many successive zeros in our sequence, the angle of rotation, or rather its fractional part, is small, because our moving point

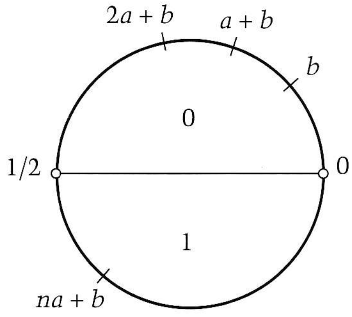


Figure 1

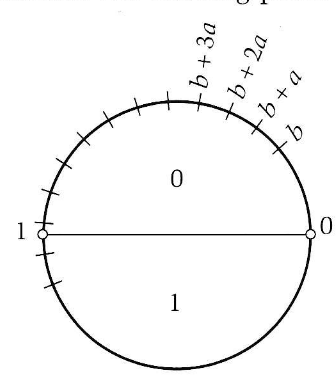


Figure 2

[1]More formally, $[x]$ is defined as the largest integer not greater than x, while $\{x\}$ is defined as $x - [x]$.—*Ed.*

Art by Mariya Konstantinova

stays on the upper semicircle under a sufficiently large number of successive rotations by the angle a. When it finally gets into the lower semicircle, it will have to stay there for a rather long time as well (because a is small—see figure 2). Therefore, a sequence in which a solitary 1 pops up in the midst of a long series of 0's can't be a word in p_n.

In particular, we can show that for $k = 5$, the word 00010 can never occur. Indeed, the fact that there are three zeros in a row in p_n means that $|a| < 1/4$ of a full turn. At the same time, the segment 010 shows that the point made a jump from the upper semicircle down to the lower one, and then back to the upper one. This is possible only when $|a| > 1/4$ of a full turn. This contradiction shows that the answer for $k = 5$ is no. As for the case $k = 4$, any of the 16 possible k-digit words of zeros and ones can occur in p_n (this can be proved by direct search, which we leave to the reader as an exercise).

Vocabulary

This olympiad problem is a good example of "symbolic dynamics." We would like to consider more general situations like this. To do so, it will be convenient for us to introduce a number of general notions. Consider a map of a certain domain—say, from the plane, although this restriction isn't obligatory—into another (plane) domain. The first domain (where the map is defined) is called the *phase space* of the map. The maps resulting from successive repetitions of the original map are called its *iterates*. More exactly, these are *positive* iterates, while the repetitions of the inverse map are *negative* iterates (with respect to the original map). Iterates are defined on a domain that, in general, is a part of the entire phase space. Throughout this article we'll use the notation $f^n(x)$ for the nth iterate of a map f. (For $n > 0$, it's f applied n times; for $n < 0$, it's the inverse map f^{-1} applied $|n|$ times; and for $n = 0$, it's the identity map.)

We define the *orbit of a point x under a map f* as the set of the images of x under all iterates f^n of f ($n \in \mathbf{Z}$). This set is defined only for points x in the intersection of the domains of all iterates, both positive and negative.

Now imagine that the phase space is split into two parts, S_0 and S_1. To each point whose orbit is well defined we assign its "fate." That is, we consider a sequence a_n of zeros and ones defined by the following rule: $a_n = 0$ if the point $f^n(x)$ belongs to S_0, and $a_n = 1$, if $f^n(x) \in S_1$. We denote this sequence by $\omega(x)$ and call it the *fate of point x*. Along with the "full fate" (or just "fate"), which is obtained when n ranges from $-\infty$ to ∞, we'll sometimes use its "segments": the future and past fates, which correspond to $n \geq 0$ and $n < 0$, respectively, or finite fates, corresponding to various finite segments of numbers $n_1 \leq n \leq n_2$.

Symbolic dynamics deals only with this information about orbits—it doesn't care about the exact position of the point $f^n(x)$, but asks only which part of the phase space this point belongs to. It studies natural questions similar to those that were discussed in the olympiad problem above: can any finite or infinite sequence of ones and zeros be realized as a fate of a certain point? If it can, how many points have a given fate? Solving our first problem, we found that some sequences can't be fates of points. The second question, about the set of points with a given fate, wasn't discussed in this setting and is left to the reader as another very useful problem.

The horseshoe map

In this setting, we can begin to talk about Smale's horseshoe. We must admit right from the start that this map doesn't look like a horseshoe at all. The origin of the name will be explained at the end of the article.

Take a unit square. Divide it into five equal vertical strips and, similarly, into five equal horizontal strips (fig. 3). Leave only the second and fourth vertical strips; denote the left one S_0, the right one S_1; similarly, denote by S_0' and S_1' the second and fourth (from the top) horizontal strips.

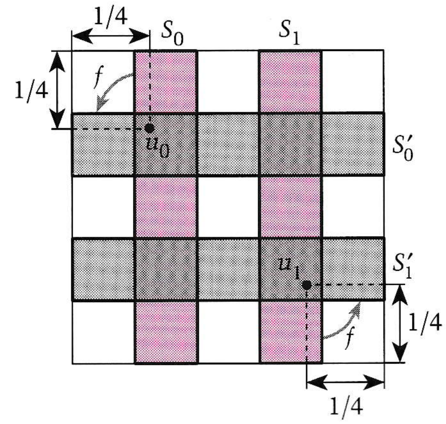


Figure 3

Now consider the map that contracts the rectangle S_0 by a factor of 1/5 vertically, stretches it fivefold horizontally, and lays what's obtained over S_0'. Some thought will show that the point in the top left corner of our unit square at a distance 1/4 from the top and left sides is left intact under this mapping. So this turns out to be the point relative to which the vertical contraction by a factor of 1/5 and the horizontal dilation by a factor of 5 are performed.

This is how our map will work in the rectangle S_0. But its domain consists of two rectangles—S_0 and S_1. On S_1 it's defined quite similarly except that the top left corner of the big square must be replaced with its bottom right corner. The map contracts S_1 five times vertically, stretches it five times horizontally, and lays it over S_1'. The point in the original square at a distance of 1/4 from its bottom and right sides remains fixed (fig. 3).

Thus, the horseshoe map is the map shown in figure 3. Its phase space is the union of the rectangles S_0 and S_1; its range is the union of S_0' and S_1'.

It turns out that the set of points for which full orbits of this map are defined is much leaner than the phase space itself. To describe it, we'll need a construction of *Cantor's perfect set*, also called the Cantor discontinuum, or just the Cantor set.

Cantor's perfect set

The definition of the Cantor set we'll give here is not exactly the usual one. Consider the segment

[0, 1]. Divide it into five equal parts, leave the second and fourth intervals, and erase the other three. Then do the same operation with each of the remaining segments: divide it into five parts, delete the parts at the ends and in the middle, and so on.[2]

The segments left after the first step, [1/5, 2/5] and [3/5, 4/5], will be called the segments of the first rank. Then, by induction, the segments of the nth rank are constructed as the corresponding pieces of the segments of the $(n-1)$st rank. Denote by W_n the union of all the segments of rank n.

We define *Cantor's perfect set C* as the intersection of all sets W_n, $n = 1, 2, 3, \ldots$.

Is this intersection nonempty? How many points does it contain?

It turns out that *the set C has exactly as many points as the entire segment* [0, 1].

To prove this remarkable statement, let's write down all numbers from 0 to 1 in the quinary number system (that is, to the base 5 rather than our ordinary 10) and consider the segments of rank n:

$$[1/5, 2/5] = [0.1, 0.2],$$
$$[3/5, 4/5] = [0.3, 0.4]$$

(fractions on the right side are *quinary*—that is, the expression $0.a_1a_2a_3\ldots$ denotes the number $a_1/5 + a_2/25 + a_3/125 + \ldots$, where $0 \le a_i \le 4$). We see that the numbers from these segments can be described as those whose first digit after the (quinary) point is 1 or 3. (The right endpoints of our segments seem to be exceptions. However, we could write them as periodic fractions $0.2 = 0.1\overline{4}$ and $0.4 = 0.3\overline{4}$, or simply ignore them—they'll be deleted in the next step anyway.) What happens when the segments of rank 1 are subdivided into segments of rank 2? Two smaller segments of length 1/25 remain from each of them:

$$[1/5 + 1/25, 1/5 + 2/25],$$
$$[1/5 + 3/25, 1/5 + 4/25],$$
$$[3/5 + 1/25, 3/5 + 2/25],$$
$$[3/5 + 3/25, 3/5 + 4/25].$$

In the quinary notation, the numbers in these segments are written as 0.11..., 0.13..., 0.31..., 0.33..., respectively, so their second digits after the point are 1 or 3 as well.

And it's clear that this will continue forever: for any number from the set W_n its first n quinary digits are only ones and threes. (You can prove this rigorously by induction.)

So, any point of Cantor's set defined above can be written as an infinite quinary fraction all of whose digits are ones and threes.

Conversely, any such fraction represents a point from the set C. Indeed, if the first digit after the point in a quinary fraction is 1, the corresponding point belongs to the left segment of the first rank; if this digit is 3, the point lies in the right segment. Since the second digit is also 1 or 3, the point belongs to one of the segments of rank 2. Then, by induction, we can show that this point belongs to W_n for any n—that is, to the intersection of all the sets W_n, which is the set C.

Now, if we replace all the ones in our fractions with zeros, and all threes with ones, we'll get all possible infinite fractions consisting of ones and zeros. But they can be thought of as *binary* representations of all the numbers from 0 to 1! This establishes a one-to-one correspondence between Cantor's set and the entire segment [0,1].

It's an interesting exercise to verify that the total length of all the intervals that are deleted from [0,1] in the course of constructing Cantor's set is 1. This means that we've built a subset C of the unit interval that contains as many points as the entire interval, but doesn't take any room on this interval at all!

Symbolic dynamics of the horseshoe map

Before we get to points that have a prescribed *infinite* fate on the horseshoe map, let's consider a simpler question: what does the set of points with a given finite fate look like?

For instance, let $a_0 = 0$ or $a_0 = 1$. What is the corresponding set of points in the phase space? The answer can be read directly from the definition. The fact that $a_0 = 0$, by definition, means that the image of x under the zeroth power of f belongs to S_0. But f^0 is the identity mapping: $f^0(x) = x$, and so the condition $a_0 = 0$ means that point x lies in S_0. And $a_0 = 1$, of course, corresponds to S_1.

Now let's examine the same question for a finite fate of length 2. What is the set of points with given values of the pair a_0a_1 (which can be 00, 01, 10, or 11)?

Look at figure 4. The points 00 are those that lie in S_0 along with their first images. This means that the image of such a point belongs to the second of the five equal squares into which the rectangle S'_0 is divided. What are the sets (pre-images) from which these squares are obtained under the map f? The answer is clear: we must cut the rectangle S_0 into five equal vertical rectangles. The map f takes the kth of these rectangles (counting from the left) into the kth square from the left in S'_0. So the set of points with the future fate 00 is the second from the left vertical strip in S_0. Similarly, the fate 01 awaits the points from the fourth strip from the left in S_0. The sets of points with the future fates

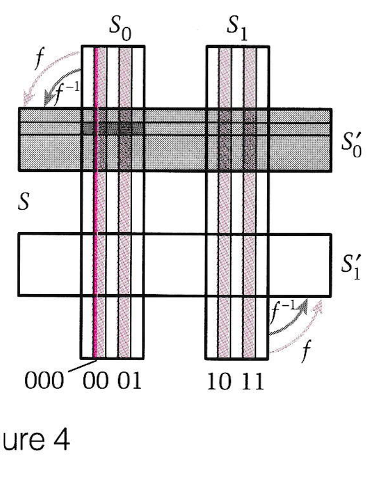


Figure 4

---

[2]The classical construction divides the segment into three parts, erases the central one, puts the other two through the same operation of erasing the central third, and proceeds in the same way to infinity. (A related construction, "Cantor's staircase," was discussed as problem 7 in "Bushels of Pairs" in the November/December 1993 issue.) We can satisfy ourselves that the "two-fifths" and "two-thirds" constructions are equivalent.

10 and 11 are the second and fourth, respectively, of the five vertical strips in S_1.

Now it's easy to draw the points whose future fate begins with three given symbols $a_0a_1a_2$. Consider, for instance, the set of points with the future fate 000. They all lie in the strip with the future fate 00. The map f takes this strip into a narrow horizontal rectangle of height 1/25 and length 1. But we want the points in question to have the fate 000. This means that their images under the second iterate f^2 must lie not only in the narrow rectangle but in S_0, too. The intersection of S_0 with this rectangle is a $1/5 \times 1/25$ rectangle, the darkest in the figure, and it is the image of the set with the fate 000 under f^3. Therefore, this set itself is the vertical narrow red strip (of width 1/125)—the second of the five strips into which the 00-strip can be divided.

These arguments can be considered indefinitely and lead to the following lemma. In its statement, we use a coordinate system whose origin is at the lower left-hand corner of the original unit square. We express these coordinates in quinary notation.

LEMMA 1. *The set of points with a given infinite future fate* $\omega^+ = a_0a_1...a_n...$ *is a vertical segment consisting of points whose x-coordinates all equal* $a(\omega^+) = 0.\alpha_0\alpha_1...\alpha_n...$, *where* $\alpha_n = 1$ *if* $a_n = 0$, *and* $\alpha_n = 3$ *if* $a_n = 1$, *and whose y-coordinates vary from 0 to 1.*

To prove this we note that, as was illustrated above, all the points with a given fate of length n have their x-coordinates in a fixed segment of the nth rank—one of those that appeared in the construction of the Cantor set. A rigorous proof is left as a problem for the reader.

Now we come to the main result of our research.

THEOREM. *Any sequence* ω *of zeros and ones that is infinite in both directions can be realized as the fate of one and only one point. The set of all points that have a well-defined infinite fate consists of the points whose x-coordinates belong to the Cantor set C on a horizontal side of the square and whose y-coordinates belong to the Cantor set on a vertical side.*

Proof. The given sequence $\omega = ...a_{-2}a_{-1}a_0a_1a_2...$ can be represented as a combination of two subsequences—one infinite to the right, the other to the left: $\omega^+ = a_0a_1a_2...$ and $\omega^- = ...a_{-3}a_{-2}a_{-1}$. We need one more lemma to conclude the proof of this theorem. Lemma 2 carries the result of lemma 1 over to sequences infinite to the left, and is proved in exactly the same way.

LEMMA 2. *The set of points with a given infinite past* $\omega^- = ...a_{-n}...a_{-1}$ *is a horizontal segment whose points have the same y-coordinate* $b(\omega^-) = 0.\beta_{-1}\beta_{-2}...$, *where* $\beta_{-n} = 1$ *if* $a_{-n} = 0$, *and* $\beta_{-n} = 3$ *if* $a_{-n} = 1$, *and whose x-coordinates vary from 0 to 1.*

Thus, we've found the set of all points with the future fate ω^+ (it's a vertical segment) and the past fate ω^- (it's a horizontal segment). So what is the set of points with the fate ω? Clearly, it's the point of intersection of these two segments! Such a point always exists and is unique.

So the theorem has been proved, and we can gather the fruits of our labor.

Corollaries

First let's consider the points with the simplest fate—the stable points shown in figure 3. The top left one, u_0, stays in its place in S_0 throughout the entire past and future. Its fate consists only of zeros. Similarly, the fate of the bottom right stable point u_1 consists only of ones.

Besides stable points, there's another kind of interesting point—periodic points. These are the points that come back to their initial positions after a number of iterations of the map f. How many such points are there?

An unexpected and remarkable property of f is that the number of these points is infinite.

COROLLARY 1. *The map f has infinitely many periodic points.*

Proof. A periodic point has a periodic fate—that is, a periodic sequence of zeros and ones. Not only that—our theorem allows us to prove the converse statement: if the fate of a point is periodic, then the point itself is periodic.

Indeed, suppose that for some positive integer p the fate of a point x satisfies the property

$$\alpha_{n+p} = a_n$$

for any $n \in \mathbf{Z}$— that is, it's periodic with period p. Then the point itself returns to its place after no more than p iterations of the map f, because points x and $f^p(x)$ have the same fate, and therefore they coincide by the uniqueness of a point with a given fate. It remains to note that there are infinitely many periodic points, since there are infinitely many periodic sequences of 0's and 1's, and each defines a periodic point.

In the course of its evolution under iterations of the map f, a periodic point x "jumps" between a finite number of positions—namely, x, $f(x)$, ..., $f^p(x)$, where p is its period. Another type of orbit is one that approaches a certain limit point as $n \to \infty$. This means that all the points of such an orbit with sufficiently large numbers stay within a neighborhood of the limit point, no matter how small it is. It's easy to see that the limit point u must necessarily be a fixed point of the map f. Indeed, if points $f^n(x)$ approach point u as $n \to \infty$, then point $f^{n+1}(x)$ approaches $f(u)$, because, by the definition of f, the distance between $f^{n+1}(x)$ and $f(u)$ is no greater than five times the distance between $f^n(x)$ and u, so it also tends to 0. But the sequences $f^n(x)$ and $f^{n+1}(x)$ geometrically are the same (they differ only in how they are numbered), so they approach the same point, and $f(u) = u$. All these considerations can be applied to the limiting behavior of an orbit as $n \to -\infty$ as well.

Combining the past and the future limiting properties of an orbit, we come up with the following definitions: an orbit is called *homoclimic* if its points approach the same fixed point both as $n \to \infty$ and $n \to -\infty$. If they approach different fixed points in the infinite past and infinite future, we say that the orbit is *heteroclimic*.

COROLLARY 2. *The map f has infinitely many homoclimic and heteroclimic orbits.*

To prove this, consider a point whose fate consists of uninterrupted zeros starting from a certain moment in the future and reaching back from a certain moment in the past. There are infinitely many such points, and so infinitely many orbits. Let's show that each of these orbits is homoclimic with the limit at the stable point u_0 from S_0.

Suppose, for instance, that "future zeros" appear at the moment n: $a_n = a_{n+1} = \ldots = 0$. Then the infinite future fate of $f^k(x)$, $k \geq n$ is the same as that of u_0 (for u_0, the fate consists only of zeros). By lemma 1, $f^k(x)$ and u_0 have equal x-coordinates. So $f^k(x)$ for $k \geq n$ always stays on the same vertical segment in S_0 and, by the definition of f, is pulled closer to u_0 by a factor of 1/5 with every application of f. This means that $f^k(x)$ approaches u_0 as $n \to \infty$. The same argument holds for the past except that the vertical segment through u_0 must be replaced with the horizontal one.

Replacing zeros with ones, we'll obtain homoclimic orbits approaching the second fixed point u_1 (in S_1). And if ones are substituted for zeros only in the past or in the future, we'll get heteroclimic orbits.

Notice that the description of homo- and heteroclimic orbits we used here is complete—that is, any of these orbits has a fate that contains uninterrupted zeros and/or ones on both ends.

The next statement generalizes the previous one.

COROLLARY 3. *For any two points x and y there exists a point whose orbit approaches the orbit of x in the future and the orbit of y in the past.*

Try to prove this fact yourself. The result you may want to use (and which was factually used in the proof of corollary 2) appeared in our proof of the theorem above: *any sequence of ones and zeros that consists of the "past" half ω^- and the "future" half ω^+ is realizable as the fate of the point with coordinates $(a(\omega^+), b(\omega^-))$, where a and b are as defined in lemmas 1 and 2.*

We conclude with one more property of periodic orbits, whose proof is also left as an exercise.

COROLLARY 4. *For any period p, there are only a finite number of periodic points.*

This means, in particular, that any periodic point with a given period has a neighborhood free of other points with the same period. Another interesting question that can be asked in this context is how many periodic orbits with a given smallest period p there are.

So why was the example we examined so sensational? Why did it change mathematical thinking with regard to dynamics?

Laplace and Smale: determinism and chaos

About two centuries ago, in his treatise *A Philosophical Essay on Probabilities*, Pierre Simon Laplace wrote: "A rational being that at any given moment knows all the forces animating nature and the relative positions of all its constituent substances could—if its mind were sufficiently comprehensive to subject all these data to analysis—embrace in one formula the motion of the greatest bodies in the universe on a par with the smallest atoms. Nothing would be left uncertain to it—it could take in at a glance both the future and the past."

These words contain a great philosophical discovery: all evolutionary processes in the universe can be described by ordinary differential equations, perhaps in a phase space of a very high dimension.

This concept is based on a mathematical fact situated not on the leading edge of the science of that time but well beyond it—the existence and uniqueness theorem for the ordinary differential equation. This theorem was proved later by Augustine Louis Cauchy.

For a long time it seemed that the philosophy based on this theorem adequately described the reality around us. Smale's horseshoe map provides grounds for a completely different point of view.

Imagine that we observe a process described by the map f above: we keep track of the evolution of points under iterations of this map. As before, we're interested in the fate of a point rather than its actual orbit. Suppose we repeat our experiment twice with the same point: both times we choose a point with the same initial coordinates and see where it's taken by our map and its iterates.

The problem is that we can't measure coordinates with complete accuracy. If initial coordinates are specified by the experimenter, then the points chosen in the two trials won't coincide exactly—they'll differ a little bit from each other. And, as we know, the smaller the difference, the longer the fates of the points will coincide. However, a time will necessarily come when the fates will part.

Imagine that the first experiment has been performed and the infinite fate of the first point has been written down. Also, imagine another experiment conducted at the same time: someone tosses a coin and writes a zero if it's heads and a one if it's tails. This results in another infinite sequence of zeros and ones. Then the two sequences are merged: we write out the first sequence from $-\infty$ up to a certain moment in the future—say, $n = 1{,}000$—and from this moment on we write the second sequence. By the theorem about the realization of an arbitrary fate, there is a point whose fate coincides with the new sequence. This point will be at a distance no greater than 5^{-1000} from the first one (their y-coordinates are the same, because they have the same past fate and their x-coordinates have the same first 1,000 quinary digits). So from the point of view of any experimenter, the two points will be indistinguishable. Nevertheless, their fates, starting from a certain moment, are different, and the difference is random in nature.

This effect also manifests itself in more complicated processes described by differential equations. It's called *experimental irreproducibility.* We can perform the same experiment, reproducing its initial conditions as accurately as possible, and after a lapse of time observe completely dissimilar results.

During the last decade, research into the chaotic behavior of deterministic systems has been a matter

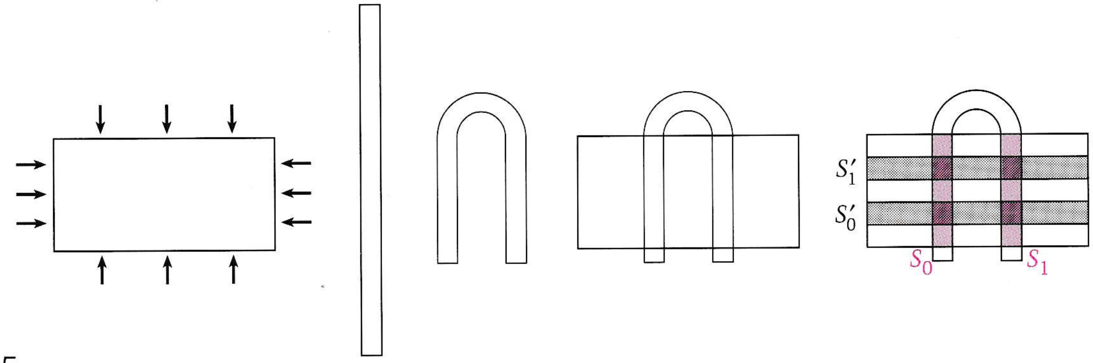


Figure 5

of the most lively interest. In many real systems described by differential equations, determinism ends up being a purely theoretical property. In practice, inevitable deviations lead to chaotically different results in seemingly identical experiments.

But let's return to our remarkable map. We haven't yet explained what the horseshoe is doing there in its name.

Smale's horseshoe

The original mapping considered by Stephen Smale was constructed as follows. Take a rectangle (fig. 5), squeeze it horizontally and stretch it vertically so as to make a tall and narrow vertical rectangle out of the low and wide initial one, then bend it into a horseshoe and superimpose over the original figure as shown.

This composition of two maps is just what was originally called the horseshoe map. At first glance it has almost nothing in common with the map we considered above. However, if we restrict the domain of the new map, we'll easily see the similarity.

Consider the intersection of the domain and range of the horseshoe map: it consists of the two red rectangles S_0 and S_1 in the figure. Suppose the inverse map is linear on S_0 and S_1 (that is, it reduces to uniform dilations parallel to the sides of these rectangles—Smale included this condition in his construction). Then the complete inverse image S_0' of S_0 is a long rectangle near the lower base of the original rectangle; similarly, from S_1 we obtain a long rectangle S_1' near the top base of the original rectangle. The map takes S_0' into S_0 and S_1' into S_1. And this reminds us of our map f above, though it isn't the same.

A useful problem is to formulate and prove an analogue of our main theorem for this (piecewise) linear map. ◘

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M141

Incompatibility with primes. A natural number is put through the following operation: its last digit is split off and multiplied by 4, then the product is added to the remaining number. (For instance, 1995 is thus transformed into 219.) The result is again subjected to the same operation, and so on. Prove that if the sequence thus obtained contains 1001, then it doesn't contain any prime numbers. (B. Ginzburg)

M142

Factored by substitution. Prove that for any polynomial $P(x)$ of degree greater than one, there exists a polynomial $Q(x)$ such that $P(Q(x))$ can be factored into two factors (all the polynomials have integer coefficients). (A. Kanel)

M143

Equal sections. (a) Three lines are drawn through a point in a triangle parallel to its sides. The segments intercepted on these lines by the triangle turn out to have the same length (see figure 1, in which the three equal segments are colored red). Given the triangle's side lengths a, b, and c, find the length of the segments. (b) Four planes are drawn through a point in a tetrahedron parallel to its faces. The sections of the tetrahedron created by these planes turn out to have the same area. Given the areas a, b, c, and d of the faces, find the area of the sections. (A. Yagubiants, V. Dubrovsky)

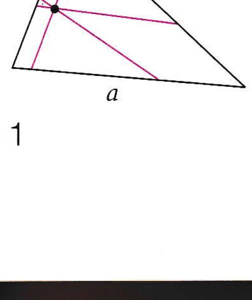


Figure 1

M144

Eccentricity of rectangles. For any rectangle, let's call the ratio of its larger side to its smaller the *eccentricity* of the rectangle. Show that if rectangle B is inscribed in rectangle A (so that each vertex of B is on a different side of A), then the eccentricity of B is not less than the eccentricity of A. (N. Vasilyev)

M145

Patriotic squares. Each square in an $n \times n$ array of squares is either red, white, or blue. Each red square borders a white square and a blue square (so that they share a common edge). Each white square borders a red square and a blue square, and each blue square borders a white square and a red square. If k is the number of squares of a given color, show that (a) $k \leq 2n^2/3$, (b) $k \geq n^2/11$. (F. Nazarov)

Physics

P141

Soft landing. A vessel contains two liquids that do not mix. Their corresponding densities are ρ_1 and ρ_2, and they form two layers in the vessel of thickness h_1 and h_2, respectively. A small streamlined body is lowered into the vessel. The body reaches the bottom at the very moment its velocity is zero. What is the body's density? Assume that the fluids are nonviscous. (M. Balashov)

P142

Bermuda Triangle fantasy. Using a special radar altimeter, astronauts in the Skylab space station found that the surface of the water in the Bermuda Triangle is 25 m lower than the normal ocean level. Assuming that this sag is caused by the existence of a spherical cavity filled with water lying just under the ocean floor, estimate the radius of this cavity. The depth of the ocean there is $h = 6$ km, and the average density of the bedrock $\rho_r = 3 \cdot 10^3$ kg/m^3. (A. Stasenko)

P143

Name that gas. A tank contains a pure gas, but it's not known which gas. An expenditure of 958.4 J is needed to increase the temperature of 1 kg of this gas by one degree Celsius at constant pressure, while an expenditure of only 704.6 J is needed to do this at constant volume. What gas does the tank contain? (K. Sergeyev)

P144

Ring in a B-field. A ring of diameter $d = 6$ mm made of very thin wire with a resistivity $\rho = 2 \cdot 10^{-8}\ \Omega \cdot$ m and a density $D = 9 \cdot 10^3$ kg/m^3 travels along the perpendicular bisector between the poles of a magnet and is not able to turn. Estimate the change in the ring's velocity if it was $v_0 = 20$ m/s before the ring entered the magnetic field. The magnetic field is perpendicular to the plane of the ring, and the velocity vector is parallel to this plane. The

CONTINUED ON PAGE 46

Airplanes in ozone

"The ozone layer is situated at an altitude of 20–50 kilometers. . . . In the absence of this ozone 'screen' protecting us from ultraviolet radiation, which is harmful in large doses, life on Earth in its present form would be impossible."—From the entry "Atmosphere" in The Young Physicist's Encyclopedic Dictionary

by Albert Stasenko

FISH IN ASPIC, LINGUINI IN clam sauce, "pigs in a blanket" —you've probably heard of them, even if you haven't actually eaten them. But "airplanes in ozone"? What on earth does *that* mean?

Let's back up a bit. As you probably know, the atmosphere in which we go to school, converse, and fly, the air we breathe, is basically composed of nitrogen and oxygen. But it also contains the so-called minor gases, whose role is actually far from minor.

One of the most important of the minor gases is ozone. Its chemical symbol is O_3, in contrast to ordinary oxygen (O_2). Perhaps your nose has been tickled by this gas after a thunderstorm, or in a coniferous forest, or in a house where an ozone generator is operating. ("Ozone" means "odorous" in Greek.) From the human point of view, the most important role ozone plays is blocking harmful ultraviolet radiation from reaching the Earth's surface. As you may recall, visible light is confined to the range of wavelengths 0.35–0.7 μm, and that the energy of a quantum of radiation (photon) is proportional to its frequency and inversely proportional to its wavelength.

As it turns out, ozone itself is produced primarily by ultraviolet radiation, but of shorter wavelengths (less than 0.2 μm). The energy of these UV photons is high enough to split the relatively stable oxygen molecule in two. Then these loose atoms attach themselves to two other oxygen molecules, producing two molecules of O_3. These new molecules (ozone) are less stable than molecular oxygen (they have a lower bonding energy), so they can be broken by photons with a somewhat longer wavelength, though still in the ultraviolet range of the spectrum (0.22–0.29 μm). Thus, acting in concert, oxygen and ozone absorb almost all the UV radiation coming from the Sun and let only a tiny fraction of it pass—which is nonetheless quite enough for a bad case of sunburn if you're not careful.

Although the spectral range where ozone "works" is narrow, the corresponding amount of absorbed energy is three times that of the rest of the spectrum. And if it weren't for ozone, the UV radiation would have nothing to contend with on its trip to the Earth's surface. It's clear now why we owe such a debt of gratitude to the ozone in our atmosphere—it's our main bulwark against the UV radiation that's so harmful for life forms on our planet.

Of course, I've simplified the mechanism of ozone formation. Different processes take place at different altitudes, and many substances participate in the synthesis (and decay) of ozone. Figure 1 shows the system of other atmospheric reactions that also produce ozone.

Recently scientists, and then reporters and all of forward-thinking humanity, were alarmed by the "ozone hole" over Antarctica. As a result, international agreements

Art by Yury Vashchenko

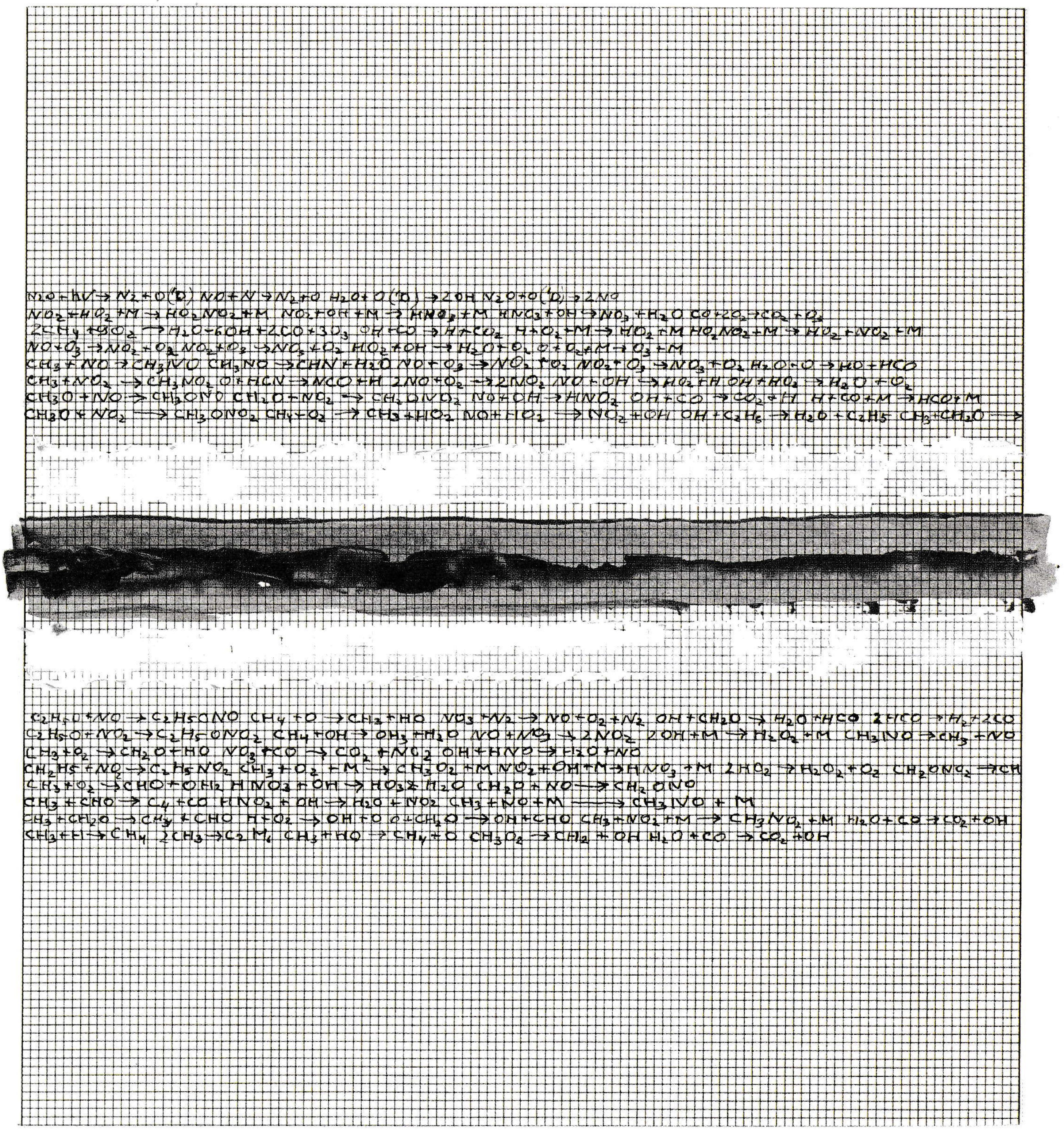


Figure 1

were concluded that prohibit the production of agents containing chlorine that are responsible for destroying ozone—for instance, the freon in refrigerators and the propellant in aerosol cans. However, the system of reactions involving such substances that "eat" ozone isn't shown in figure 1. This figure is shown not to intimidate but to illustrate what will be discussed below. To reassure the hesitant reader, I'll just say that this isn't the entire set of reactions where the drama of the birth and death of ozone is played out. We'll use only what is most necessary for our discussion.

First of all, the total density of the atmosphere (a mixture of many gases) decreases monotonically with altitude. If the atmospheric

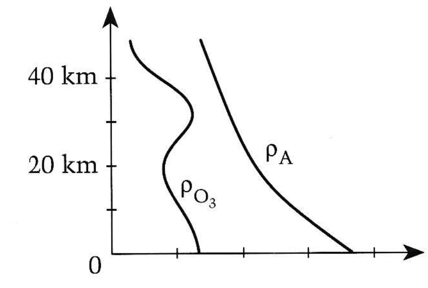


Figure 2

temperature were constant and equal to T, the pressure p and density ρ of a gas with a molecular mass M would vary with altitude y according to the Boltzmann barometric equation

$$\frac{\rho}{\rho_0} = \frac{p}{p_0} = e^{-Mgy/RT}. \tag{1}$$

The subscript 0 corresponds to pressure at sea level.

However, the ozone's density does not vary monotonically and assumes a maximum value at an altitude of about 20–30 km (fig. 2), depending on the geographical location, season, and time of day. Of course, the fact that the variation is not monotonic has almost no effect on the dependence of the total density of the mixture of gases constituting the atmosphere on altitude, since the ozone's density is less than the total density by many orders of magnitude. However, it is just at these altitudes that airliners will be flying in the near future. At first glance, there's no problem: Fly in, fly out—bye-bye! It reminds me of an old Russian riddle about a boat slicing through the water: "I cut, I cut, but there's no pain; I go along, but leave no trace." But this saying is totally inappropriate for an airliner from the ecological point of view.

The middle of figure 1 schematically shows the exhaust of an aircraft engine as well as the chemical reactions going on in and around it. Again, they are not shown to intimidate you: even those who solve this system with computers don't know everything about the coefficients in these reactions. So we can just contemplate the big picture. The point is, the airliners will deliver this heap of chemicals into the very heart of the ozone layer—that is, at altitudes where the ozone concentration is the greatest. Mind you, it will be the most peaceful civilian aviation. It will be used by business travelers with briefcases full of agreements and contracts, tourists and relatives flying to other continents, and exchange students traveling to other countries. So "disarmament" treaties are out of the question here—this is the aviation that we all use.

For commercial aviation to be profitable, the airline companies need not a few jet planes but hundreds of them. In sum, these planes will burn 50–100 million tons of fuel in the atmosphere, producing among other things about one million tons of the nitrogen oxides NO, NO_2, … . For convenience, we'll denote this set of oxides as NO_x. Surprisingly, even nitrogen "burns" (oxidizes) in the high-temperature regions of jet engines. Some of the substances expelled by a jet engine are harmful to the ozone, and first among them are the nitrogen oxides themselves (fig. 1). These reactions are directly responsible for ozone depletion. However, this figure also shows that there are ozone-producing reactions—both in the jet engine itself and in the atmosphere. Since the production of extra amounts of nitrogen oxides may shift the chemical equilibrium and decrease the amount of ozone in the air, it's very tempting to decrease the concentration of NO_x by any means possible. But how?

First of all we need to take a close look at the trail left by a jet airplane.

A solitary round stream

Why does a jet plane need a "jet"—the high-speed stream of air that its engines produce? Clearly, it's the jet that contains the backward momentum necessary to obtain the plane's reactive (forward) force, or "thrust." A jet airplane "swallows" the air that comes in, consumes some of it (oxygen) to burn fuel (kerosene, as a rule), and then expels the heated mixture backward, producing the thrust necessary for flight. When kerosene or any other hydrocarbon is burned in the presence of nitrogen, a gaseous mixture is produced, as shown in figure 1. If hydrogen were burned instead of kerosene, no carbon radicals would be present in the jet. But nitrogen oxides will nevertheless appear in even greater amounts.

Now let's consider the mechanics of a jet. We denote the velocity of the jet relative to the airplane as u, the exhaust velocity of the jet from the engine as u_A (the subscript A comes from the German word *Ausgang*, which means "exit"), and the jet's velocity at a great distance from the plane as u_∞ (it's equal to the velocity of the air relative to the airplane, or simply to the velocity of the airplane itself). Bursting into the atmosphere with a velocity of $|u_A - u_\infty|$,

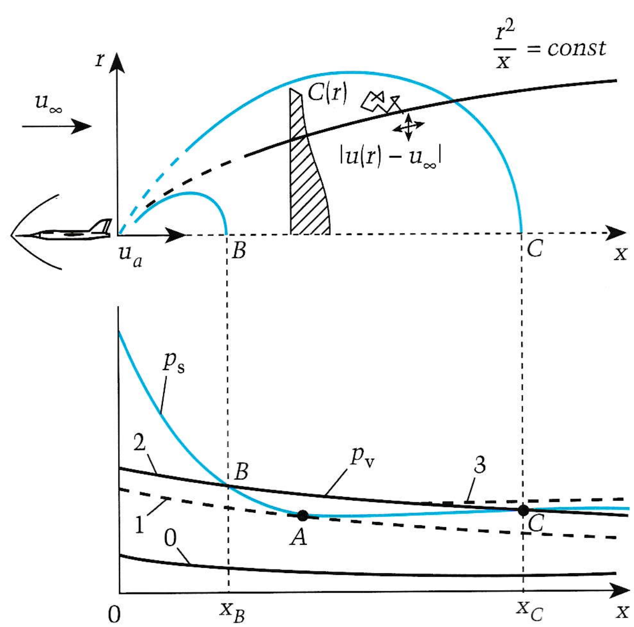


Figure 3

the jet begins to move relative to the air with a small and monotonically decreasing velocity $|u - u_\infty| \ll u_A, u_\infty$. While this is happening, the gases in the jet mix continuously with ever new regions of the atmosphere. In this new mixture of gases, the concentration of the initial gases will smoothly decrease from the jet's axis to the periphery (see the curve $C(r)$ in the top part of figure 3).

Similarly, the velocity $u(r) - u_\infty$ will decrease in the radial direction, so that layers of the stream moving at different distances from the axis will "rub" on one another. The distant layers will slow the motion of the nearest ones. A reader well versed in physics will have restated it already: "The axial component of momentum is being transferred radially." Such a reader will certainly have introduced the corresponding coefficient of dynamic viscosity (internal friction), and that of diffusion to describe the mass transfer, and finally the coefficient of thermal conductivity for the transfer of heat. Good work! But maybe we're not so far advanced. Let's try to describe these phenomena more simply.

The transfer of all the values mentioned and also the motion of the jet's particles radially are described in terms of stochastic (random) motion. From the theoretical point of view, this kind of motion is explained in detail in, for instance, the *Textbook on Physics* by Richard P. Feynman, where you'll find the famous example of the drunken sailor who arbitrarily chooses one of four directions at each intersection in an unfamiliar town (backward or forward, to the right or to the left). Let's get to the crux of the matter. A particle that comes to a certain point (fig. 3) will interact with another particle and then move in any direction with equal probability. However, after traveling a characteristic distance l, it collides with the next particle and again changes direction stochastically.

It can be shown that, on average, each step increases the square of the distance by l^2. Consider a particle that has made N steps and whose position is described by the radius-vector $\mathbf{r}_N$. The particle came to this point from some other point $\mathbf{r}_{N-1} = \mathbf{r}_N - \mathbf{l}$. This equation tells us that the vectors $\mathbf{r}_{N-1}$ and $\mathbf{l}$ are not necessarily parallel. Let's raise this equation to the second power:

$$r_N^2 = (\mathbf{r}_{N-1} + \mathbf{l})^2$$
$$= r_{N-1}^2 + 2\mathbf{r}_{N-1} \cdot \mathbf{l}.$$

The motion of the chosen particle is accurately described, because this is its own history, represented by some zigzag line. But since we are interested in the mean value of the coordinates of a vast number of particles, we sum these squares of the displacements and divide the sum by the number of particles. In this way we obtain the mean value. In doing so we observe that in the last item the displacements $\mathbf{l}$ are directed along and counter to vector $\mathbf{r}_{N-1}$ with equal probability. Thus, averaging over a large number of particles yields

$$\langle r_N^2 \rangle = \langle r_{N-1}^2 \rangle + l^2 + 0.$$

Then, beginning with $N = 1$, we obtain the next equation by mathematical induction:

$$\langle r_N^2 \rangle = Nl^2$$

—that is, the average square of the particle's displacement during random motion is proportional to N (hence, the mean square displacement $\sqrt{\langle r_N^2 \rangle}$ is proportional to $\sqrt{N}$ and not to N, as would be true for uniform motion).

If the particle's velocity between collisions is v, then the time necessary for a step is $\tau = l/v$, so in a time t the particle makes $N = t/\tau$ steps. Finally, denoting r_N simply as r, we get

$$\langle r^2 \rangle = tlv.$$

The magnitude D, which is proportional to lv, is called the diffusion coefficient. The exact result for the case under consideration looks similar:

$$\langle r^2 \rangle = 4Dt. \qquad (2)$$

Let's denote the concentration of certain particles in a stream as n—they can be particles of soot, for instance. Multiplying this concentration by their travel speed, we get the density of these particles in the airplane's frame of reference: nu_∞ (it's assumed here that the jet's velocity relative to the atmosphere is negligibly small compared to u_∞). Now multiply both sides of equation (2) by the flux density and by the number π:

$$\pi\langle r^2 \rangle nu_\infty = 4\pi Du_\infty tn.$$

The left-hand side of this equation is the total flux of all the particles through a circle of area $\pi\langle r^2 \rangle$. If the particles are neither attaching themselves to one another nor splitting themselves up—that is, when the total number of particles doesn't change—the left-hand side is constant, too. On the right-hand side we see the combination $u_\infty t = x$—that is, the distance from the airplane. Consequently,

$$n \sim \frac{1}{x}. \qquad (3)$$

Thus, we almost know the "mechanism" for a jet in the dynamical frame of reference attached to the airplane. The jet's particles diffuse radially as they are carried backward from the plane with an almost constant velocity $u_\infty = x/t$. So on average they move along the parabolas $x \sim r^2$. At the axis this results in a hyperbolic decrease in the concentration (equation (3)).

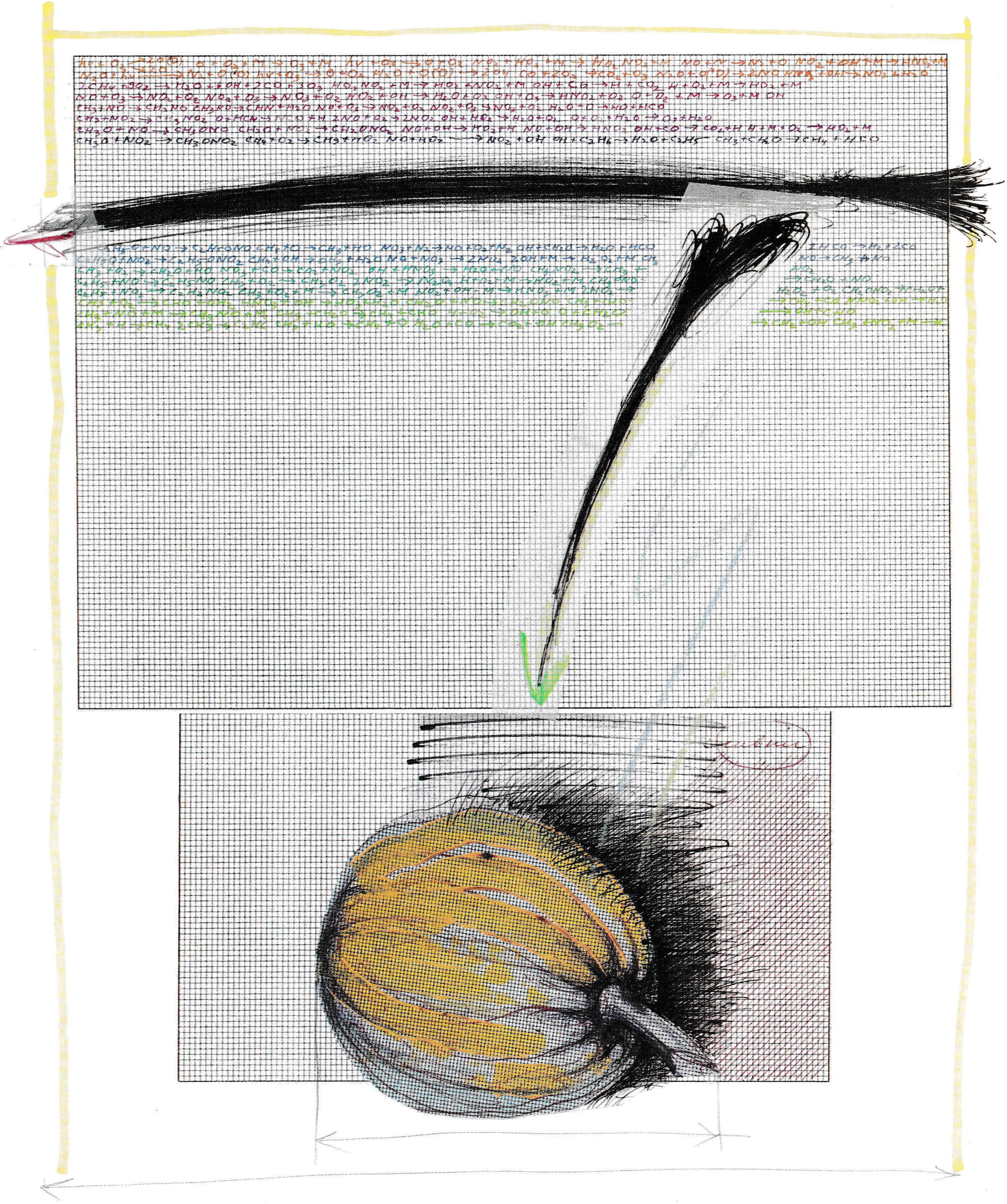
Readers who are further along in physics will quickly grasp how the concentration varies with both coordinates:

$$n(x, r) \sim \frac{1}{x/u_\infty} e^{-r^2/(4Dx/u_\infty)}. \qquad (4)$$

(Recall that x/u_∞ is the time t.)

Both the momentum and the heat content will change in the same way:

$$\frac{n(x, r)}{n_A} = C(x, r) = \frac{T(x, r) - T_\infty}{T_A - T_\infty}$$
$$\cong \frac{u(x, r) - u_\infty}{u_A - u_\infty}. \qquad (4')$$



Equations (4) and (4′) will not be of much use, however, in what follows. All they can do is give the equation for the lines where all the dimensionless parameters in equation (4′) take constant values—for example, 10^{-1}, 10^{-2}, 10^{-3}, and so on; that is, the lines where the stream is "diluted" by atmospheric air down to a concentration of one tenth, one hundredth, one thousandth, and so on, compared to the initial value. These lines have the characteristic shape shown in the top part of figure 3.

Drops behind

Now let's find the conditions leading to the condensation of water vapor in the stream. First of all, the stream must become rather cool. However, this is not enough to begin condensation. A certain amount of water is necessary to form a saturated vapor—only then will the "dew" appear. Strictly speaking, the pressure of the water vapor at some point in the stream p_v must not be less than the saturated pressure p_s. Previously we saw how p_v changed with the location. What about p_s?

Now it's time to recall equation (1). In the argument of the exponential function there is a relation between two energies: the potential energy of a molecule (or one mole M) at altitude y above sea level and its characteristic thermal energy kT at a temperature T (or RT for one mole): $mgy/kT = Mgy/RT$. It turns out that this formula is a special case of Boltzmann's more generalized assertion: if a system composed of a vast number of identical molecules is in thermodynamic equilibrium, and if these molecules can be characterized by certain energy levels (say, E_1 and E_2), the relation between the numbers of molecules at these levels obeys the formula

$$\frac{n_2}{n_1} = \frac{e^{-E_2/kT}}{e^{-E_1/kT}} = e^{-\frac{E_2-E_1}{kT}}.$$

But what is evaporation if not a process of "fishing out" a molecule from a liquid into a gas? Perhaps you've encountered the concept of the heat of vaporization L—that is, the energy needed to extract one kilogram of vapor from the liquid phase (the corresponding value for one mole is ML. This energy can be taken as a measure of the depth of the potential energy well where the molecules of liquid are located and from which they must be taken to form gas. So, according to Boltzmann's principle and by analogy with equation (1), we can write the relation of the densities of saturated vapor and liquid as $\rho_s/\rho^0 \sim e^{-L/RT}$, or for pressure values as

$$p_s = p^0 e^{-L/RT}. \qquad (5)$$

Note that this dependence on temperature (the exponential one!) is much steeper than the hyperbolic decrease ($\sim 1/x$) of the jet's parameters along its axis (equations (3) and (4)).

Now let's sketch the change in the water vapor pressure p_v and saturated vapor pressure p_s along the jet's axis. There are several possibilities, as shown in the bottom part of figure 3. For curve 0, the amount of vapor at every point in the stream is less than the amount necessary for saturation. In curve 1, the condensation condition is met only at a single point A and the drops evaporate before they begin to grow. Curve 2 crosses the saturation curve at two points, B and C; the vapor pressure between these points is greater than that necessary for saturation, which means that in this region the formation of drops is possible. However, after starting to grow at point B, the drops may disappear after point C if the water vapor pressure in the stream becomes less than the saturation pressure, because they diffuse away from the jet into the "dry" atmosphere. Finally, in curve 3 there is the necessary amount of vapor in the atmosphere itself and this vapor is close to saturation—$p_{v\infty} \cong p_s T_\infty$—but it does not condense, perhaps because there are no foreign particles to serve as "seeds," perhaps for some other reason. But the jet just happens to contain such particles, and the vapor, which begins to condense on them at point B, will not evaporate now, so that the drops will not disappear, but may even grow at the expense of atmospheric water vapor if $p_{v\infty} > p_s T_\infty$.

Let's consider the case when all the water vapor produced in the aircraft engine has condensed as drops. It's known that a great number of soot particles (carbon resulting from incomplete combustion) are expelled from the exhaust nozzle into the jet. Measurements show that their density at the nozzle's opening varies from 10^{13} to 10^{17} m^{-3}. When these foreign particles are present, vapor condenses specifically on them and doesn't wait until a significant supersaturation (and hence supercooling) has been achieved. Indeed, this is the very reason why people build fires to combat frost in orchards and gardens: the vapor condenses on the smoke particles and the phase heat dissipates into the air, thus preventing the frost from forming on the plants and damaging the crops.

So, each soot particle acquires "its own portion" of water vapor. From this one can obtain the characteristic size of the water droplets:

$$\frac{4}{3}\pi\rho^0 a^3 n_A = \rho_v^a,$$

$$a = \sqrt[3]{\frac{3\rho_v^a}{4\pi n_A \rho^0}}. \qquad (6)$$

Here we neglect the volume of the soot particle itself, considering this center of condensation to be very small.

Now let's make some estimates, for, as the old Russian saying has it, a theory without estimates is like soup without salt. Assume that the atmospheric pressure and density are an order of magnitude less than the corresponding values at sea level (you can use equation (1) if you like). Let the temperature at the nozzle be three times that in the surrounding atmosphere; hence, the jet's density is one third that in the surrounding air: $\rho_A = \rho_\infty/3 = \rho_0/30$ (here we assumed that the pressure in the jet exactly equals the atmospheric pressure, in

CONTINUED ON PAGE 48


HOT
COLD
S. Ivanov

Jesse James discovers the heat equation

Viewing the redistribution of wealth as a diffusion process

by Kurt Kreith

LEGEND HAS IT THAT JESSE James was a Robin Hood of the Wild West. According to one ballad,

Jesse James was a lad, who
killed many a man,
He robbed the Glendale train;
He stole from the rich and he
gave to the poor,
He'd a hand and a heart and
a brain.

Had Jesse used his brain to study physics instead of train schedules, he might have considered alternate means for implementing the social goals attributed to him. In particular, he might have noted that the spread of dye in a petri dish, the dispersion of fumes in the atmosphere, and the flow of heat in a rod all represent "diffusion processes"—ones that bear an interesting relationship to Jesse's economic agenda.

Monetary diffusion

To see how such physical concepts might relate to Jesse's situation, let's consider eight individuals of varying wealth arranged in a row (other arrangements will be considered later). I'll label these individuals with integers 1, 2, ..., 8, and denote their assets by $u(1)$, $u(2)$, ..., $u(8)$, respectively. This situation is represented in the chart shown in figure 1.

individual	1	2	3	4	5	6	7	8
$u(x)$	\$160	\$30	\$80	\$110	\$200	\$80	\$60	\$40

Figure 1

One way of "sharing the wealth" is to begin by recording the differences in assets between each individual listed and that person's two immediate neighbors. For example, individual 3, with assets of \$80, has \$50 more than one neighbor and \$30 less than the other; individual 5 (who may not like the next set of instructions!) has \$90 more than one neighbor and \$120 more than the other.

Having recorded such pairs of differences for each individual in the row, Jesse might (with some friendly persuasion) impose the following rule:

"At my command, let there be a transfer of funds, from richer neighbor to poorer, equal to 10% of the differences just recorded."

For example:

Individual 3 is to receive \$3 from individual 4, but at the same time must pay \$5 to individual 2, leading to a net loss of \$2;

Individual 5 must pay \$9 to individual 4 and \$12 to individual 6, leading to a net loss of \$21;

Individual 8, having only one neighbor, receives \$2 from individual 7.

When all these transfers are completed, there exists a new social order represented by the list of assets in figure 2.

At first glance, this method of sharing the wealth leaves much to be desired. Individual 4, who started out with assets of \$110, grew wealthier, while individual 3, who started out with \$80, lost money.

individual	1	2	3	4	5	6	7	8
$u(x)$	\$147	\$48	\$78	\$116	\$179	\$90	\$60	\$42

Figure 2

Art by Sergey Ivanov

individual	1	2	3	4	5	6	7	8
wealth	\$137.10		\$78.80	\$118.50				

Figure 3

Individual 7, who started out with only \$60, saw no change. Finally, we should note the need of "special rules" for the individuals at the ends of the row. Having only one neighbor, they engaged in only one transfer of funds.

The virtue of this particular rule becomes clear only if we are willing to repeat it a number of times. While it's tedious to do many iterations, doing one more may help fix the ideas in your mind. In figure 3, I've started to enter the results of applying Jesse's rule to the chart in figure 2. (I'll leave it to you to complete.)

Jesse's rules restated

In anticipation of having to repeat this process many times, it may help to express the underlying rules in functional notation:

$u(1)$ becomes $u(1) + 0.1[u(2) - u(1)]$;

$u(x)$ becomes $u(x) + 0.1[u(x-1) - u(x)] + 0.1[u(x+1) - u(x)]$ or $u(x) + 0.1[u(x-1) - 2u(x) + u(x+1)]$ for $2 \le x \le 7$;

$u(8)$ becomes $u(8) + 0.1[u(7) - u(8)]$.

Making use of spreadsheet software (see "Look, Ma—No Calculus!" in the November/December 1994 issue of Quantum or the sidebar on page 31), we can program these rules to yield the results in figure 4 (showing six iterations). In fact, with modern software like Microsoft® Excel, we can easily repeat these same rules 40 times and then obtain a graphical representation of the result (fig. 5).

The process represented above is, in the physical world, called diffusion. It occurs in a variety of situations, such as the diffusion of toxic substances in a lake, diffusion of heat in a rod, and the diffusion of smoke in the air. A more theoretical example concerns the location of a small particle subject to Brownian motion. While we may know exactly where such a particle starts out, with time our knowledge becomes one of probabilities. This "diffusion of certainty" is another example of the kind of process under consideration.

To make our discussion more specific, we'll focus on the diffusion of heat in a rod. However, the possibility of other interpretations should also be kept in mind.

The heat equation

Underlying Jesse's scheme for diffusion of wealth was a repeated application of the rule

$u(x)$ becomes $u(x) + 0.1[u(x-1) - 2u(x) + u(x+1)]$ for $x = 2, 3, \ldots, 7$

(the modifications at the ends of the rows will be discussed later). If we measure time in minutes (or any other unit of time) and represent the wealth of individual x at time t by $u(x, t)$, this rule can also be written

$$u(x, t+1) = u(x, t) + 0.1[u(x-1, t) - 2u(x, t) + u(x+1, t)]$$

or

$$u(x, t+1) - u(x, t) = 0.1[u(x-1, t) - 2u(x, t) + u(x+1, t)].$$

Those who have studied calculus may recognize this latest formulation of Jesse's rule as the finite-difference form of the "heat equation"

$$u_t = 0.1u_{xx}.$$

That is, $u(x, t+1) - u(x, t)$ corresponds to "the change of $u(x)$ in a unit of time," or u_t. Similarly,

$$u(x-1, t) - 2u(x, t) + u(x+1, t) = [u(x+1, t) - u(x, t)] - [u(x, t) - u(x-1)]$$

is called "the second difference of u in x" and corresponds to u_{xx}. However, calculus won't be needed to follow this discussion. Rather, I'll show how spreadsheets can be used to represent discrete versions of the heat equation and to obtain excellent approximations of its solution.

Jesse's rules were based on a transfer of money between individuals and their immediate neighbors. In our physical interpretation, such a flow of money corresponds to "the flux of heat" (formally defined as the flow of energy across a unit of cross-sectional area). We're all familiar with the fact that a cup of hot tea will cool faster in a refrigerator than at room temperature—that is, that the rate of heat flow depends on the difference in temperature between the tea and its environment. Our particular rule—making the flow of money proportional to the differences in wealth between individuals and their immediate neighbors—corresponds to Fourier's law: the flux of heat at any point is proportional to the *temperature gradient* at that point.

The fact that money flows from richer to poorer corresponds to the assumption that heat flow is proportional to the temperature gradient multiplied by a *negative* constant. In the notation of partial differential equations, this corresponds to

$$\text{flux} = -cu_x.$$

Jesse's rule was specific in calling for a transfer of 10% of the difference in wealth between

Iteration	1	2	3	4	5	6	7	8
0	160.00	30.00	80.00	110.00	200.00	80.00	60.00	40.00
1	147.00	48.00	78.00	116.00	179.00	90.00	60.00	42.00
2	137.10	60.90	78.80	118.50	163.80	95.90	61.20	43.80
3	129.48	70.31	80.98	119.06	152.48	99.22	62.93	45.54
4	123.56	77.29	83.72	118.59	143.81	100.92	64.82	47.28
5	118.94	82.56	86.57	117.63	137.00	101.60	66.68	49.03
6	115.30	86.60	89.27	116.46	131.52	101.65	68.40	50.80

Figure 4

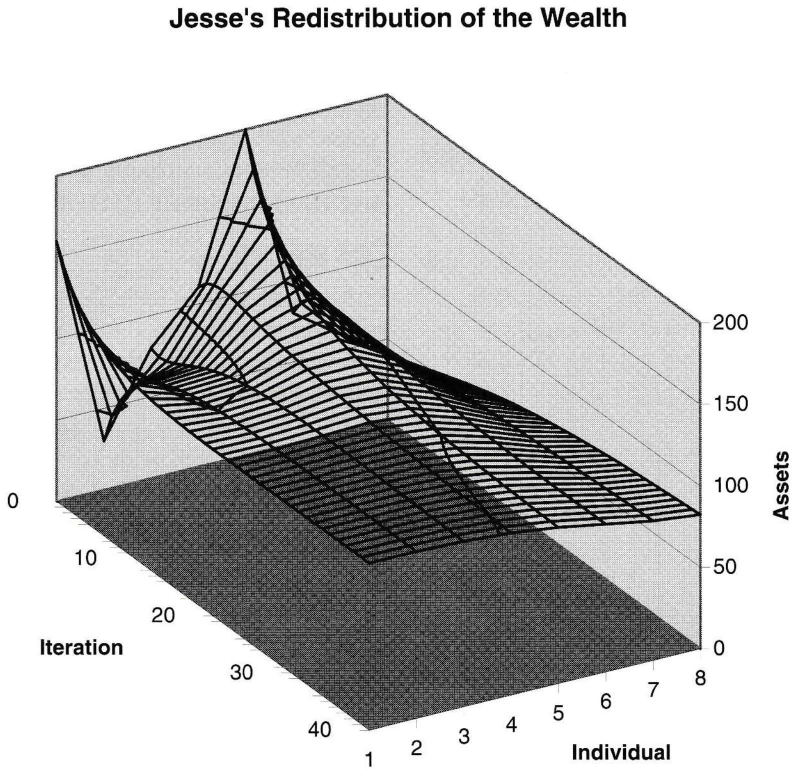


Figure 5

adjoining individuals in each unit of time. This corresponds to setting $c = 0.1$ in the above equation. In terms of heat flow, c represents the *conductivity* of the material through which heat is flowing. As anyone who has picked up a hot frying pan knows, c is quite large for metals and considerably smaller for pot-holders.

This interpretation raises the question of whether Jesse could have expedited the distribution of wealth by increasing c from 0.1 to 0.6, or even larger. It's at this point that some differences between our discrete model and the corresponding differential equation appear—ones that the reader can discover by experimenting with the spreadsheet model discussed above. It turns out that making c too large prevents an orderly "sharing of wealth." Indeed, this fact is closely related to *chaos theory*, but that's another story. For now, let's just note that if Jesse is in a hurry, he should consider increasing the *frequency* at which money is transferred rather than making c inordinately large. For example, rather than transferring 60% of the wealth every minute, he could insist on a transfer of 1% of the wealth every second. It's this sort of issue that distinguishes the difference equations we're considering here from the partial differential equations traditionally used to model the diffusion of heat.

The ends of the row

Interesting issues of a different kind arise at the ends of the row. Here individuals 1 and 8 were subject to the special rules

$$u(1, t+1) = u(1, t) + 0.1[u(2, t) - u(1, t)]$$

and

$$u(8, t+1) = u(8, t) + 0.1[u(7, t) - u(8, t),$$

respectively. In the case of heat flow in a rod, such special rules correspond to a particular *boundary condition* at the ends of the rod.

The fact that there is no transfer of money at the left of individual 1 or to the right of individual 8 makes Jesse's rule correspond to a rod that is *insulated* at both ends. One consequence of this boundary condition is that the total amount of money owned by the individuals in the row remains constant.

In order to include other boundary conditions in our monetary version of the heat equation, it will be convenient to introduce two fictional characters at the ends of our row—individuals numbered 0 and 9. Individual 0 is located at the left of individual 1, while individual 9 is at the right of individual 8. Suppose now that the assets of individual 0 are, at all times, postulated to be identical to the assets of individual 1, and that individual 1 (who now has two neighbors) implements the same rule as individuals 2–7. That is,

$$u(0, t) = u(1, t) \qquad (1)$$

and

$$u(1, t+1) - u(1, t) = 0.1[u(2, t) - 2u(1, t) + u(0, t)]. \qquad (2)$$

This situation (readily represented in a spreadsheet) is in fact equivalent to the insulated boundary condition previously considered (why?). However, this boundary condition now corresponds to a special rule being imposed on the *fictional* individual 0 rather than requiring that we impose a special rule for the (less fictional) individual 1. Are you able to impose a special rule on the fictional individual 9 to achieve a corresponding situation at the right end of the row?

The advantage of including such fictional individuals is that they enable us to represent other kinds of boundary conditions. For example, suppose that we change the rule for individual 0 from

$$u(0, t) = u(1, t) \text{ for all } t$$

to

$$u(0, t) = 0 \text{ for all } t,$$

while individual 1 continues to abide by equation (2). Individual 0 now plays the role of a "tax collector" who, at each transfer of funds, takes 10% of individual 1's assets. These rules can also be represented on a spreadsheet.

In this situation the total amount of money in the row is *not* conserved. Rather, the tax collector drains off funds every time money changes hands. (One could also impose a tax

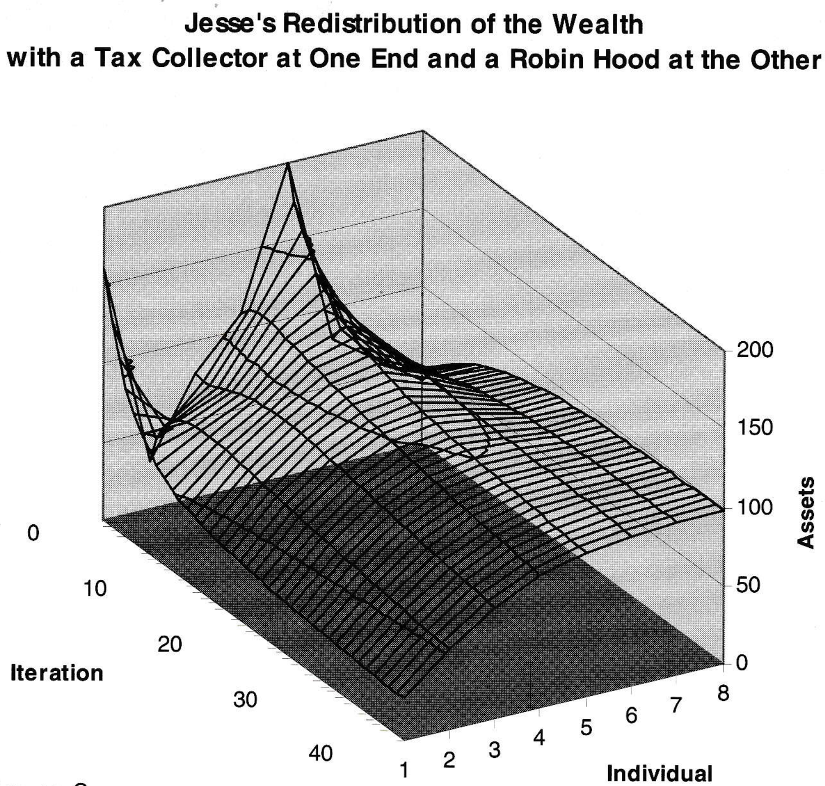


Figure 6

on money exchanged within the row. Can you think of a physical analogy for this in the case of heat flow?)

With a tax collector situated at the left end of the row, let's now consider putting a "Robin Hood" at the right end—that is, an individual who gives or takes money from individual 8 depending on the amount of money concentrated at the right end of the row. This can be accomplished by setting $u(9, t) = 100$ for all values of t, thereby subjecting individual 8 to the rule

$$u(8, t+1) - u(8, t) = 0.1[u(7, t) - 2u(8, t) + 100].$$

In physical terms, these rules correspond to a rod with conductivity 0.1 whose left end is in ice water and whose right end is in boiling water at 100°C. The graph in figure 6 shows the outcome of applying Jesse's rules 40 times with these boundary conditions.

Note that this form of diffusion doesn't seem to be leading us to Jesse's goal of economic equity. As in the real world, proximity to the tax collector is a decided disadvantage in this scheme of things!

Cups and caps

Now that we're able to relate monetary diffusion to a physical phenomenon—namely, heat flow in a rod—let's return to the "bottom line" in the original economic model: "Will I get richer or poorer?" Answering this question in the long run requires that one know the outcome of this process for large values of t, and this may be difficult to predict. However, the implications of a single transfer of funds for individual x at time t depends only on the value of the "second difference"

$$D^2u(x, t) = u(x-1, t) - 2u(x, t) + u(x+1, t).$$

If $D^2u(x, t)$ is positive, then individual x's assets will increase as $t \to t+1$; if $D^2u(x, t)$ is negative, they will decrease; if $D^2u(x, t) = 0$, they will remain unchanged.

Noting the importance of the quantity $D^2u(x, t)$ for our diffusion process, we'll refer to it as the "degree of cuppedness of $u(x, t)$ at time t." This curious terminology turns out to be appropriate for both social and geometric reasons. That is, a cup may

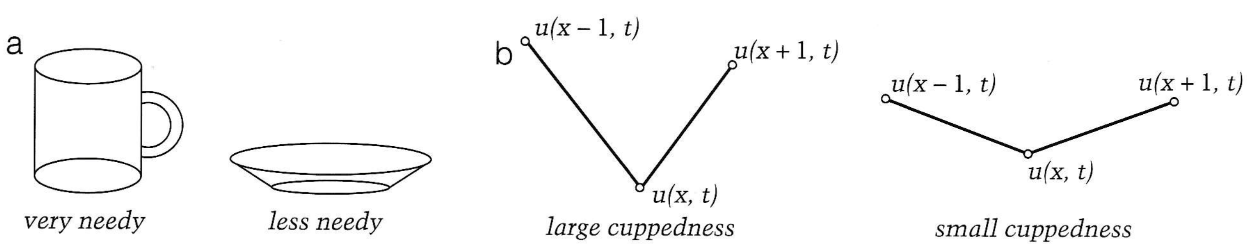


Figure 7

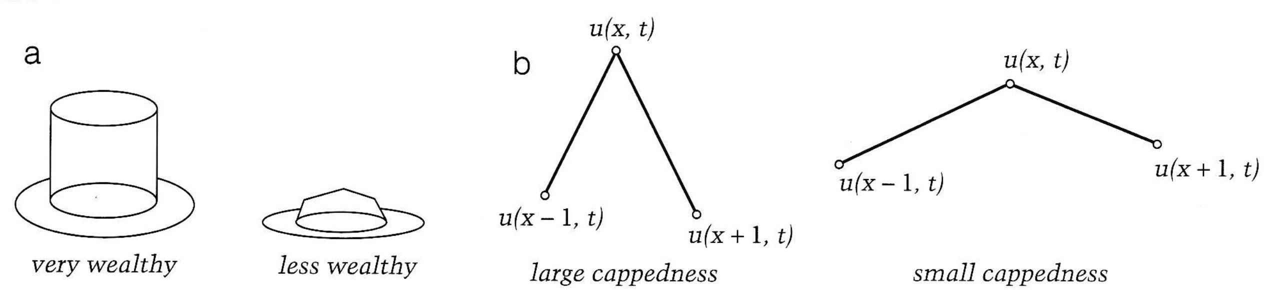


Figure 8

be considered a symbol of need, with its shape representing the degree of need (see figure 7a). At the same time, a graphical representation of $u(x, t)$ indicates that "positive cuppedness" corresponds to a distribution that is convex upward (that is, cup-shaped) at x (fig. 7b).

On the other hand, negative cuppedness (which will also be referred to as cappedness) may be regarded as a symbol of relative wealth, with the shape of the cap representing the degree of wealth (fig. 8a). Also, cappedness corresponds to a distribution that is convex downward (or cap-shaped) at x (fig. 8b).

Where are we heading?

Assuming that the diffusion process is approaching a steady state, we would expect it to have an "equilibrium solution"—that is, one that has zero cuppedness and zero cappedness at every value of x. As you may have guessed by now, the condition that

$$u(x-1) - 2u(x) + u(x+1) = 0 \quad \text{for } 1 \le x \le 8$$

requires that the graph of $u(x)$ consist of 10 points $(0, u(0))$, $(1, u(1))$, ..., $(9, u(9))$ all of which lie on a straight line in the (x, u)-plane. As in Euclidean geometry, this line will be determined by two endpoints $(0, u(0))$ and $(9, u(9))$. In the case of heat flow, this

CONTINUED ON PAGE 51

How to make a spreadsheet

(SEE ALSO "LOOK, MA—NO CALCULUS!" IN THE NOVEMBER/DECEMBER 1994 ISSUE)

IN ORDER TO APPLY JESSE'S RULE TO THE original eight individuals, we can use a nine-column spreadsheet. The first column will keep track of how many times we have applied Jesse's rule. We'll label the remaining columns 1, 2, ..., 8 across the top to keep track of the wealth of the eight individuals. Just below these numbers we enter each person's assets:

	A	B	C	D	E	F	G	H	I
1	Iteration	1	2	3	4	5	6	7	8
2	0	160.00	30.00	80.00	110.00	200.00	80.00	60.00	40.00
3									
4									
5									

The third row of the spreadsheet will now be used to enter the rules:

$$\begin{aligned} &u(1) \text{ becomes } u(1) + 0.1[u(2) - u(1)]; \\ &u(x) \text{ becomes } u(x) + 0.1[u(x-1) - 2u(x) + u(x+1)] \\ &\qquad \text{for } 2 \le x \le 7; \\ &u(8) \text{ becomes } u(8) + 0.1[u(7) - u(8)]. \end{aligned}$$

The fact that we are about to "program" a rule is signaled by entering an equal sign in front of the rule itself:

	A	B	C	D
1	Iteration	1	2	3
2	0	160	30	80
3	=A2+1	=B2+(C2-B2)/10	=C2+(B2-2*C2 + D2)/10	=D2+(C2-2*D2 + E2)/10
4				
5				

Continuing these rules down five more rows yields

	A	B	C	D
1	Iteration	1	2	3
2	0	160	30	80
3	=A2+1	=B2+(C2-B2)/10	=C2+(B2-2*C2 + D2)/10	=D2+(C2-2*D2 + E2)/10
4	=A3+1	=B3+(C3-B3)/10	=C3+(B3-2*C3 + D3)/10	=D3+(C3-2*D3 + E3)/10
5	=A4+1	=B4+(C4-B4)/10	=C4+(B4-2*C4 + D4)/10	=D4+(C4-2*D4 + E4)/10
6	=A5+1	=B5+(C5-B5)/10	=C5+(B5-2*C5 + D5)/10	=D5+(C5-2*D5 + E5)/10
7	=A6+1	=B6+(C6-B6)/10	=C6+(B6-2*C6 + D6)/10	=D6+(C6-2*D6 + E6)/10
8	=A7+1	=B7+(C7-B7)/10	=C7+(B7-2*C7 + D7)/10	=D7+(C7-2*D7 + E7)/10

whose values are

	A	B	C	D
1	Iteration	1	2	3
2	0	160.00	30.00	80.00
3	1	147.00	48.00	78.00
4	2	137.10	60.90	78.80
5	3	129.48	70.31	80.98
6	4	123.56	77.29	83.72
7	5	118.94	82.56	86.57
8	6	115.30	86.60	89.27

Now, in accordance with the discussion on page 29, let's introduce the fictional character at the left end of the row (in Excel there's a command "Insert" for doing this). The program for an insulated rod becomes

	A	B	C	D	E
1	Iteration	0	1	2	3
2	0	=C2	160	30	80
3	=A2+1	=C3	=C2+(B2-2*C2 + D2)/10	=D2+(C2-2*D2 + E2)/10	=E2+(D2-2*E2 + F2)/10
4	=A3+1	=C4	=C3+(B3-2*C3 + D3)/10	=D3+(C3-2*D3 + E3)/10	=E3+(D3-2*E3 + F3)/10
5	=A4+1	=C5	=C4+(B4-2*C4 + D4)/10	=D4+(C4-2*D4 + E4)/10	=E4+(D4-2*E4 + F4)/10
6	=A5+1	=C6	=C5+(B5-2*C5 + D5)/10	=D5+(C5-2*D5 + E5)/10	=E5+(D5-2*E5 + F5)/10
7	=A6+1	=C7	=C6+(B6-2*C6 + D6)/10	=D6+(C6-2*D6 + E6)/10	=E6+(D6-2*E6 + F6)/10
8	=A7+1	=C8	=C7+(B7-2*C7 + D7)/10	=D7+(C7-2*D7 + E7)/10	=E7+(D7-2*E7 + F7)/10

A spreadsheet that includes the rule for the tax collector looks like this:

	A	B	C	D	E
1	Iteration	0	1	2	3
2	0	0	160	30	80
3	=A2+1	=B2	=C2+(B2-2*C2 + D2)/10	=D2+(C2-2*D2 + E2)/10	=E2+(D2-2*E2 + F2)/10
4	=A3+1	=B3	=C3+(B3-2*C3 + D3)/10	=D3+(C3-2*D3 + E3)/10	=E3+(D3-2*E3 + F3)/10
5	=A4+1	=B4	=C4+(B4-2*C4 + D4)/10	=D4+(C4-2*D4 + E4)/10	=E4+(D4-2*E4 + F4)/10
6	=A5+1	=B5	=C5+(B5-2*C5 + D5)/10	=D5+(C5-2*D5 + E5)/10	=E5+(D5-2*E5 + F5)/10
7	=A6+1	=B6	=C6+(B6-2*C6 + D6)/10	=D6+(C6-2*D6 + E6)/10	=E6+(D6-2*E6 + F6)/10
8	=A7+1	=B7	=C7+(B7-2*C7 + D7)/10	=D7+(C7-2*D7 + E7)/10	=E7+(D7-2*E7 + F7)/10

Converting these spreadsheets into graphs required an advanced spreadsheet program and a sizable computer. Consult the reference manual for your software to see if and how this can be done. Good luck!

How's your as

"Astronomy, considered most generally, is t
consists of defining celestial motions that app

EVEN IF YOU HAVEN'T TAKEN any "official" astronomy courses, you've certainly had occasion to acquaint yourself with the outstanding achievements in astronomy. Naturally, your physics course touched on mechanics, of which celestial mechanics is a part. And there's natural history, and geography, and ancient and medieval history . . . lots of opportunities to learn about this great branch of science. In all these subject areas, astronomical problems were undoubtedly discussed in passing—topics like planetary motion and star observation, navigation and eclipses, and the mechanics and properties of worlds beyond our own. This certainly testifies to the ancient ties connecting humans with the cosmos, as well to a continuing interest in what Johannes Kepler called "celestial physics."

Today you'll take part in only a few acts of the majestic show that nature presents daily on the celestial stage.

Questions and problems

1. How many angular minutes does the Earth rotate every minute?

2. What is the altitude of the Sun if a vertical object casts a shadow as long as it is tall?

3. When does the altitude of the stars above the Earth's horizon not change during the course of a day?

4. For an observer situated at one of the Earth's geographic poles, the Sun is above the horizon for half a year and below the horizon for the other half. What about the Moon?

5. Why does Venus's terminator (that is, the line separating day and night) look like an elliptical arc from the Earth?

6. Is it possible to see a reflection of the Sun in the water of a deep well?

7. As observed from the Earth, the Moon takes at least two minutes to rise completely above the horizon. How long does it take the Earth to rise above the horizon for an observer on the Moon?

8. The clouds on Venus are so thick that the stars can't be seen. If you were on Venus, could you be sure that the planet was rotating on its axis? If so, could you determine in which direction it was rotating?

9. What does Saturn's ring look like to observers at the equator and at the poles of that planet?

10. If there is a total lunar eclipse on Earth, what does an astronaut on the Moon see?

11. Why do total eclipses of the Sun in the northern hemisphere occur more often in summer than in winter?

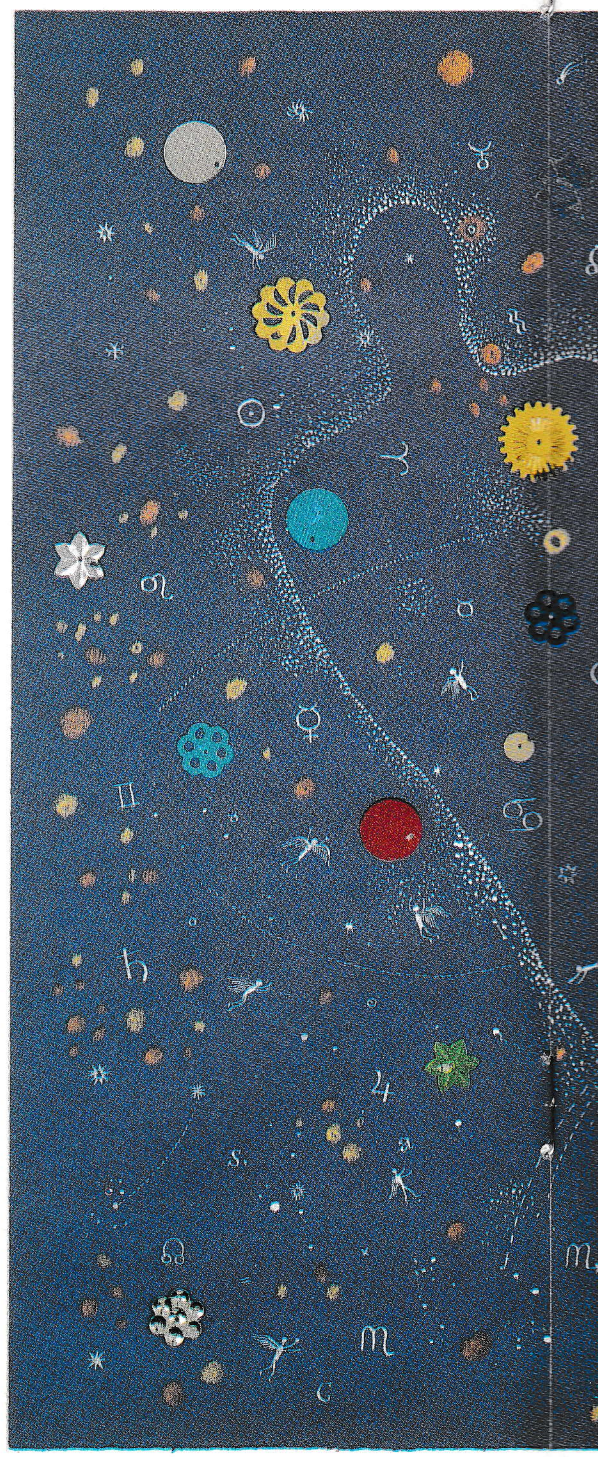
12. A white wall illuminated by the setting Sun looks brighter than the surface of the Moon at the same height above the horizon as the Sun. Does this mean that the lunar soil consists of dark rocks?

13. Does the Sun look the same from the Moon as from the Earth?

14. Would the apparent position of the stars change if the Earth's atmosphere suddenly disappeared?

15. What observations prove that comets don't travel in the Earth's atmosphere, as was believed in ancient times?

16. Why are more meteorites observed from midnight to dawn than from evening to midnight?



Art by Sergey Barkhin

astronomy?

, is the great problem of mechanics, which appear arbitrary."—Pierre Simon de Laplace

Microexperiment

The disks of the Sun and Moon near the horizon seem to be larger than when they are higher in the sky. How one can prove experimentally that this discrepancy is only apparent?

It's interesting that . . .

. . . one of the oldest known observatories is Stonehenge in Great Britain. This site is about four thousand years old. The first "real" astronomical observatory didn't appear in Europe until the 16th century.

. . . the detailed study of the heavens was stimulated by astrology. For example, about 2,500 years ago Assyrian priests could predict the dates of eclipses.

. . . the name of the inventor of the telescope is unknown. We know only this much: in 1604 a dealer in glass for spectacles, a Dutchman by the name of Janssen, "made a copy" of a telescope that belonged to an Italian who remained nameless.

. . . telescopes don't produce enlarged images of stars at all. A telescope serves to increase the angular distance between stars and the amount of light reaching the eye from a distant object. This is why people construct giant telescopes with reflectors several meters in diameter.

. . . optical instrument making became one of the first areas where the direct participation of physicists raised an empirical cottage industry to the level of technological production.

. . . the first complete astronomical textbook appeared in 1618: Kepler's *Epitome of Copernican Astronomy*.

. . . the famous English astronomer William Herschel quite seriously believed that the Sun is inhabited. He thought that the Sun's surface is cool enough to support life and that only the clouds floating above it are very hot.

. . . astronomical observations have lent support to some of the most important theories in physics. For example, the way light bends in the Sun's gravitational field, which was observed during a solar eclipse, or the deviation in Mercury's orbit, which was discovered in 1845, couldn't be explained by classical science, but fit quite naturally in the new conceptual framework provided by the general theory of relativity.

What to read in *Quantum*

♋ "Mushrooms and X-ray Astronomy," July/August 1994, p. 10

♌ "An Act of Divine Providence," May/June 1993, p. 41

♍ "The Inevitability of Black Holes," March/April 1993, p. 26

♊ "Late Light from Mercury," November/December 1993, p. 40

♒ "The Universe Discovered," May/June 1992, p. 12

♓ "Optics for a Stargazer," September/October 1994, p. 18

♎ "What Little Stars Do (And the Big Old Planets Don't)," March/April 1994, p. 22

—Compiled by A. Leonovich

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Editor's note: The Kaleidoscope in the September/October issue was also compiled by A. Leonovich.

The first photon

"Have other eyes, new light! And look! This is my glory, unveiled to mortal sight."—Bhagavad Gita

by Arthur Eisenkraft and Larry D. Kirkpatrick

AN ENVIRONMENTALLY rich village inhabited by curious people had been fully explored. All of the interesting corners and crevices of this remarkable land had revealed their secrets. There were no mysteries left. Oh, of course, some of the villagers remarked that a finer microscope may reveal a little more detail. But most were pleased with the comfort they felt in the familiar surroundings.

Such was the state of physics at the turn of the 20th century. The great syntheses of Newton and Maxwell remarkably explained so much about forces and motion, electricity, magnetism, and optics. Albert Michelson, America's first physics Nobel laureate, remarked that physics was complete and the following years would be devoted to simply increasing the precision of the experiments. Yet there existed a problem or two that appeared to be stumbling blocks. One was the photoelectric effect—the ability of light to free electrons from a metallic surface.

Maxwell and Hertz provided the theoretical and experimental evidence to convince the entire physics community that light was an electromagnetic wave. Almost a century earlier, Thomas Young had argued that light was a wave phenomenon and even measured the wavelength of this light. Yet a wave picture of light presented numerous difficulties in explaining the photoelectric effect. If the light is very dim, it should take hours to free an electron. Contrary to this prediction, the electrons are freed almost instantly. If the light is very intense, the expectation was that many electrons would be freed. However, intense, bright red light does not free even a single electron. Finally, when the light is able to free electrons, the kinetic energies of the electrons do not depend on the energy reaching the surface of the metal in a given time.

In what now seems like physics folklore, the young patent clerk Albert Einstein stepped onto the scene to propose that light behaves like a particle (known as a photon) and that each photon has an energy that depends on its frequency. More precisely, Einstein attributed an energy to each photon of light according to the equation

$$E = h\nu,$$

where h is Planck's constant ($6.63 \cdot 10^{-34}$ J · s) and ν is the frequency of the light.

The electrons are bound to the metal with a certain energy defined as the work function ϕ. When a photon strikes an electron in the metal, the electron acquires all the photon's energy and the photon disappears. The maximum kinetic energy a freed electron can attain is equal to the difference between the energy of the photons of light and the work function. (Some of the electron's kinetic energy is lost in getting to the surface. In fact, if it loses too much, the electron cannot escape the surface.)

Assume that the work function for a given metal is 3 eV. (One electron-volt equals $1.6 \cdot 10^{-19}$ J, the energy acquired by an electron falling through an electrical potential difference of 1 V. Using this energy unit, $h = 4.14 \cdot 10^{-15}$ eV · s.) A photon of red light (wavelength = 620 nm) has an energy equal to 2 eV. Since all photons of this light have energies of 2 eV, no single photon can provide the 3 eV required to free the electrons from the surface. No matter how intense the red light (no matter how many 2-eV photons are present), the electron will not be freed. If ultraviolet light of wavelength 310 nm impinges on the metal, each photon has a corresponding energy of 4 eV and electrons come flying off the metal. The maximum kinetic energy K these electrons can have is the difference between the two energies:

$$K_{\max} = h\nu - \phi.$$

An electron freed by this ultraviolet light may have a kinetic energy as

Art by Tomas Bunk

E=h(nu)
FREE AT LAST!!
POTAT
CHIPS
10¢
Tomás Bunk

large as 1 eV.

A vending machine provides a useful analogy for what is happening here. Assume we have a vending machine that can't accept multiple coins. You can submit only one coin at a time—a penny, a nickel, a dime, or a quarter. Potato chips cost 10 cents. If you put a penny in the machine, the penny is returned or lost. If you put a nickel in the machine, the nickel is also returned or lost. However, if you put a dime in the machine, a bag of potato chips is released to you. If you put a quarter in the machine, one bag of potato chips is released and up to 15 cents is returned in change. Note that when 25 cents is deposited, *two* bags of potato chips are not released. One coin can yield one bag or none—that's the rule.

The work function of the metal is the price of the potato chips. The energy of the incident photon is the value of the coin dropped into the machine. If an electron is freed, the kinetic energy it attains is represented by the change the machine releases. A low-energy photon of light (a coin less than 10 cents) is unable to free the electron (or a bag of potato chips.) No matter how many nickels you have, you will not be able to free the potato chips. No matter how intense the light (lots and lots of low-energy photons), no electrons will be released. If a high-energy photon is incident on the metal surface, electrons will be emitted. (If a quarter is placed in the machine, a single bag of potato chips will be released. More photons (more quarters), more electrons.

The photon (particle) nature of light satisfactorily explains the experimental results of the photoelectric effect. However, it is not able to explain the wave aspects of light so clearly demonstrated by Young, Maxwell, and others. This leaves us with the question: What is the true nature of light?

More insight into the nature of light was attained in the next decades, as it was shown that light behaves like a particle in an elastic collision between light and an electron, where the momentum of the photon of light is shown to be $h/\lambda = h\nu/c$. Arthur H. Compton earned recognition for these experimental studies in 1922.

The problems for this month focus on the particle nature of light and the associated energy and momentum of the photons.

A. Monochromatic light sources with a variety of wavelengths were incident on lithium and the maximum kinetic energies of the emitted electrons were recorded by Millikan (of the famous Millikan oil drop experiment, which measured the charge on an electron) as follows:

Wavelength (nm)	Kinetic energy (eV)
433.0	0.55
404.7	0.73
365.0	1.09
312.5	1.67
253.5	2.57

Plot these data and find the numerical values of Planck's constant and the work function for lithium.

B. Show that a free electron cannot completely absorb a photon. (In a metal, the surrounding atoms can participate in the collision, allowing energy and momentum to be simultaneously conserved.)

C. (i) The human eye is so sensitive that it can detect single photons of light. If the pupil of the eye has a diameter of 0.5 cm, at what distance would you place a 50-W light source (of wavelength 500 nm) so that the number of photons reaching the pupil is one per second on average? (ii) At what distance should the light source be placed so that the density of photons is on average 1 photon per cubic centimeter?

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

Superconducting magnet

Our contest problem in the November/December 1994 issue was adapted from the XXV International Physics Olympiad held in China and was concerned with the operation of a superconducting magnet using a superconducting switch.

A. Because the resistance r of the superconducting switch is zero, the voltage drop across the switch must be zero—that is, $V_s = 0$. Because the coil is in parallel with this switch, the voltage drop across the coil must also be zero—that is, $V_c = 0$. Therefore,

$$V_c = 0 = -L\frac{\Delta I_c}{\Delta t}.$$

This means that the current I_c through the coil cannot change (see figure 1). Conservation of charge (or equivalently Kirchhoff's current rule) requires that

$$I = I_c + I_s.$$

Therefore, any change in the total current I must result in a change in the current I_s through the superconducting switch, and I_s must drop linearly to $-\frac{1}{2}$ A. At t = 6 min, the current is flowing clockwise around the loop containing the coil and the superconducting switch.

B. Because the current through the superconducting switch I_s is zero, the voltage across the switch V_s remains zero even when the switch changes to the normal state and returns to the superconducting

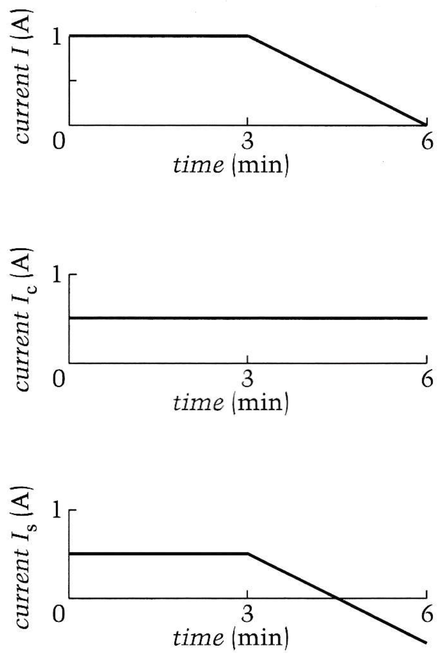


Figure 1

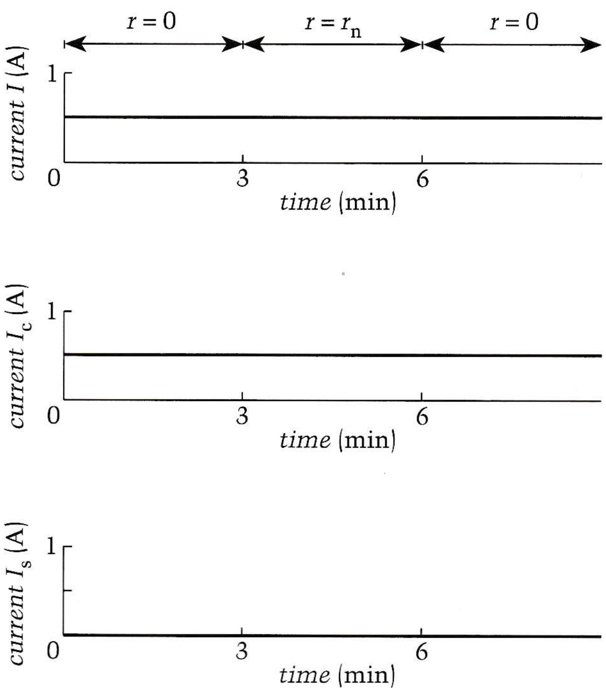


Figure 2

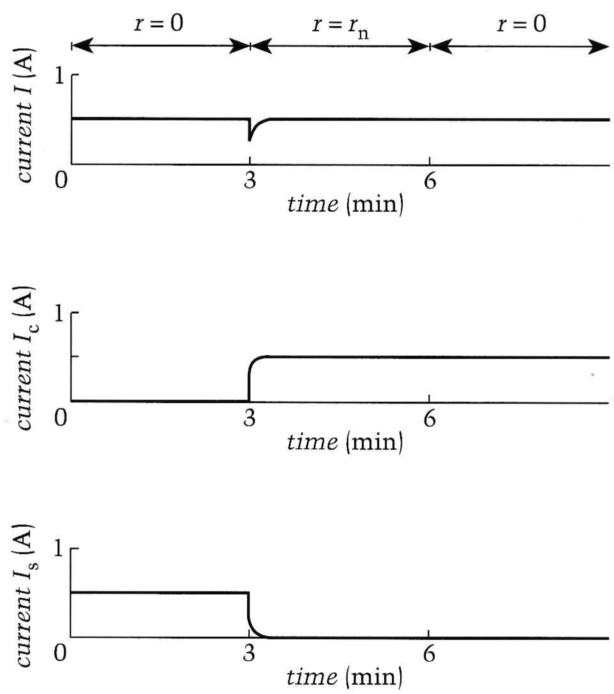


Figure 3

state. This means that the voltage across the coil V_c must remain zero, and the current I_c through the coil cannot change. Therefore, the total current I will also not change (see figure 2).

C. At $t = 3$ min, the resistance of the superconducting switch suddenly jumps from 0 to $r_n = 5\ \Omega$. Because the current I_c through the coil cannot change instantaneously due to its inductance, the total current I (and thus the current I_s through the superconducting switch) must drop from E/R to $E/(R + r_n)$. With $R = r_n$, both currents will drop to one half their original values, as shown in figure 3.

The currents now approach their steady-state values exponentially with a time constant $\tau = L/R_t$, where R_t is the total resistance connected to the inductance. Since the two resistances are in parallel, we have

$$\tau = \frac{L}{R_t} = \frac{10\ \text{H}}{2.5\ \Omega} = 4\ \text{s}.$$

At steady state, there cannot be any current through the superconducting switch. Otherwise, there would be a voltage drop across the coil, necessitating a changing current through the coil in violation of the steady-state condition. Therefore, the current from the power source returns to its original value and all of this current must pass through the coil.

As in part B, at $t = 6$ min there is no current through the superconducting switch and, therefore, there are no changes in the currents when the superconducting switch returns to zero resistance.

D. Unfortunately, we made an error in the statement of this part of the problem. It should have stated that "we will destroy the switch if the current through the switch *in the normal state* exceeds 0.5 A." The parenthetical remark that follows is correct, but is not relevant to the problem.

Begin by closing the power switch K and increasing the total current I to 20 A. Note that

CONTINUED ON PAGE 50

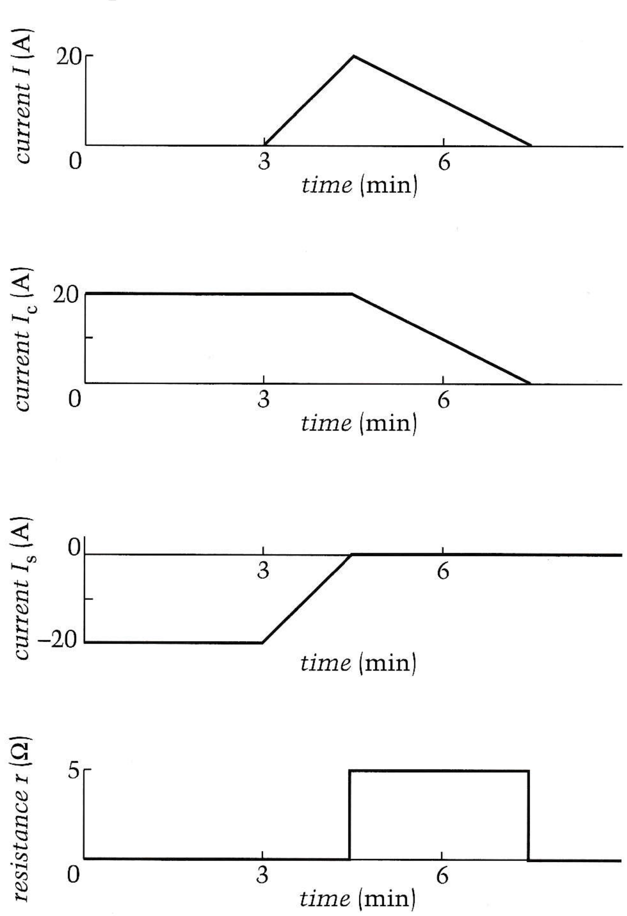


Figure 4

READER SURVEY

Free buttons for *Quantum* readers!

We'll send an attractive Quantum *button to everyone who completes and returns this questionnaire. (Photocopies are acceptable.) Anyone who reads* Quantum *regularly—whether a subscriber or not—is eligible to complete this survey and receive a button. All information is completely confidential and strictly for research purposes.*

1. Sex
- ❐ Male
- ❐ Female

2. Age
- ❐ Under 15
- ❐ 15–24
- ❐ 25–34
- ❐ 35–44
- ❐ 45–54
- ❐ 55–64
- ❐ 65+

3. Marital status
- ❐ Single
- ❐ Married
- ❐ Separated/divorced/widowed

4. Education
- ❐ Attending elementary school
- ❐ Attending middle school
- ❐ Attending high school
- ❐ High school graduate
- ❐ Attending college
- College graduate:
  - ❐ Bachelor's degree
  - ❐ Postgraduate study without degree
  - ❐ Master's degree
  - ❐ Doctorate or equivalent

5. Occupation
- ❐ Student
- ❐ Teacher/professor
- ❐ Other educational professional
- ❐ Researcher/scientist
- ❐ Professional/managerial
- ❐ Professional/nonmanagerial
- ❐ Retired

6. Do you have access to a computer?
- ❐ Yes, at home
- ❐ Yes, at school/work
- ❐ No

7. Do you use the Internet? (Check all that apply)
- ❐ Yes, for electronic mail
- ❐ Yes, for downloadable information (FTP, telnet)
- ❐ Yes, for the World Wide Web (multimedia)
- ❐ No

8. How do you get *Quantum*?
- ❐ Subscribe
- ❐ Newsstand
- ❐ From the library
- ❐ From a teacher
- ❐ From a colleague/friend
- ❐ Other________________

9. Approximately how much time do you spend reading each issue?
- ❐ Half-hour or less
- ❐ Half-hour to one hour
- ❐ One hour to two hours
- ❐ More than two hours

10. Would you still read *Quantum* if it had fewer illustrations?
- ❐ Yes
- ❐ No

11. What is your favorite department in *Quantum*?
- ❐ Brainteasers
- ❐ Challenges in Physics and Math
- ❐ Physics Contest
- ❐ Math Investigations
- ❐ Kaleidoscope
- ❐ Anthology
- ❐ In the Lab
- ❐ At the Blackboard
- ❐ Toy Store
- ❐ Happenings
- ❐ Don't have a favorite

12. What types of periodicals do you read regularly?
- ❐ Science magazines (*Scientific American, Discover, Omni,* etc.)
- ❐ Literary magazines (*The New Yorker, Harper's, Atlantic Monthly,* etc.)
- ❐ Education journals (*Education Week, Teacher, Scholastic, The Science Teacher,* etc.)
- ❐ Computer magazines (*Byte, PC Magazine, Mac Week,* etc.)
- ❐ Other________________

13. ***Optional***—please include any comments you may have about *Quantum*, positive or negative (content, design, price, delivery, etc.). Attach a separate sheet if necessary.

To receive your button, complete the following (please print neatly):

Name

Address

City

State

Zip

Please return this form **no later than JULY 1, 1995,** to:

Elisabeth Tobia
Quantum magazine, NSTA
1840 Wilson Boulevard
Arlington, VA 22201-3000

Or you can **fax** it: **703 522-6091.**

Maximizing the greatest

Quod differtur non aufertur

by George Berzsenyi

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

THE LATIN SAYING IN THE subtitle might be loosely translated as "You can put off solving problems, but you can never really get rid of them." To demonstrate the truth of this, we will revisit a problem area first called to my attention by F. David Hammer, with whom I enjoyed a lively mathematical correspondence throughout the 1980s.

In one of his letters David suggested the problem of finding the maximum value of the greatest common divisor of $n^3 + 1$ and $(n + 1)^3 + 1$, as n ranges through the set of positive integers. I posed it in the Third Texas Mathematical Olympiad in 1981; my first challenge to my readers is to **show that the answer is 7, attained when $n \equiv 5 \pmod 7$**. David also found that the corresponding answer for $n^4 + 1$ and $(n + 1)^4 + 1$ is 17; I used this recently as a problem in *Math Horizons*.

It turns out that the situation gets more complicated for exponents greater than 4, and a lot more complicated if we add an arbitrary integer k, rather than 1, to each integral power. Nevertheless, in view of the wide availability of powerful computer algebra systems, it should be possible to shed some light on the relationship of m and k to $G(m, k)$, the greatest common divisor of $n^m + k$ and $(n + 1)^m + k$, as n ranges through the positive integers. After some initial investigations by my colleague Allen Broughton (using the computer program Maple), I found some answers and many more questions (using Mathematica). I hereby challenge my readers to surpass our findings.

It turns out that $G(2, k) = 4k + 1$, as one can readily deduce from the identity

$$(2n + 3)(n^2 + k) + (-2n + 1)((n + 1)^2 + k) = 4k + 1.$$

This was the basis for Problem 13 of the 1985 American Invitational Mathematics Examination. No such identity is known for $G(m, k)$ for $m > 2$. By applying the Euclidean Algorithm to the polynomials $n^m + 1$ and $(n + 1)^m + 1$, I managed to derive similar expressions for $G(m, 1)$ for $3 \le m \le 7$, but didn't get further. I also made some tabulations of $G(m, k)$ for $3 \le m, k \le 10$, but didn't manage to discern any patterns. In particular, it would be of interest to know **for what values of m and k is $G(m, k) = 1$?**

To whet your appetite, the formulas behind $G(3, 1)$, $G(4, 1)$, and $G(5, 1)$ are displayed in the box below. Perhaps you will succeed in finding similar formulas for other values of m and k and some patterns among them.

$$(3n^2 - 6n + 5)((n + 1)^3 + 1) - (3n^2 + 3n - 4)(n^3 + 1) = 2 \cdot 7$$

$$(20n^3 - 10n^2 - 12n + 23)((n + 1)^4 + 1) - (20n^3 + 70n^2 + 68n - 5)(n^4 + 1) = 3 \cdot 17$$

$$(120n^4 - 85n^3 + 15n^2 - 30n + 91)((n + 1)^5 + 1) - (120n^4 + 515n^3 + 790n^2 + 395n - 159)(n^5 + 1) = 11 \cdot 31$$

The Great Art

The controversial origins of "Cardano's formula"

by Semyon Gindikin

THIS ARTICLE IS DEVOTED TO the major achievement of sixteenth-century mathematics: the discovery of the formulas for solving algebraic equations of the third and fourth degrees. The events surrounding this discovery still have the power to hold us spellbound, as the fates of four scholars—del Ferro, Tartaglia, Cardano, and Ferrari—become capriciously intertwined. The title of this article refers to Cardano's *Ars Magna*, published in 1545. In the words of Felix Klein, "this extremely valuable work contains seeds of modern algebra that transcend the bounds of the old mathematics."

The 16th century marked the revival of European mathematics after its medieval hibernation. At first, European scientists tried to understand and study what was done by their classical and oriental (Indian and Arabic) predecessors. The first achievements of sixteenth-century mathematicians themselves were in algebra. (This is because algebra was in its infancy then, whereas geometry was already a fully developed science.)

The state of algebra at the end of the 15th century was summarized in one of the first mathematics books printed, *Summa de Artithmetica, Geometria, Proportione et Proportionalita*, published in Venice in 1494. The book was in Italian, and so it was one of the first scientific books that was not written in Latin. Its author was Luca Pacioli, a monk and a friend of great Leonardo da Vinci. At the end of the book Pacioli writes: for solving cubic equations, "the art of algebra has not yet given a method, as it has not given a method for squaring a circle." These words were apparently taken as a statement of the impossibility of a formula for solving cubic equations.

Scipione del Ferro

However, there was a person who wasn't deterred by Pacioli's opinion. Scipione del Ferro (1465–1526), a professor of mathematics in Bologna, managed to find a method for solving the equation

$$x^3 + ax = b. \qquad (1)$$

Since negative numbers weren't used at that time, the letter coefficients in equation (1) (and throughout the article) are assumed to be positive. So equation (1) and the equation

$$x^3 = ax + b \qquad (2)$$

were considered completely different equations! Del Ferro's own explication has not been preserved. He conveyed his method to his son-in-law and successor in his professorship Annibale della Nave and to his student Antonio Maria Fior. After his teacher's death, Fior decided to take advantage of the secret entrusted to him to become invincible in problem-solving debates ("scientific duels") that were the custom back then. At the end of 1534 he challenged Niccolò Tartaglia, a mathematician from Venice, to a debate.

Niccolò Tartaglia

Tartaglia was born in 1499 in Brescia into the family of a poor mounted postman named Fontane. In his childhood, when his town was seized by the French, he received a wound in his larynx and thereafter spoke with difficulty. That's where his nickname "Tartaglia" ("stammerer") comes from. From his early years, Niccolò was left in the care of his mother. They were so poor that he went to school for only two weeks. In his writing class this was enough time for him to reach the letter K. Tartaglia was forced to leave school without ever having learned to write his own name. However, he continued to study on his own and became a "master of the abacus" (something like a teacher of arithmetic in a private commercial school). After 1534, Tartaglia lived in Venice.

Tartaglia's scientific studies were stimulated by his contact with engineers and artillery officers from the famous Venetian arsenal. In 1537 he published *New Science*, a book devoted to mechanical problems. This book played an important role in the development of ballistics. In 1546 he published another book, *Problems* *and Diverse Inventions*. In the first of these books Tartaglia followed Aristotle in maintaining that a body projected at an angle first flies along a slanting line, then in a circular arc, and, finally, falls vertically down. But in the second book he states that the trajectory "has no segments that would be absolutely straight." Tartaglia translated Archimedes and Euclid into Italian, which he called "popular" (vernacular) in contrast to Latin.

When Tartaglia received Fior's challenge, he thought he'd score an easy victory. He wasn't even concerned when he discovered that all thirty of Fior's problems involved equation (1) for different a and b. Tartaglia assumed that Fior himself couldn't solve his problems: "I thought that none of them could be solved because Brother Luca assures us in his treatise that this sort of equation cannot be solved with a general formula."

The "combatants" had 50 days to present their solutions to a notary public. When that time had almost run out, Tartaglia heard a rumor that Fior knew of some mysterious method for solving equation (1) after all. The prospect of treating 30 of Fior's friends to a banquet—for those were the rules of the battle (one friend fêted for each problem solved)—didn't appeal to Tartaglia. He made a titanic effort and, eight days before the deadline (on February 4, 1535), fortune smiled on him: he found the method he needed. Thus armed, Tartaglia solved all his opponent's problems in two hours, whereas Fior solved none of Tartaglia's problems in time. (Strangely enough, he didn't even manage to solve one problem that could have been solved by del Ferro's method.)

The only known portrait of Niccolò Tartaglia.

Soon Tartaglia discovered a method for solving equation (2). The Fior–Tartaglia "duel" and Tartaglia's victory became widely known. He was asked to reveal his secret, but he refused. Then someone turned up who managed to persuade him—Gerolamo Cardano.[1]

Gerolamo Cardano

Cardano was born in Pavia on September 24, 1501, into a lawyer's family. After graduating from the university, Gerolamo decided to devote himself to medicine. He was an illegitimate child, so at first he had to practice in the provinces for a long time. It wasn't until August 1539 that the physicians' college of Milan accepted Cardano, after its rules had been changed especially for that purpose. Later he even became the rector of this college. Cardano was one of the most renowned physicians of his time, second only perhaps to his friend Vesalius.

In the twilight of his life Cardano wrote an autobiography, *Book of My Life*. There he mentions his mathematical work only a few times, but writes in detail about his medical studies. Cardano claims that he described methods of treating as many as five thousand diseases; that the number of medical problems he diagnosed reached forty thousand; and that his therapeutic suggestions numbered almost two hundred thousand. We should, of course, treat these figures with a healthy dose of skepticism. Nonetheless, Cardano's fame as a physician is indisputable. He asserts that he experienced only three failures in his medical practice.

But his medical work didn't take up all his time. In his leisure moments he engaged in all kinds of intellectual activities: philosophy, astrology, physics, mechanics, mathematics.

Cardano worked up the horoscopes of both the living and the dead (including Jesus Christ, Petrarch, Dürer, Vesalius, and Luther). The Pope made use of Cardano's services as an astrologer. (One nasty legend has it that Cardano took his own life to confirm his own horoscope.)

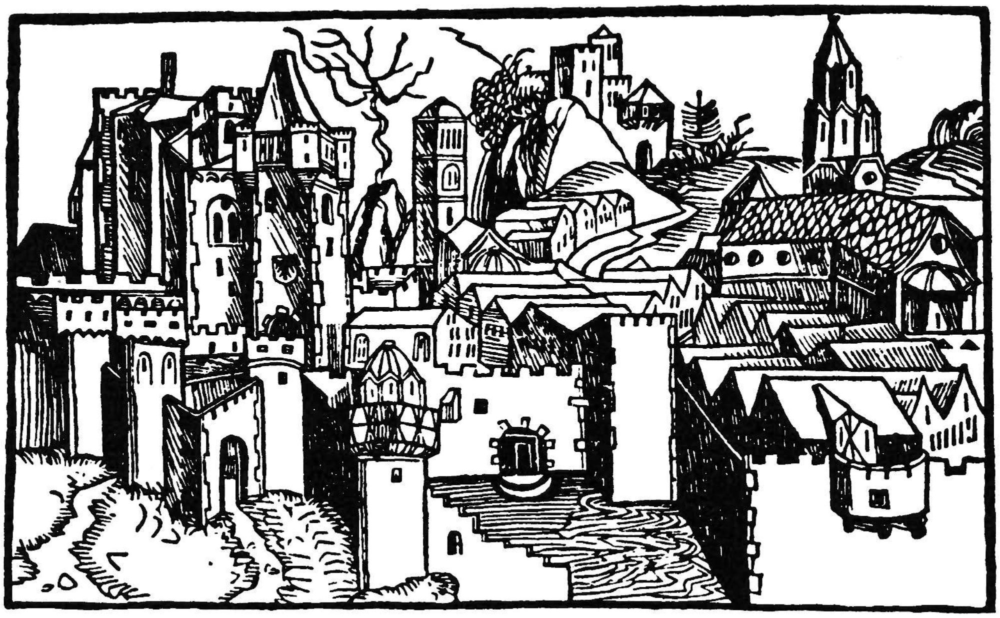
Cardano's book *The Subtlety of Things* (1550) was translated into French, and throughout the 17th century it was a popular textbook on statics and hydrostatics. When Galileo observed the oscillation of a natural pendulum (a chandelier in a cathedral), he followed Cardano's advice to use one's own pulse for measuring the time. Cardano wrote about the impossibility of perpetual motion. Some of his remarks can be interpreted as the virtual work principle. Cardano established experimentally the ratio of the densities of air and water. He invented the system of coupling two shafts for the king's coach, now called the cardano joint (or cardano shaft) and widely used in cars. (To be fair, I should point out that the idea of such a joint dates back to antiquity; also, one of Leonardo's drawings depicts a compass with a cardano joint.)

Some of Cardano's works were encyclopedic in scope. During the Renaissance encyclopedias were written by individual scholars. The

[1] Known variously in the literature as Gerolamo Cardano, Geronimo Cardano, and Jerome Cardan.—*Ed.*

first encyclopedias that were the result of collective effort didn't appear until a century and a half later.

Cardano wrote an enormous number of books (some were published, some remained in manuscript form, and some he destroyed—see the last section). Just a description of them filled up an entire book, *On My Own Works*. His books on philosophy and ethics were popular for years. His book *On Consolation* was translated into English and influenced Shakespeare. Some Shakespearean scholars even assert that Hamlet recited his famous monologue "To be or not to be . . ." with this book in his hand.



Milan in 1493 (woodcut).

For forty years Cardano played chess ("I would never be able to express in a few words how much damage, without any compensation, it caused in my domestic life . . ."); for twenty five years he played at dice (". . . but dice caused me even greater harm than chess did"). From time to time he would forsake all other activities to engage in gambling. As a by-product of this passion Cardano wrote *A Book on Dice-playing* in 1526 (not printed until 1663). This book studies problems of probability and combinatorics; it also contains some observations about the psychology of gamblers.

Cardano and Tartaglia

By 1539 Cardano had finished his first entirely mathematical book, *The Practice of General Arithmetic*. His intention was that it replace Pacioli's *Summa*. When he heard about Tartaglia's secret, he was consumed by the desire to enhance his book with it.

In January 1539 Cardano asked Tartaglia to send him the rule for solving equation (1) either for publication in his book or on the condition that it be kept secret. Tartaglia refused: "Begging His Lordship's pardon, but if I wish to publish my discovery I'll do it in my own book and not in someone else's." On February 12, Cardano repeated his request. Tartaglia didn't budge. On March 13, Cardano invited Tartaglia to visit him in Milan and promised to introduce him to the governor of Lombardy. It seems Tartaglia found this prospect enticing: he accepted the invitation. The decisive conversation took place at Cardano's home on March 25.

Here's an excerpt from the notes of this conversation (it should be kept in mind that they were taken down by Tartaglia; Cardano's student Ferrari says they don't completely correspond to the facts):

Niccolò. I am telling you, I turned you down not only because of this chapter and the discovery made in it, but because of the things that can be discovered knowing it, for this is the key that unlocks the door to countless other areas for investigation. I would have found a general rule for many other problems long ago if I were not so busy at present with translating Euclid into the vernacular (by now I have reached the end of the 13th book). But when this work, which I have already begun, is finished, I am going to publish a work for practical application along with the new algebra . . . If I reveal it to any theoretician (such as Your Lordship), he will easily be able to make use of this explanation to write other chapters (because this explanation is easily applicable to other questions) and publish the fruits of my discovery under his own name. This would destroy all my plans.

Signor Gerolamo. I swear to you by the Lord's Holy Gospel and not only pledge my word as an honest man never to publish this discovery of yours, if you entrust it to me, but also promise—and let the conscience of a true Christian be your guarantee—to encipher it such that nobody will be able to read this writing after my death. If I am trustworthy, in your opinion, do it; if not, let us drop the matter.

Niccolò. If I did not believe this oath of yours, I would certainly deserve to be considered a nonbeliever myself.

So Tartaglia allowed Cardano to persuade him. It's difficult to understand from the notes above what made him change his mind. Was he really so moved by Cardano's oaths? After divulging his secret, Tartaglia immediately left Milan—he even declined to meet with the governor, which was why he took the trip in the first place. Did Cardano hypnotize him, or what?

When on May 12 Tartaglia received the freshly printed *Practice of General Arithmetic* without his "recipe" (the method of solution was rendered in the form of a Latin poem—they couldn't write formulas back then), he calmed down somewhat.

Cardano received from Tartaglia a finished method for solving equation (1) without a trace of a proof. He spent a great deal of effort on carefully verifying and substantiating the rule. From our vantage point it's hard to understand what the problem was: just substitute into the equation and check it! However, without well-developed algebraic notation, things that can now be done automatically by any high school student were then accessible only to a few select people. Without acquainting ourselves with the original texts of that time, it's impossible to evaluate the extent to which algebraic techniques "economize" our thinking. The reader must always bear this in mind so as not to be misled by the apparent "triviality" of problems that aroused such feverish passions in the 16th century.

Ludovico Ferrari

Cardano had a young assistant in his mathematical work, Ludovico Ferrari (1522–1565). In a list he made of his fourteen students, Cardano counted Ferrari as one of the three most outstanding.

In 1543 Cardano and Ferrari went to Bologna, where della Nave allowed them to acquaint themselves with the papers of the late del Ferro. There they verified that del Ferro had known Tartaglia's rule.

Apparently they and their contemporaries knew almost nothing about del Ferro's formula. Cardano would hardly have pursued Tartaglia so relentlessly if he had known that the same information could have been obtained from della Nave.

Today almost all historians of mathematics agree that del Ferro invented the formula, Fior knew about it, and Tartaglia rediscovered it knowing that Fior had it. But none of these facts has been rigorously proved!

At the end of his life Tartaglia wrote: "I can assure you that the described theorem has been proved neither by Euclid nor by anyone else, except only Gerolamo Cardano, to whom we showed it. . . . In 1534 in Venice I found the general formula for the equation. . . ."[2]

[2]Another source gives the date as February 4, 1535.

It's hard to make ends meet in this story.

By 1545 Cardano had learned how to solve not only equations (1) and (2), but also the equation

$$x^3 + b = ax, \tag{3}$$

as well as a "complete" cubic equation that contains a term with x^2. By the same time Ferrari had created a method for solving fourth-degree equations.

Ars Magna

Either his acquaintance with del Ferro's papers, or heavy pressure from Ferrari, or most likely an unwillingness to bury the results of his work of many years made Cardano include everything he knew about cubic equations in the book *Ars Magna, sive De Regulis Algebraicis* (*The Great Art, or On Rules of Algebra*) published in Nuremberg in 1545.

In his preface Cardano presents the history of the issue:

> In our time Scipione del Ferro discovered a formula according to which the cube of the unknown plus the unknown is equal to the number. It was a very beautiful and remarkable work. Because this art exceeds all human adroitness and all mental clarity of a mortal, it must be considered a gift of heavenly origin, and also an indication of the mind's power, and this is so glorious a discovery that he who achieved it can be expected to achieve anything. Competing with him, Niccolò Tartaglia from Brescia, our friend, having been challenged by del Ferro's student by the name of Antonio Maria Fior, solved the same problem in order to not to be defeated, and after repeated requests made over a long period of time, passed it on to me. I was misled by the words of Luca Pacioli, who said that there is no general solution to equations of this kind, and, although I possessed many discoveries made by myself, I nevertheless did not despair of finding what I did not dare look for. However, when I received this chapter and reached its solution, I saw that it can be used to do much more: and, already with greater confidence in my deeds, I made further discoveries during my investigation, partly on my own, partly together with Ludovico Ferrari, my former student.

Cardano's method of solving equation (1) can be presented in an updated form as follows. We'll look for a solution to equation (1) in the form $x = \beta - \alpha$. Then $x + \alpha = \beta$ and

$$x^3 + 3x^2\alpha + 3x\alpha^2 + \alpha^3 = \beta^3. \tag{4}$$

Since $3x^2\alpha + 3x\alpha^2 = 3x\alpha(x + \alpha) = 3x\alpha\beta$, equation (4) can be rewritten as

$$x^3 + 3\alpha\beta x = \beta^3 - \alpha^3. \tag{5}$$

Let's try to choose the pair $\langle\alpha, \beta\rangle$ knowing the pair $\langle a, b\rangle$ such that equation (5) coincides with equation (1). To this end, the pair $\langle\alpha, \beta\rangle$ must be a solution of the simultaneous equations

$$\begin{cases} 3\alpha\beta = a, \\ \beta^3 - \alpha^3 = b \end{cases}$$

or the equivalent system of equations

$$\begin{cases} \beta^3(-\alpha^3) = -\dfrac{a^3}{27}, \\ \beta^3 + (-\alpha^3) = b. \end{cases}$$

By the property of the roots of a quadratic equation, β^3 and $-\alpha^3$ are the roots of an auxiliary quadratic $y^2 - by - a^3/27 = 0$. Since we are looking for *positive* roots of equation (1), $\beta > \alpha$. Thus, by the quadratic formula

$$\beta^3 = \frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}$$

and

$$-\alpha^3 = \frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}.$$

Therefore,

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} - \sqrt[3]{-\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}}.$$

If a and b are positive, x is positive too.

This calculation follows Cardano's train of thought only ideally.

He reasons in geometric terms: if a cube with side length $\beta = \alpha + x$ is cut by planes parallel to its faces into a cube with side length α and a cube with side length x, then in addition to the two cubes, three rectangular parallelepipeds measuring $\alpha \times \alpha \times x$ and three others measuring $\alpha \times x \times x$ are obtained. The relation between their volumes yields equation (4); to pass to equation (5), the parallelepipeds of different types are united in pairs.

"Since I was aware that the part that Tartaglia gave me was discovered by means of a geometric proof," Cardano wrote, "I thought that this is indeed the royal way leading to all the other parts."

Equation (2) can be solved by substituting $x = \beta + \alpha$, but here we might run into the case where the initial equation has three real roots, while the auxiliary quadratic has no real roots. This is the so-called *irreducible case.*[3] It caused Cardano a lot of trouble (and probably Tartaglia as well).

Cardano solved equation (3) by means of reasoning that was very daring for that time: he turned the negativeness of a root to good account. No one before him used negative numbers so resolutely, though even Cardano himself was still far from free in his treatment of them: he considers equations (1) and (2) separately!

Cardano also gave a complete account of the general cubic equation $x^3 + ax^2 + bx + c = 0$ (and Tartaglia surely had nothing to do with this problem!). In modern terms, substituting $x = y - a/3$ in this equation eliminates the term with x^2.

Cardano dared to consider not only negative numbers (he called them "purely false"), but also complex numbers (these he called "truly sophistic"). He noticed that if we operate on them according to certain natural rules, then any quadratic equation without real roots can be thought of as having complex roots. Perhaps Cardano arrived at complex numbers via the "irreducible" case.

[3]We'll offer a special treatment of this case in an upcoming issue.—*Ed.*

The Great Art also reflected Ferrari's personal contribution—a method for solving quartic equations.

In modern terms, Ferrari's method for the equation

$$x^4 + ax^2 + bx + c = 0 \qquad (6)$$

(to which a complete quartic equation is easily reduced)[4] consists of the following.

We introduce an auxiliary variable t and rewrite equation (6) in the equivalent form

$$\left(x^2 + \frac{a}{2} + t\right)^2 = 2tx^2 - bx + \left(t^2 + at - c + \frac{a^2}{4}\right). \qquad (7)$$

Then we choose t such that the two roots of the quadratic polynomial in x on the right side of equation (7) coincide—that is, its discriminant is zero:

$$b^2 - 4 \cdot 2t \cdot \left(t^2 + at - c + \frac{a^2}{4}\right) = 0.$$

Thus we get an auxiliary cubic equation for t. Let t_0 be any of its roots—it can be found by Cardano's formula. Then equation (7) can be rewritten as

$$\left(x^2 + \frac{a}{2} + t_0\right)^2 = 2t_0\left(x - \frac{b}{4t_0}\right)^2. \qquad (8)$$

This equation breaks down into two quadratic equations that yield four roots of the original equation.

So Ferrari's method reduces a quartic equation to one cubic and two quadratic equations.

However, the place of *The Great Art* in the history of mathematics is primarily due not to its particular results about cubic and quartic equations but to the fact that certain general algebraic notions (for instance, the multiplicity of a root) and statements (if the equation $x^3 + ax^2 + bx + c = 0$ has three real roots, then their sum is $-a$) appeared here for the first time.

[4]Quartic equations are also discussed in "What You Add is What You Take" in the November/December 1994 issue of *Quantum.* —*Ed.*

Ferrari and Tartaglia

It's not hard to imagine the impression *The Great Art* made on Tartaglia when it appeared in 1545. In the last part of his book *Problems and Diverse Inventions* (1546), Tartaglia published his correspondence with Cardano and the notes of their conversation. He attacked Cardano with curses and reproaches. Cardano didn't respond to the attack. On February 10, 1547, Ferrari rather than Cardano responded to Tartaglia. He objected to Tartaglia's reproaches, pointed out the flaws in his book, and in one case blamed him for appropriating a result obtained by someone else and for having a faulty memory (apparently a grave accusation in those days). In the end, Tartaglia was challenged to a public debate "on geometry, arithmetic, or disciplines connected with them such as Astrology, Music, Cosmography, Perspective, Architecture, and so on."

In his reply of February 19, Tartaglia tried to draw Cardano himself into the squabble: "I wrote in such a heated and insulting tone in order to force His Lordship (not you) to write something in his own hand, because I have some old scores to settle with him."

Wrangling over the conditions of the debate dragged on. Tartaglia began to understand that Cardano would remain on the sidelines. Then he began to point up Ferrari's dependence, calling him "Cardano's creature" (as Ferrari called himself in his challenge). The "Questions" that Tartaglia, according to tradition, sent in reply to the challenge were addressed to both: "You, Signor Gerolamo, and you, Signor Ludovico . . ."

In the end, Tartaglia agreed to a debate with Ferrari. It took place in the presence of many noble persons, but with Cardano absent, in Milan on August 10, 1548. Only Tartaglia's brief notes about the debate have

Gerolamo Cardano at the age of 46 (woodcut).

survived, and it's almost impossible to restore any real picture from them. It seems, though, that Tartaglia suffered a defeat.

However, the debate had nothing to do with the problem that aroused the controversy. In general, debates of this sort were as little related to clarifying the truth as real duels.

For each actor, a closing scene

In 1556, a year before Tartaglia died, his *General Treatise on Number and Measure* began to come out. It contained the results in combinatorics and probability that Tartaglia obtained during the years of his dispute with Cardano. The *Treatise* says very little about cubic equations. The final parts of the *Treatise* appeared after the author's death.

Ferrari achieved widespread fame after the debate. He gave public lectures at Rome, led the cadastral administration in Milan, and took part in the education of the king's son. But he left no other traces in science! He died in 1565 at the age of 43.

Cardano outlived them both. But his life wasn't easy. One of his sons poisoned his wife out of jealousy and was put to death. Another son became a tramp and robbed his own father. In 1570 Cardano himself was put in prison and his property was confiscated. (The reason for the arrest remains unknown. We know only that while he was waiting to be arrested he destroyed 120 of his books.) Cardano ended his days at Rome in the position of "private person" (his expression), receiving a modest pension from the Pope. He devoted his remaining days to writing the autobiographical *Book of My Life*. The last event mentioned in this book is dated April 28, 1576, and on September 21, 1576, Cardano died.

In his last book Cardano mentions Tartaglia four times. He writes that Tartaglia preferred to have him as a "rival and victor rather than a friend and person indebted to him for his good deeds." Elsewhere he numbers Tartaglia among his critics who "did not range beyond the scope of grammar." However, in the closing pages he writes: "I confess that in mathematics I borrowed a few things, albeit an insignificant number, from brother Niccolò."

Apparently his conscience was bothering him!

The Cardano–Targaglia controversy died away and was all but forgotten. The "cubic formula" was linked to *The Great Art* and gradually came to be called Cardano's formula, though for some time del Ferro's name was also mentioned—after all, del Ferro's authorship was emphasized by Cardano himself. However, such injustices in naming aren't all that rare in mathematics.

Historians of mathematics returned to the controversy at the beginning of the 19th century, after rediscovering the existence of the offended party, Tartaglia, who had been practically forgotten by that time. The story attracted attention once again, and amateurs as well as professionals were ready to fight for Tartaglia's honor. As the story was repeated countless times and worked its way into popular culture, Cardano was sometimes made out to be an adventurer and a scoundrel who stole Tartaglia's discovery and took full credit for it.

By the end of the 19th century some of the discussion took the form of serious studies in the history of mathematics. Mathematicians understood the important role played by Cardano's works in sixteenth-century science. What Leibniz said two centuries earlier became clear to many: "With all his faults, Cardano was a great man; without them he would have been perfect."

The great historian of mathematics Moritz Cantor (1829–1920) (not to be confused with Georg Cantor, the creator of set theory!) repeated the conjecture, voiced by Ferrari many years earlier, that Tartaglia didn't rediscover del Ferro's rule, but got it ready-made from other sources.

Over the course of a century and a half, passions died away, then flared up anew. But perhaps this is one of those questions that even now cannot be answered unambiguously.

And the formula for solving cubic equations remains forever "Cardano's formula."

Readers write . . .

Prof. Doris Schattschneider of Moravian College (Bethlehem, Pennsylvania) points out that the proof of Napoleon's theorem by tiling in "Suggestive Tilings" (July/August 1994) appeared earlier, in an article by John F. Rigby ("Napoleon, Escher, and Tesselations," *Mathematics Magazine*, Vol. 64, 1991, pp. 242–46) of which the author of the *Quantum* article was not aware. Rigby's article also gives many interesting historical references.

"HOW DO YOU FIGURE?" CONTINUED FROM PAGE 19

dependence of the magnetic field B on the x-coordinate (along which the ring moves) is shown in figure 2, where $B_0 = 1$ T, $a = 10$ cm. (V. Afanasyev)

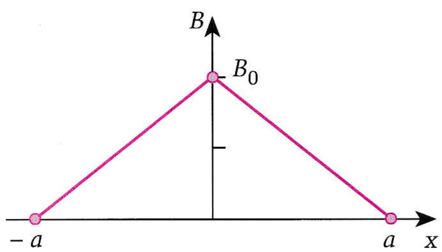


Figure 2

P145

Nuclear disintegration. A nucleus of mass M traveling with velocity $\mathbf{v}$ breaks into two identical pieces. The internal energy of the nucleus is E_1, and that of each piece is E_2 ($E_1 > 2E_2$). Determine the maximum possible angle between the velocity vector of one piece and the vector $\mathbf{v}$. (I. Slobodetsky)

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

"MAGNETIC MONOPOLY" CONTINUED FROM PAGE 9

monopoles exist, since no other significant events were recorded in the experiment.

In 1983 a group at Berkeley put together an apparatus that tried to combine the signals from a number of superconducting loops, each having an area of the order of 1 m². No monopole events were observed. The method holds great promise, however, as it is very sensitive and is independent of the mass and speed of the passing monopole.

Faraday's law would require the loop to produce a current that would produce a flux equal and opposite to the total flux of the monopole passing through the loop. Since the whole monopole passes through, the entire flux of the monopole must be balanced by the loop. Calculating this total flux is easy if we imagine a spherical surface of radius r about the monopole. The flux is then

$$\phi = \frac{\mu_0}{4\pi}\frac{q^*}{r^2}\cdot 4\pi r^2$$
$$= \mu_0 q^*.$$

The astute physics student should have an objection to this. Why do I calculate the total flux of a monopole here when, in most situations, it is only the flux due to the component of the magnetic field perpendicular to the face of the current loop that contributes to induction effects? The difference here is that we're not simply changing a magnitude or a direction of flux with respect to the current loop—we're actually passing an isolated source of field lines through a loop.

To see how this makes a difference, imagine breaking each field line into components perpendicular and parallel (and in the radial direction) to the face of the current loop. As the monopole approaches the current loop, the magnitude of the perpendicular field lines increases and an induced emf will result in the loop due to Faraday's law. However, the radial field lines parallel to the face also contribute an induced emf. To understand this, imagine the motion of the current loop from the frame of reference of the monopole. As the actual wire passes over the radial field lines, the free charge carriers in the wire will experience a Lorentz force, and this will cause a current to flow in the same direction as that due to the induced current due to the perpendicular field lines. And so it really is the total flux of the monopole that must be balanced by the superconducting loop.

Now comes the really amazing part. It was discovered experimentally in 1961 that the flux through a superconducting current loop is quantized and can take on only multiples of some finite minimum flux. Remember that in our experimental apparatus, the passage of a monopole would leave a permanent current flowing in the superconducting loop. Since the flux through the loop depends on this current, which is in turn due to the motion of a charge carrier, it means that the charge carrier may take on only discrete energy states. This should sound familiar—a charge moving in a circular orbit taking on only discrete energy states? The Bohr model of the hydrogen atom gives just such a result. Bohr insisted that an electron moving in a circular orbit of radius R should satisfy

$$2\pi R = n\lambda = n\frac{h}{p},$$

where λ is the de Broglie wavelength of the electron, given by $\lambda = h/p$. Here $p = mv$ is the momentum of the electron and h is Planck's constant.

In our case, we can think of an electron charge carrier as moving in a circular orbit about a uniform magnetic field B. The Lorentz force will result in a centripetal acceleration, and so

$$evB = m\frac{v^2}{R}.$$

Combining these two results gives

$$B\left(\pi R^2\right) = n\frac{h}{2e}.$$

The flux through the loop is just

$$\phi = B\left(\pi R^2\right) = n\frac{h}{2e}.$$

Therefore, the flux through a current loop is quantized. The flux quantum (or fluxon) is very small, with a value of about $2 \cdot 10^{-15}$ T · m^2. This quantization would not normally be observable in a macroscopic system—that is to say, not every electron traveling in a circular orbit must be treated quantum-mechanically. Superconductors, however, are a rare breed of materials where quantum-mechanical behavior becomes apparent on a macroscopic scale. This should be the topic of another article in *Quantum*, but in processes like superconductivity and superfluidity (fluid flow with no viscosity), macroscopic-scale quantum correlations between single electrons are very real and very apparent. The passage of a monopole through a superconducting loop would result in the loop having a flux of only a couple of fluxons. It's remarkable that current technology (pun intended) can accurately measure a flux as low as one fluxon, and this is only possible with the development of superconducting semiconductor devices.

Let's combine our results for the total flux that must be balanced by the current loop due to the passage of a monopole and the flux quantization condition. We finally arrive at

$$\mu_0 q^* e = n\frac{h}{2}.$$

This is the famous Dirac quantization condition for the magnetic charge of a monopole. It's expressed here in S.I. units—something that isn't normal for quantum-mechanical results, which is why it may not look immediately familiar to those who have seen it before expressed in other unit systems. Dirac first derived this result in 1931, and the monopoles it describes are to this day called Dirac monopoles. What it says is truly remarkable. If at least one monopole of any magnetic charge q^* exists in the universe, then electric charge must necessarily be quantized. It's a great mystery of physics why electric charge is, in fact, quantized. There's no reason for this to be so. It just is. The existence of a Dirac monopole would provide a reason for charge quantization, and physicists like knowing the reasons for things.

We can actually estimate a mass for a Dirac monopole from the quantization result. You know that an electron has mass but no measurable size. The mass of an electron is considered to be tied up in its electric field energy. The electric and magnetic field energies obey

$$U_e \propto E^2,$$

$$U_m \propto \frac{B^2}{c^2},$$

where c is the speed of light and E and B are electric and magnetic fields. Since the fields of an electron and a monopole are proportional to the electric and magnetic charge, respectively, we can estimate the mass of a Dirac monopole according to

$$m_m = \frac{q^{*2}}{c^2 e^2} m_e.$$

Using our quantization condition, we find the minimum mass for a Dirac monopole to be

$$m_m \geq \frac{h^2 m_e}{\mu_0 c^2 e^4} \geq 4{,}700 m_e.$$

Thus, a Dirac monopole would have a mass of a little more than twice that of a proton. It is conceivable that a Dirac monopole could be produced in a modern particle accelerator, but this has not turned out to be the most efficient way of searching for them.

The large magnetic charge and relatively small mass of a Dirac monopole means that it's possible for one to be accelerated to very high velocities by galactic magnetic fields and to interact strongly with matter. A high-speed monopole crashing into material would be like a bowling ball thrown through a china shop. So why don't scientists see monopoles? It must be that there are very few of them around, if they exist at all.

The inflationary model of the universe may offer an explanation. It's thought that if primordial monopoles were produced in quantity by the big bang, the rapid inflation of the universe would have reduced the monopole density in a way that's in agreement with the upper level of the density that can be predicted based on the non-observance of monopoles.

Finally, the Dirac theory of a monopole isn't the only contender. Other theories, most notably grand unified theories (GUTs for short) also predict the existence of monopoles. These GUT monopoles would have very different properties from a Dirac monopole—most notably, their mass could be as much as 10^{16} times that of a Dirac monopole. This huge mass would mean that a GUT monopole could only have been produced at the time of the big bang. If this were the case, then it's quite likely that there is some higher meaning for the quantization of both electric and magnetic charge. In physics, mysteries beget mysteries.

Question

We have, up to now, only considered the motion of an electric charge in the presence of a monopole. Now imagine that, as in figure 7, a monopole of pole strength $-q^*$ and mass m is moving with an initial velocity v_0 and finds itself in the vicinity of a stationary and fixed monopole of pole strength $+Q^*$. Let the initial

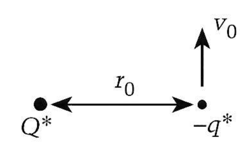


Figure 7

separation between the two monopoles be r_0. Describe the motion of the moving monopole. Are there any limits placed on this motion by the magnitude of the initial velocity?

ANSWER ON PAGE 60

"AIRPLANES IN OZONE" CONTINUED FROM PAGE 25

which case the jet will be neither compressed to the axis nor sprayed aside—this is the so-called "standard mode").

Of course, water vapor constitutes only a fraction of the gaseous mixture in the jet—the jet also contains atmospheric nitrogen (which does not burn), carbon dioxide, and other agents (refer again to figure 1). Let the vapor concentration be 5% = 1/20; we take $\rho_v^a = \rho_A/20 = \rho_0/600 \cong 2 \cdot 10^{-3}$ kg/m³. The density of soot particles (condensation centers) is assumed to be $n_A \sim 10^{13}$ m^{-3}. Now the size of a drop can be obtained from equation (6):

$$a = \sqrt[3]{\frac{3 \cdot 2 \cdot 10^{-3}}{4\pi \cdot 10^{13} \cdot 10^3}} \cong 0.4 \ \mu\text{m}.$$

No doubt you've seen the white trails left by high-flying airplanes—they consist of these droplets. Sometimes they stretch out for hundreds of kilometers. But what does all this have to do with ozone?

Chemisorption of nitrogen oxides

Here's what it has to do with ozone: the droplets can absorb the nitrogen oxides, which turns the water in the drops into a nitric acid solution. The process by which substances are taken in is called absorption (from the Greek word for "devour"), and when accompanied by a chemical reaction, it's known as chemisorption. Such processes are used in chemical plants that produce nitric acid, for instance.

In many respects a jet containing drops isn't at all similar to towering structures used to produce nitric acid. First, the supply of water isn't limited in a land-based industrial plant, so a continuous flow of water absorbs the nitrogen oxides. Second, the pressure of the nitrogen oxides is an order of magnitude greater in an industrial plant than in the jet, and this higher pressure increases the rate at which the oxides are dissolved (true, the temperature in the stratosphere is rather low—$T_{min} \cong 217$ K, which promotes dissolution). We can estimate the maximum concentration of nitric acid in the jet's drops. As the concentration of gaseous nitrogen oxides at the nozzle's outlet is about one hundredth that of water, the concentration of nitric acid will not exceed one percent. And third, the airliner's jet contains other substances in addition to nitrogen oxides.

Thus, generally speaking, all the reactions occurring in the gas phase (fig. 1) can take place in a jet's drops as well (at different rates, of course). These reactions must be taken into account if we are to decide the question whether absorption of substances in the airplane exhaust by water drops increases or decreases the percentage of ozone in the atmosphere.

These drops actually gave rise to the hope that the harmful exhaust gases could be hidden away somewhere. How? Well, the drops could collide with one another and merge to form larger drops, which freeze. They would gain weight and fall. They would thus carry the oxides to the lower layers of the atmosphere, where the oxides would return to the air after the drops evaporate. But, as is well known, the addition of these oxides to the lower regions of the atmosphere actually increases the amount of ozone—which is what happens in the famous London smog.

"Hey, wait!" the thoughtful reader will say. "I don't think it's at all clear that a jet's white trail, with all its drops, actually makes the trip downward." True, it's not clear at all. But this doesn't prove that the drops disappear—we can suppose that, as the drops grow larger, the tail becomes transparent. Recall that fog, which consists of small drops, is opaque, while rain, which contains much more water per unit volume, allows one to see for quite a distance.

Indeed, to estimate the average visibility along some beam of light, let's circumscribe the beam with a cylindrical surface whose radius is equal to the average radius a of the drops. If the drop's center lies within the cylinder, it would block the beam. The number of centers that are located within a cylinder of length L is $N = n\pi a^2 L$, where n is the concentration of the drops and $\pi a^2 L$ is the cylinder's volume. Therefore, the average visibility is

$$\hat{l} = \frac{L}{N} = \frac{1}{\pi a^2 n} = \frac{1}{\pi a^2 \rho_v / m}$$

$$= \frac{\frac{4}{3}\pi a^3 \rho^0}{\pi a^2 \rho_v} \cong a\frac{\rho^0}{\rho_v},$$

where ρ^0 is the density of water. Thus, the larger the drops (the greater their radius a) at a fixed density of water vapor ρ_v, the more transparent the cloud.

But we mustn't forget that the jet of exhaust streaming behind an airliner isn't strictly round, and it doesn't have a strictly horizontal axis. To study in greater detail the possible future fate of drops that have absorbed nitrogen oxides, we need to take another step forward and consider how a flying machine disturbs the atmosphere.

Free vortices and warm streams

Let's begin with a fly. Maybe you've encountered this problem in the course of your studies. A closed box with a fly standing on the bottom rests on a scale—this is state number one. State number two: the fly takes off and hovers somewhere inside the box without touching the walls. Is it possible to distinguish between these two states by looking at the reading on the scale?

The answer is no—the scales will show the same weight in both states. As the fly supports itself in the air, it produces a flow of momentum—that is, a force directed downward—that is exactly equal to the fly's weight.

However, as the fly throws the air mass downward, air must return to the fly's location from above. The air begins to circulate inside the box, as shown qualitatively in figure 4a by the blue and red lines. It looks very much like the air flow around a hovering helicopter.

Now let the fly (or helicopter) not

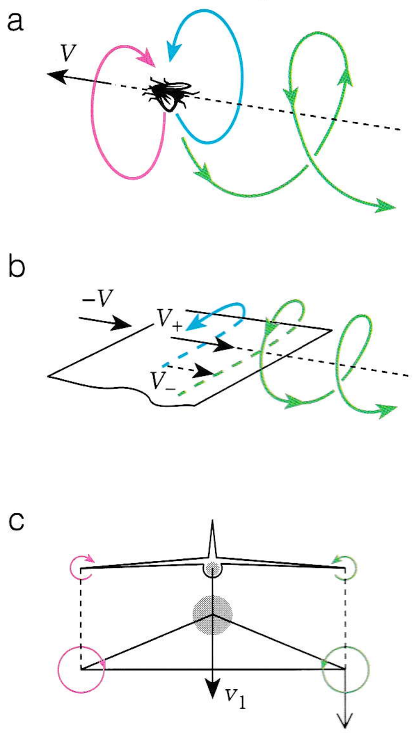


Figure 4

just hang in the air but move forward with a constant velocity V. Then in the reference frame attached to the center of mass, air is blown backward with velocity $u_\infty = -V$. Drawing the flow lines, we can see that two vortices with almost parallel axes arise in the plane perpendicular to the direction of flight. (So as not to overload figure 4a with details, we have shown only one vortex, which is located to the right of the fly.)

An airplane produces almost the same air flow as a fly or helicopter, but it doesn't combine the lifting and thrust functions: the wings create the lift and, unlike the fly, the thrust is produced not by the wings but by a propulsion device (propeller or jet engine).

We can explain vortex generation at the wing tip qualitatively in another way. As the air flowing around a wing produces lift, its pressure on the bottom surface of the wing is greater than that acting on its upper surface: $p_- > p_+$. According to Bernoulli's law, the air velocity above the wing is greater than that below: $V_+ > V_-$ (fig. 4b).

In addition, at the wing tip the air tries to flow upward due to this pressure difference. If we draw the trajectory of an air particle, we get a helix that runs off from the wing tip—this is a so-called trailing (wing tip) vortex. Figure 4c shows the rear view of a plane with two trailing vortices. Thus, the emergence of these two vortices is directly connected with the generation of lift by a moving wing. (This explanation is suitable only for subsonic flow around a wing, but it's quite sufficient for our purposes here.)

There is yet another force that lifts the jet: the buoyant force F_b. The jet is warmer than its surroundings and looks like a dirigible filled with a gas (air) of somewhat less density (indeed, according to Clapeyron's law, $\rho \sim 1/T$ at constant pressure). To estimate the upward velocity due to buoyancy, we change the continuous bell-shaped radial distribution of temperature and density in the jet for a stepwise distribution. These values are assumed to be constant and equal to T_m and ρ_m inside a cylinder of a certain radius r_e and equal to T_∞ and ρ_∞ in the undisturbed atmosphere outside this cylinder (see figure 5).

Thus, we'll imagine that a part of the exhaust jet of length Δx is placed in a cylindrical cellophane package with insulated walls and that this package (a sort of dirigible) is lifted in the cool atmosphere by the buoyant force

$$\Delta F_A = (\rho_\infty - \rho_m) g\pi r_e^2 \Delta x.$$

Here $\pi r_e^2 \Delta x$ is the volume of that portion of the exhaust jet of length Δx where the jet can be considered approximately a cylinder, although we know that a real stream of exhaust expands slightly. However, since $\rho_\infty = p_\infty M/RT_\infty$, and $\rho_m = p_\infty M/RT_m$, we get

$$\Delta F_b = \rho_\infty g\pi r_e^2 \Delta x \frac{T_m - T_\infty}{T_m}.$$

Thus, the buoyant force is proportional to the square of the exhaust jet's cross-sectional radius and to the temperature difference between the jet and the atmosphere.

Let this force result in movement of the jet upward with a constant velocity $v_\uparrow$. Then a counterbalancing drag force will arise that is (as has been said many a time) proportional to the density of air flowing around the object, the square of the velocity, and the cross-sectional area $S_\perp = 2r_e\Delta x$ perpendicular to the vector $\mathbf{v}_\uparrow$:

$$\Delta F_d = \rho_\infty v_\uparrow^2 S_\perp = \rho_\infty v_\uparrow^2 2r_e \Delta x.$$

Setting F_b equal to F_d, and taking into account that $T_m(x) \sim 1/x$ and $r_e \sim \sqrt{x}$, we get

$$v_\uparrow^2 \sim \frac{1}{r_e} \sim \frac{1}{\sqrt{x}} \Rightarrow v_\uparrow \sim x^{-1/4}.$$

If the exhaust jet "escapes" due to this buoyant force from the influence of the vortices, its further motion will be determined basically by the buoyant force. We can write this as

$$\frac{dy^j}{dx} = \frac{v_\uparrow}{u_\infty} \sim \frac{x^{-1/4}}{u_\infty},$$

which after a bit of integration gives us $y^j \sim x^{3/4}$. The resulting axis of the jet will look like the solid curve shown qualitatively in figure 6 (on the next page).

Now it's time to look back at the mess I've made here and, as they say in polite society, "beg your pardon" for everything I've neglected so far.

First, the vortex thread itself, like the exhaust jet, will

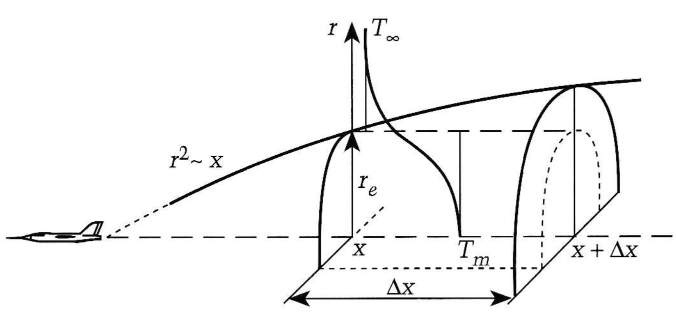


Figure 5

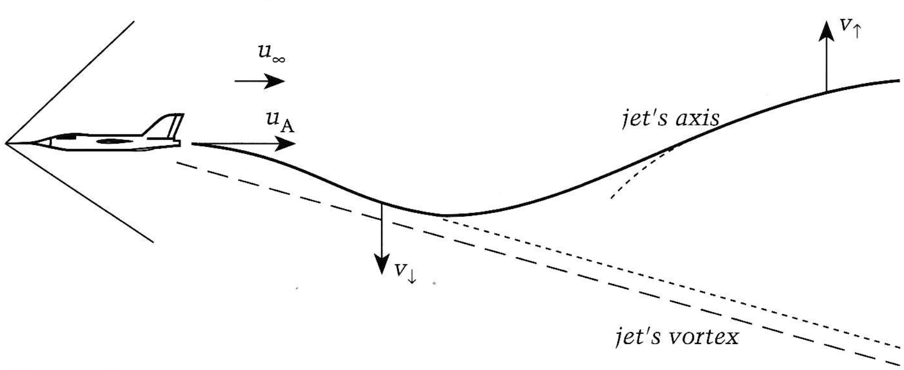


Figure 6

"diffuse" into the surroundings because of the friction of layers against one another—they rotate with different circular velocities about the axis. This means that a vortex also diffuses radially, so the radial dependence of its circular velocity becomes more complicated than equation (2).

Second, the field of vortex velocities will affect not only the jet's axes but their periphery as well, because a jet isn't a line—it has a certain cross-sectional size. The individual elements in the exhaust jet will be "gone with the wind" each in its own way depending on its location relative to the axis.

Third, after being slowed by the atmosphere, the jet may disintegrate into individual pieces that will flow upward as individual clumps, not as a cylinder.

I also neglected many other things that may affect the evolution of an individual drop, the possibility that drops might fuse together, and the rate of precipitation of the entire trail of drops together with the absorbed nitrogen oxides.

But then, maybe there are those among our readers who can condense these thoughts into equations.

Conclusions

So—will the stratospheric airliners of the future be harmful for the atmospheric ozone? Perhaps, yes—but almost everything that human beings do is harmful in one way or another, including breathing. Indeed, when we breathe, we produce carbon dioxide (animals do the same)—plants ingest it and return oxygen to the atmosphere, and this is an equilibrium process. So, before drawing any conclusions, it's important to compare the expected harm with what goes on in nature.

It's known that up to the present time volcanic eruptions have put more nitrogen oxides into the atmosphere than all the airplanes in the world. Even in the future, when the contribution of harmful substances from commercial airliners will be comparable to that from natural sources, it will nevertheless be difficult to determine by means of measurements just who is doing the polluting. So both now and in the future the important instrument in determining the contribution of aviation to this process will be theoretical estimates based on physical, chemical, and mathematical models of the sort considered here.

Further studies may show that this ozone problem isn't as tragic as it was presented by some nervous journalists who were frightened by the "ozone hole" over Antarctica. Some scientists consider that there is no problem at all: the solar radiation, they say, is absorbed by the entire atmosphere and not just by the ozone. And folk wisdom points to the value to crops of a weak solution of nitric acid—stated in different terms, naturally: "If there are thunderstorms in the spring or early summer, the harvest will be abundant." Why? Well, plants need nitrogen, and even though the atmosphere consists mostly of nitrogen, they can't take it directly from the air—it has to be combined with other elements. That's where the thunderstorms come in. When lightning discharges, the atmospheric nitrogen forms chemical compounds that are dissolved in the raindrops to turn them into weak solutions of nitric acid. The nitric acid solution reacts with the minerals in the soil and frees up certain substances—phosphorus and potassium, among others. These elements are necessary for the rapid growth of plants.

Sounds familiar, doesn't it? This is much like what happens in the exhaust jet from an airplane. So we can't exclude the possibility that, in the third millennium A.D., a new saying will appear: "If a supersonic airliner flies over your garden in the spring, expect a large pumpkin in the autumn."

"THE FIRST PHOTON" CONTINUED FROM PAGE 37

this is the value of the circulating current. Because the superconducting switch has $r = 0$, I_c cannot change and I_s must increase by 20 A. In other words, I_s changes from –20 A to zero.

Because there is no current through the superconducting switch, we can now change it to the normal state $r = r_n$. We now gradually reduce the total current I to zero while keeping $I_s < 0.5$ A. Because

$$V_s < (0.5\ \text{A})(5\ \Omega) = 2.5\ \text{V},$$

the current through the inductor must obey

$$-\frac{\Delta I_c}{\Delta t} = \frac{V_s}{L} < \frac{2.5\ \text{V}}{10\ \text{H}} = 0.25\ \text{A/s}.$$

Therefore, the current must be dropped to zero over a minimum of 80 s. These conditions are satisfied in figure 4 (on page 37) with $\Delta I/\Delta t \cong 0.1$ A/s.

As a final step, we can return the superconducting switch to the superconducting state and open the power switch K.

"JESSE JAMES" CONTINUED FROM PAGE 31

corresponds to the fact that "a solution of the heat equation is determined by two boundary conditions at the ends of the rod."

There are, however, variations in this problem that lead to different kinds of equilibrium solutions. For example, at the same time as the eight individuals are obeying Jesse's rule, they may also be receiving income. Assuming $f(x)$ denotes the income of individual x at time $t = 1, 2, \ldots$ (income received just *after*, say, a monthly transfer of funds has occurred), we find that

$$u(x, t+1) = u(x, t) + 0.1[u(x, t-1) \\ - 2u(x, t) + u(x, t+1)] + f(x),$$

which corresponds to the partial differential equation $u_t = 0.1u_{xx} + f(x)$. Here the steady-state solution will again be determined by $u(0)$ and $u(9)$, but it would in general *not* be linear.

Another possibility is for Jesse to impose different transfer rules on the individuals in a row. This corresponds to a heat flow in a nonhomogeneous rod whose conductivity varies with x according to a function $c(x)$. It leads to the differential equation $u_t = (c(x)u_x)_x$ whose steady-state solution would in general also not be linear.

Such problems can, however, be programmed in a spreadsheet to determine the nature of an equilibrium solution, if one exists.

A two-dimensional model

With these perspectives in mind, we can see another approach to Jesse's problem of equalizing the wealth. If the number of townsfolk is large, we might choose to arrange them in a rectangular grid rather than a straight line. Displaying the assets of 24 townsfolk could then lead to a grid such as figure 9.

A reasonable rule here is to require each individual to compare assets with *four* neighbors—that is, the individuals above, below, to the right, and to the left. This comparison would be followed by a transfer of assets of 10% of each difference from richer to poorer.

In figure 9, the individual with \$70 would receive a total of \$11 from two horizontal neighbors while receiving \$13 from two vertical neighbors. Analogy suggests that such a process should correspond to two-dimensional heat flow (for example, in a thin metal plate).

The reader is invited to explore this phenomenon in the context of the problems that follow.

Problems

1. Show that in one dimension, an individual of "zero cuppedness" is one whose assets equal the *average* of the assets of the two immediate neighbors.

2. As an alternative to Jesse's rule for basing one-dimensional transfer of funds on differences, consider the following: **At $t = 1, 2, 3, \ldots$, each individual passes 10% of his or her assets to each neighbor**. How does this compare with Jesse's rule?

3. If in a rectangular array the wealth of an individual is denoted by $u(x, y)$, what is an "index of cuppedness" corresponding to the "reasonable rule" cited above—that is, one in which an individual located at (x, y) compares assets with four neighbors located at $(x-1, y)$, $(x, y+1)$, $(x+1, y)$, and $(x, y-1)$?

4. Rather than defining "zero cuppedness" for two-dimensional arrays, we'll say that an individual whose assets remain constant is "at harmony with her or his neighbors," or simply "harmonic." Can you find a harmonic individual among the 24 townsfolk arrayed in figure 9?

5. Show that a harmonic individual is one whose assets equal the average of those of his or her *four* immediate neighbors.

\$50	\$100	\$80	\$140	\$120	\$40
\$90	\$200	\$160	\$70	\$90	\$20
\$40	\$30	\$130	\$130	\$130	\$80
\$50	\$120	\$200	\$20	\$50	\$80

Figure 9

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

Student inventors show their stuff

Duracell/NSTA Scholarship Competition awards $90,000 in savings bonds

WOULDN'T LECTURES BE more interesting if students could record what the teacher said and draw diagrams of the material at the same time? And wouldn't it be helpful for drivers to know when another vehicle is in their blind spots? How about having a small robot to sort your recyclables? Or a cane that gets longer for going down stairs? What about a bike taillight that signals when a biker is braking, and a stop sign for school crossing guards that automatically lights up the corner stop sign to warn drivers on dark mornings?

These handy devices do exist—but currently each is one of a kind. They were invented by six high school students—the top winners in the Duracell/NSTA Scholarship Competition, which awarded $90,000 in prize money.

High-tech note-taker takes first place

Ara Knaian of Newton, Massachusetts, designed and produced Note-Taker, a portable computer with a built-in microcassette tape recorder. While recording what is being said, this device also lets the user draw charts or diagrams. "I originally conceived of it as a tool for reporters," the Newton North High School senior said. "Its main application, education, occurred to me later." Knaian says the combined analog/digital device "allows the user to simultaneously take notes *and* think about what is being said." For his ingenuity, Knaian won first place and a $20,000 US savings bond. Knaian has been accepted at MIT, where he will major in physics or electrical engineering. His sponsor on the Duracell project was Richard Duffy, a physics teacher at Newton North High School.

Mara Lynn Carey of Palm Beach, Florida, was the second-place winner for The Determinator, a robot that sorts recyclables. "The Determinator began by researching digital logic," the Martin County High School senior said, and it resulted in an invention that sorts plastic, glass, steel, and aluminum. Carey received a $10,000 savings bond for her efforts. She has been accepted at Georgia Tech, where she plans to study electrical engineering.

Other second-place winners who received $10,000 savings bonds are Benjamin Ihas of Gainesville, Florida; Scott Jantzen of Shoreham, New York; Mohammed Omer Khan

Judges in the 13th annual Duracell/NSTA Scholarship Competition (l–r): Mary Nalbandian (seated), Arthur Eisenkraft, Huk Yuk Cheh, Morris Shamos. The first- and second-place winning devices are displayed on the table—back row (l–r): The Determinator, Advanced Bike Tail Light, S4–Super Scintillating Stop Sign; front row (l–r): Note-Taker, Easy Step, Ultrasonic Eye.

of Gaithersburg, Maryland; and James Zajkowski of Harrington Park, New Jersey.

Ben Ihas improved upon existing taillights by inventing the Advanced Bike Tail Light, which blinks during pedaling, then brightens and burns steadily when the rear brakes are on. "A plus of this light is that since the circuit has been kept simple and the most expensive parts are a transistor and a miniature transformer, it would be easy and inexpensive to mass-produce," the Eastside High School sophomore noted.

Scott Jantzen came up with the idea of building the Easy Step after he broke his leg. "When on crutches, sometimes I felt like I was going to fall down the stairs because of the instability," he said. Easy Step is an adjustable cane that extends and contracts and is especially useful for climbing stairs. The Shoreham–Wading River High School junior keeps busy with soccer, track, and lacrosse.

Mohammed Khan created the Ultrasonic Eye after personally experiencing a need for a device that detects and identifies objects in a driver's blind spot. "I would like the Ultrasonic Eye to be placed in every truck in America," declared the Watkins Mill High School senior. "Ironically, I came up with this device when I was searching for a ranging system to be part of a robot that I was making. The robot did not know how to avoid people in its path. While brainstorming how to solve this problem, I came up with reflecting sound to determine the distance of objects ahead of the robot. That's when it dawned on me that it would be nice if a car had such a device."

James Zajkowski said his idea for the S4–Super Scintillating Stop Sign "originated when I, a fledgling driver, noticed how difficult it was to see school crossing guards during dawn's early light or in inclement weather." The Northern Valley Regional High School junior added, "If S4 can stop even one accident, I believe all crossing guards should be equipped with one." By illuminating not only itself but also a second stationary stop sign when it is raised, the S4 heightens awareness around crosswalks. Moreover, the device has several applications in road repair.

Awards ceremony in Philadelphia

The first- and second-place winners, their parents, and their science teachers were guests of Duracell at an awards ceremony in Philadelphia on March 23, moderated by NASA Teacher in Space designee Barbara Morgan. The winners demonstrated their devices for a luncheon audience and exhibited them for thousands of science teachers at the National Science Teachers Association's 43rd annual convention.

Now in its thirteenth year, the Duracell/NSTA Scholarship Competition also named ten third-place winners, who each received a $1,000 savings bond; 25 fourth-place winners, who received a $200 savings bond; and 59 finalists, who were eligible for a $100 savings bond.

Judging began in late January, when a panel reviewed 781 entries and selected 100 for final judging at Duracell headquarters in Bethel, Connecticut, at the end of February. Four judges examined and operated each device during final judging, looking for inventiveness, precision, and energy efficiency.

Every student who entered the competition received a sports water bottle from Duracell and a certificate of participation from NSTA. Many of the finalists will have their devices displayed at conventions and exhibits around the country.

Administered by NSTA, the Duracell/NSTA Scholarship Competition has awarded more than $500,000 in scholarships, savings bonds, and cash awards to 65 students over the past thirteen years.

Bulletin Board

Russian-American student exchange

For the fifth consecutive year, the American Regions Mathematics League (ARML) is organizing an exchange program for American and Russian high school students interested in mathematics. Partially funded by the United States Information Agency, the trip will take 25 American high school students to Moscow and St. Petersburg this summer for four weeks of mathematical studies together with cultural experiences and sightseeing. Students will live with Russian families, and the language of instruction will be English. The dates of the program will be July 13 through August 10, and the cost for each student will be $1,400, including airfare from Washington, D.C.

For more information, write to Eric Walstein, Montgomery Blair Magnet Program, 313 Wayne Avenue, Silver Spring MD 20910 (e-mail: Walstein@vax.mbhs.edu).

Quantum WWWinner

Visitors to *Quantum*'s World Wide Web home page had an advance opportunity to solve one of the brainteasers that appear in this issue. It's impossible to know how many declined to take part, but Matt Nehring, a graduate student in physics at the University of Colorado, took up the challenge—and electronically submitted a correct answer to brainteaser B144. For his efforts Matt will receive a *Quantum* button and a copy of this issue, as well as a mention at our Web site.

All our readers are invited to try and solve the next cyberteaser. Our home page is at http://www.nsta.org/quantum.

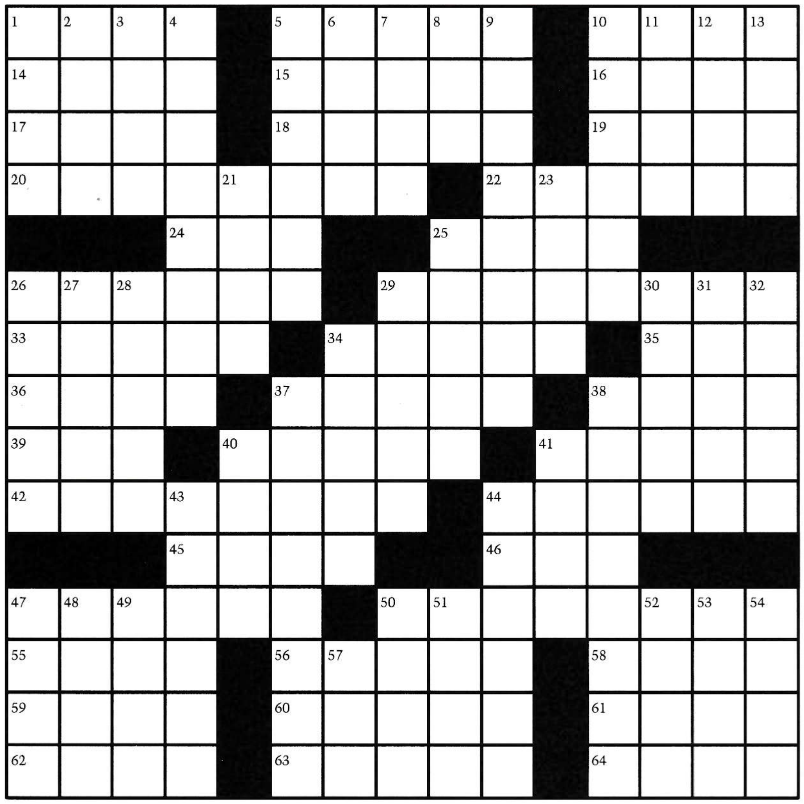
What's happening?

Help us fill Happenings and the Bulletin Board with short news items, firsthand reports, and announcements of upcoming events. Send the information to

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

Criss × cross science

by David R. Martin



Across

1 Fruit rot
5 Titles of reverence
10 English chemist Frederick Augustus ___ (1827–1902)
14 Rim
15 Type of intestine
16 Relocate
17 ___ wax (ozocerite)
18 More arid
19 "S" shaped curve
20 CHX_3
22 Atlantic and Pacific
24 Sibling: abbr.
25 Type of foundation
26 Of the sea
29 ___ vertebrae
33 Organic compound
34 Attack
35 Expert
36 Trig. function
37 "The ___ bird catches . . ."
38 60 coulombs: abbr.
39 Printer's measures
40 Line of talk
41 Sport shoe part
42 Oolite
44 Hypothetical massless particles
45 ___ in line
46 Edge
47 ___ gage (measures pressure)
50 Near the apex
55 In a violent rage
56 Philistine god
58 Hawaiian goose
59 ___ oil (juniper tar oil)
60 *Burseraceae* tree resin
61 Shady mountain side
62 City in Iowa
63 Frozen raindrops
64 Order to dogs

Down

1 Jacob's wife
2 Icelandic literary work
3 ___ fiber (from the gebang palm)
4 Petroleum product
5 Class of sugars
6 Art historian Alfred ___
7 Edge
8 Epoch
9 Science of blood
10 Rhizopod protozoan
11 Town near L. Albert
12 Type of parity
13 Dregs
21 ___-structure constant
23 Concern
25 Nut container
26 Microwave source
27 Type of acid
28 Wash lightly
29 Scottish island
30 Supporting role
31 Expert: suff.
32 Pennies
34 Holy person
37 Cyclic ethers
38 Element 13
40 British gun
41 Illegal football block
43 *Serpentes* members
44 1967 Physiol. Nobelist Ragnar ___
47 South Am. rodent
48 Muslim religious leader
49 Was carried
50 Novelist James ___
51 Arch
52 Philippine island
53 Trivia books
54 ___ Walesa
57 Everyone

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MARCH/APRIL PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M141

Our operation takes a number $x = 10a + b$ $(0 \le b < 10)$ into $y = a + 4b$. Since $x = 10y - 39b$, the numbers x and y simultaneously are or are not divisible by 13 (since $39 = 3 \cdot 13$). Therefore, if the sequence in question contains $1001 = 13 \cdot 77$, all its terms must be divisible by 13. The only prime divisible by 13 is 13. But it's immediately verifiable that 13 is a stable point in our operation $(1 + 4 \cdot 3 = 13)$, so it cannot precede 1001; on the other hand, all "descendants" of 1001 are also easy to find: $1001 \to 104 \to 26 \to 26 \to \ldots$, and the number 13 is not among them either.

M142

Let $P(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$, $a_n > 0$, $n \ge 2$. Put $Q(x) = P(x) + x$. Then, for each integer k between 1 and n, $(Q(x))^k = (P(x) + x)^k = B_k(x)P(x) + x^k$, where $B_k(x)$ is a certain polynomial (it can be written out using the binomial formula, but we don't need it at all).

So,

$$\begin{aligned} P(Q(x)) &= (a_nB_n(x)P(x) + a_nx^n) + \\ &\quad (a_{n-1}B_{n-1}(x)P(x) + a_{n-1}x^{n-1}) \\ &\quad + \ldots + a_0 \\ &= R(x) \cdot P(x) + \\ &\quad (a_nx^n + a_{n-1}x^{n-1} + \ldots + a_0) \\ &= (R(x) + 1)P(x), \end{aligned}$$

where $R(x)$ is a polynomial of degree no less than 2, and we're done.

M143

(a) The answer is $2/(a^{-1} + b^{-1} + c^{-1})$, or, more symmetrically,

$$\frac{2}{r} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c},$$

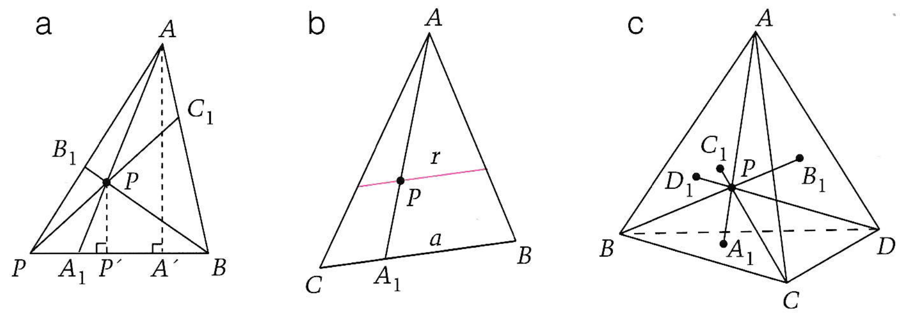


Figure 1

where r is the unknown length. We'll show this using a method that can be applied in the three-dimensional case as well.

Let's first prove that *if three segments joining the vertices A, B, and C of a triangle to points A_1, B_1, and C_1 on their opposite sides meet at point P (fig. 1a), then*

$$\frac{PA_1}{AA_1} + \frac{PB_1}{BB_1} + \frac{PC_1}{CC_1} = 1. \qquad (1)$$

Let PP' and AA' be the altitudes of triangles PBC and ABC, respectively. From similar triangles A_1PP' and A_1AA', we see that the ratio PA_1/AA_1 is equal to the ratio PP'/AA'. This ratio is in turn equal to the ratio of the area of PBC to that of ABC (since these triangles have a common base).

For similar reasons, the left side of equation (1) can be rewritten (using absolute-value signs to denote area) as

$$\frac{|PBC|}{|ABC|} + \frac{|APC|}{|ABC|} + \frac{|ABP|}{|ABC|} = \frac{|ABC|}{|ABC|} = 1,$$

completing the proof.

Now, if P is the point considered in the problem, then by the similarity of the triangles cut off by the lines through P and the original triangle (fig. 1b) we get $AP/AA_1 = r/a$, so $PA_1/AA_1 = (AA_1 - AP)/AA_1 = 1 - r/a$, and in the same way

$$\frac{PB_1}{BB_1} = 1 - \frac{r}{b}, \quad \frac{PC_1}{CC_1} = 1 - \frac{r}{c}.$$

Together with equation (1) this yields the equation

$$3 - r\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c}\right) = 1$$

and the expression for r given above.

(b) The solution for the three-dimensional case repeats the previous one almost without any changes. Four segments in a tetrahedron $ABCD$ meeting at a point P (fig. 1c) always satisfy

$$\frac{PA_1}{AA_1} + \frac{PB_1}{BB_1} + \frac{PC_1}{CC_1} + \frac{PD_1}{DD_1} = 1. \qquad (2)$$

The only difference in the proof is that we must replace the areas of triangles by the volumes of tetrahedrons $(PA_1/AA_1 = \mathrm{vol}(PBCD)/\mathrm{vol}(ABCD)$, and so on).

Then, if x is the unknown area and A is the vertex opposite the face of the area a, then $x/a = (AP/AA_1)^2$, for instance, because the ratio of the areas of similar figures is the square of their ratio of similarity, and the planes considered in the problem cut off from the given tetrahedron similar tetrahedrons with the ratios of similarity AP/AA_1, BP/BB_1, and so on.

Expressing the left side of equation (2) in terms of the ratios x/a, x/b, x/c, x/d, we arrive at the equation

$$4-\left[\left(\frac{x}{a}\right)^{\frac{1}{2}}+\left(\frac{x}{b}\right)^{\frac{1}{2}}+\left(\frac{x}{c}\right)^{\frac{1}{2}}+\left(\frac{x}{d}\right)^{\frac{1}{2}}\right]=1$$

which yields the answer

$$x=\left(\frac{3}{a^{-\frac{1}{2}}+b^{-\frac{1}{2}}+c^{-\frac{1}{2}}+d^{-\frac{1}{2}}}\right)^2.$$

(V. Dubrovsky)

M144

Let α be the angle between the larger sides of A and B (fig. 2). It's

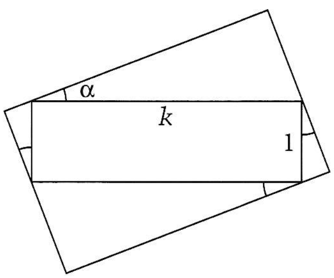


Figure 2

not difficult to see that $\alpha \le \pi/4$. Let the sides of rectangle B be equal to k and 1 (where $k > 1$), so that the eccentricity of B is equal to k. Then the sides of rectangle A are equal to $k \cos \alpha + \sin \alpha$ and $\cos \alpha + k \sin \alpha$. Note that the first is not less than the second, since $\alpha \le \pi/4$, so $\cos \alpha \ge \sin \alpha$.

Now $k^2 \sin \alpha \ge \sin \alpha$ (since $k > 1$). It follows that

$$\frac{k\cos\alpha+\sin\alpha}{\cos\alpha+k\sin\alpha}\le k,$$

which says that the eccentricity of A is not greater than that of B.

M145

Suppose R, W, and B are the numbers of red, white, and blue squares, respectively. We will prove that the result for R, the number of red squares. First we note that $R \le 3W$. To prove this we take any white

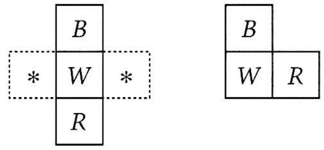


Figure 3

square and put a check mark on each red square bordering it. Then we take any other white square and put a check on any red square bordering it that has not yet been checked. We continue this until there are no more white squares left. Since each white square borders on at least one blue one, at each step we have checked no more than three red squares. But this means that

$$R \le 3W. \qquad (1)$$

In the same way we can show that $W \le 3B$ and $B \le 3R$.

Take any red square x_1 and construct a chain of three squares, consisting of a white square x_2 bordering on square x_1, then a blue square x_3 bordering on x_2. Wherever we can, we construct a "corner chain" (see figure 3). Note that we can construct a corner chain so long as one of the squares marked with an asterisk in figure 3 is blue.

In each "straight" chain, we put a check mark on the white square, and in each corner chain we put a check mark on each red square. We construct a chain, and check a square, for each red square (the chains may overlap).

Now the number of checked white squares is equal to the number of straight chains. Also, each blue square that is not part of a corner chain is checked not more than four times (it cannot belong to more than four corner chains). It follows that

$$R \le W + 4B. \qquad (2)$$

Similarly, $W \le B + 4R$ and $B \le R + 4W$. Adding inequalities (1) and (2), we find that $2R \le 4(W + B)$, and finally that $n^2 = R + W + B \le 11R$, which proves part (b).

It is left to the reader to show that these bounds on R are the best possible—that is, that 2/3 cannot be replaced by a smaller number or 1/11 by a larger number.

Physics

P141

We take advantage of the fact that the initial potential energy of the body on the surface of the upper liquid is expended during its motion to overcome the force of resistance. Therefore,

$$mg(h_1 + h_2) = W_1 + W_2,$$

where m is the body's mass, g is the acceleration due to gravity, W_1 is the work performed by resistance in the upper liquid, and W_2 is the analogous value for the lower liquid. Since the body is streamlined, the resistance is the buoyant force: $F_1 = \rho_1 Vg$ in the upper liquid and $F_2 = \rho_2 Vg$ in the lower liquid (where V is the body's volume). Thus,

$$W_1 = \rho_1 Vgh_1$$

and

$$W_2 = \rho_2 Vgh_2.$$

Substituting these values for W_1 and W_2 in the first equation above and taking into account that $m = \rho V$ (where ρ is the body's density), we get $\rho(h_1 + h_2) = \rho_1 h_1 + \rho_2 h_2$, from which we get

$$\rho=\frac{\rho_1 h_1+\rho_2 h_2}{h_1+h_2}.$$

However, most readers of *Kvant* (*Quantum*'s sister magazine, in which this problem first appeared) solved the problem in a more complicated way.

Two forces act on an object moving in a fluid: gravity ($mg = \rho Vg$) and buoyancy ($F_1 = \rho_1 Vg$ and $F_2 = \rho_2 Vg$, respectively). Since the object is streamlined, the resistance in each liquid does not change as it sinks. This means that the object moves with a uniform acceleration. Let's write the equation for this motion. For the upper liquid, $mg - F_1 = ma_1$ (the acceleration is directed downward); and for the lower liquid, $F_2 - mg = ma_2$ (the acceleration is directed upward). It follows from these equations that in the upper liquid (density ρ_1), the body sinks with an acceleration

$$a_1 = g\frac{\rho - \rho_1}{\rho},$$

while in the lower liquid it sinks with an acceleration

$$a_2 = g\frac{\rho_2 - \rho}{\rho}.$$

After passing through the upper liquid, the object acquires a velocity $V_0 = a_1 t_1$, where t_1 is the time it takes to drop through that layer. This value can be found from the kinematic equation

$$h_1 = \frac{a_1 t_1^2}{2}.$$

Therefore,

$$V_0 = \sqrt{2a_1 h_1}.$$

This value is in turn the initial velocity of the object in the lower liquid. Since the final velocity is zero, $V_0 - at_2 = 0$, where t_2 is the time it takes to drop through the lower layer. It follows that

$$t_2 = \frac{V_0}{a_2}.$$

In this amount of time the object will cover the distance h_2 in the lower liquid. Thus,

$$h_2 = V_0 t_2 - \frac{a_2 t_2^2}{2}.$$

Inserting the expressions for V_0, a_2, and t_2 in the last equations yields

$$h_2 = \frac{V_0^2}{a_2} - \frac{1}{2}\frac{V_0^2}{a_2} = \frac{1}{2}\frac{V_0^2}{a_2} = \frac{a_1 h_1}{a_2},$$

from which we get

$$\frac{h_2}{h_1} = \frac{a_1}{a_2} = \frac{\rho - \rho_1}{\rho_2 - \rho}.$$

Solving the last equation, we arrive at the same result as that above:

$$\rho = \frac{\rho_1 h_1 + \rho_2 h_2}{h_1 + h_2}.$$

Note that accelerated motion of a body (even in an ideal liquid) generally involves something called "apparent mass," which diminishes the acceleration. The effect is negligible, however, when $\rho \gg \rho_1, \rho_2$.

P142

The ocean surface is everywhere perpendicular to the force of gravity, which is directed radially to the Earth's center. The existence of the cavity beneath the ocean floor causes the surface to be curved, because the total gravitational force acting on water due to the cavity and due to the rest of the Earth's mass is, generally speaking, not directed along the planet's radii.

In the reasoning below we'll use an analogy between the electrostatic field of a point charge Q and the gravitational field of a point mass M. (Recall that the gravitational field of a sphere of mass M is equivalent to that of a point mass M placed at the center of the sphere.)

The force acting on a test charge q from the charge Q and the force attracting the test mass m by mass M are correspondingly

$$F_e = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r^2}$$

and

$$F_g = G\frac{Mm}{r^2},$$

where r is the distance between Q and q (or M and m). These formulas are quite similar: $F_e \sim 1/r^2$ and $F_g \sim 1/r^2$, $F_e \sim q$, and $F_g \sim m$—that is, the dependence of these forces on distance is the same, as is their dependence on the test charge and test mass. By this analogy, we introduce the gravitational potential ϕ_g for the gravitational field of a point mass M, which is equal to the potential energy of a unit mass in a field of mass M. The expression for ϕ_g is analogous to that for the potential $\phi_e = (1/4\pi\varepsilon_0)(Q/r)$ of the electric field of a point charge Q:

$$\phi_g = -G\frac{M}{r}.$$

The minus sign in this formula reflects the fundamental difference between F_e and F_g: the bodies are always attracted by gravitational forces. Since the potential energy is taken to be zero at infinity (that is, at very large r), it is negative everywhere due to the attraction in the gravitational field.

The force of gravity is everywhere perpendicular to the water's surface, so the work performed by this force when a particle is shifted along the water's surface is zero, which means that the potential energy is constant at any point on the surface. Therefore, the surface of the ocean (which is to say, the surface of the Earth) is an equipotential surface (analogous to an equipotential surface in an electrostatic field).

Let's determine the potential of the Earth's gravitational field at point A far from the cavity (on the opposite side of the Earth) and at point B right above the cavity (fig. 4). For point A the field distortion caused by the cavity can be neglected, so the potential is

$$\phi_A = -G\frac{M}{R},$$

where R is the Earth's radius and $M = (4/3)\pi R^3 \rho_r$ is its mass. To find the potential at point B, we apply the superposition principle. Figure 5 (on the next page) shows the superposition of the gravitational field of the Earth and the cavity. It's clear from the figure that the potential at point B is composed of the Earth's gravitational field (with no cavity) and that of the spherical mass whose density is $\rho_w - \rho_r$,—that is, the point mass $\mu = (4/3)\pi r^3(\rho_w - \rho_r)$, where r is the radius of the cavity and ρ_w is the density of water. Thus,

$$\phi_B = -G\frac{M}{R-\delta} + \left(-G\frac{\mu}{r+h-\delta}\right)$$

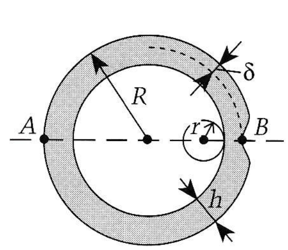


Figure 4

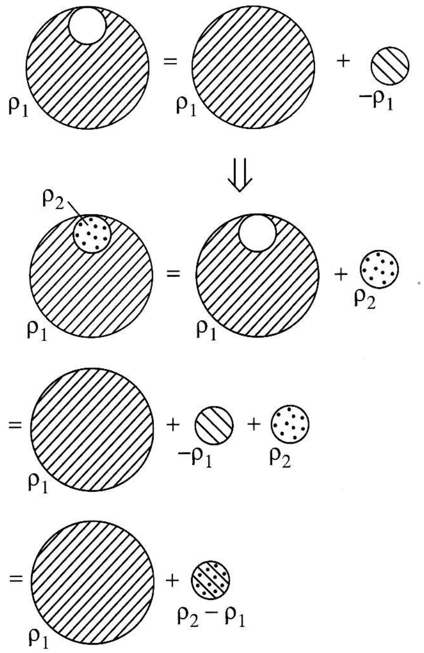


Figure 5

(see figure 5). As the ocean's surface is an equipotential, $\phi_A = \phi_B$, or

$$\frac{M}{R} = \frac{M}{R-\delta} + \frac{\mu}{r+h-\delta}.$$

Rearrangement of this equation gives us

$$-M\frac{\delta}{R(R-\delta)} = \frac{\mu}{r+h-\delta}.$$

Therefore,

$$-\frac{M}{\mu}\left[(r+h)\delta - \delta^2\right] = R(R-\delta).$$

Neglecting δ^2 on the left-hand side and δ on the right-hand side yields

$$-\frac{M}{\mu}(r+h)\delta \cong R^2,$$

or, substituting

$$-\frac{M}{\mu} = -\frac{R^3}{r^3}\frac{\rho_r}{\rho_w - \rho_r},$$

we obtain

$$\frac{r+h}{r^3}\frac{R\delta\rho_r}{\rho_r - \rho_w} \cong 1.$$

Therefore,

$$\frac{r^3}{r+h} \cong \frac{\rho_r}{\rho_r - \rho_w}R\delta$$

$$= \frac{3}{2}\cdot 6.4\cdot 10^6 \cdot 25\ \text{m}^2$$

$$= 240\ \text{km}^2.$$

Here we have used the numerical values $R = 6.4\cdot 10^6$ m, $\rho_r = 3\cdot 10^3$ kg/m^3, $\rho_w = 1\cdot 10^3$ kg/m^3, and $\delta = 25$ m. We won't solve this equation for r—we'll just estimate its value. Remembering that $h = 6$ km from the statement of the problem, we can get a first estimate by taking the square root of $240 \cong 15.5$ km. Using this value in the left-hand side yields a value less than 240 km^2. Therefore, we choose a larger value until we find that r is about 18 km.

P143

When heating a gas at constant volume, the energy is spent only on increasing the gas's internal energy, while heating it at constant pressure requires energy to perform work as well. Let's write down the law of conservation energy in both cases:

$$mc_v\Delta T = \Delta U, \qquad (1)$$

$$mc_p\Delta T = \Delta U + W, \qquad (2)$$

where c_p is the specific heat capacity of the gas at constant pressure (that is, the amount of energy necessary to raise the temperature of 1 kg of gas by 1°C at constant pressure), c_v is the specific heat capacity at constant volume, ΔT is the change in temperature, ΔU is the change in the internal energy of the gas, m is the mass of the gas, and $W = P\Delta V$ is the work performed during gas expansion (ΔV is the change in volume and P is the pressure).

Since an increase in temperature by the same number of degrees at either constant pressure or constant volume corresponds to the same increase in the gas's internal energy, we can write

$$c_p m\Delta T = c_v m\Delta T + P\Delta V.$$

By applying the ideal gas law ($PV = nRT$), we can express the amount of work performed by a gas in terms of its molecular mass μ and the gas constant R:

$$P\Delta V = \frac{m}{\mu}R\Delta T.$$

Substituting this formula in equation (1) yields $c_p = c_v + R/\mu$, from which we get

$$\mu = \frac{R}{c_p - c_v} \cong 32.7\ \text{kg/kmole}.$$

From this we deduce that the unknown gas is oxygen with a small impurity of a heavier gas.

P144

The ring is affected by an accelerating force from the magnetic field. It's not an easy task to calculate it "head-on," but it can be elegantly found by invoking energy relations. As the ring travels between the poles of the magnet, a certain amount of heat is dissipated in it, which is just equal to the change in the ring's kinetic energy. This is because the interaction energy of the ring with the magnetic field is zero both before the ring enters the field and after it leaves the field. Our calculations will be greatly simplified if the change in velocity is small. We'll make this assumption and see later whether it's justified or not. And we'll make one other assumption: the ring is small compared to size of the magnetic field, so we won't be interested in what goes on when the ring enters or leaves the field.

Thus, the current induced in the ring is

$$I = \frac{\Delta\Phi/\Delta t}{R} = \frac{B_0 S v_0/a}{R} = \text{const.}$$

The time it takes the ring to travel through the field is

$$t = \frac{2a}{v_0}.$$

The thermal energy dissipated equals

$$Q = I^2Rt = \frac{2B_0^2S^2v_0}{aR}.$$

The change in velocity can be found from the equation

$$\frac{mv_0^2}{2} - \frac{m(v_0 - \Delta v)^2}{2} \cong mv_0\Delta v = Q,$$

from which we get

$$\Delta v = \frac{2B_0^2S^2}{maR} = \frac{B_0^2d^2}{8D\rho a} = 0.25 \text{ m/s} \ll v_0.$$

P145

In the reference frame attached to the center of mass of the nucleus and moving with velocity v relative to the Earth, the fragments fly off in opposite directions with equal velocities. The kinetic energies of these fragments are equal to the difference in the internal energies of the nucleus and the fragments. If the velocity of each fragment is v_1, then

$$2\frac{mv_1^2}{2} = E_1 - 2E_2.$$

From this we get the velocity of each fragment:

$$v_1 = \sqrt{\frac{E_1 - 2E_2}{m}} = \sqrt{\frac{2(E_1 - 2E_2)}{M}}.$$

The velocity of a fragment in the laboratory reference frame (attached to the Earth) is the vector sum of $\mathbf{v}$ and $\mathbf{v}_1$. The velocity $\mathbf{v}_1$ of the fragment in the reference frame of the nucleus can be directed anywhere. However, the vector $\mathbf{v}'$ of the fragment's velocity in the laboratory system forms the maximum angle with the vector $\mathbf{v}$ when the vector $\mathbf{v}_1$ is perpendicular to the vector $\mathbf{v}'$ (fig. 6). In this case the vector $\mathbf{v}'$ of the fragment's velocity forms an angle α with vector $\mathbf{v}$ that satisfies the equation

$$\sin\alpha = \frac{v_1}{v} = \sqrt{\frac{2(E_1 - 2E_2)}{Mv^2}}.$$

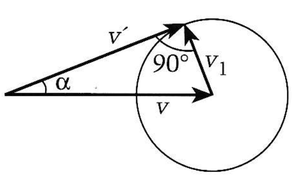


Figure 6

Brainteasers

B141

The problem can be solved in the standard way by using equations of motion. Here's a more elegant solution.

Let C be the point upstream from A at the same distance from A as B ($CA = AB$). Imagine a third motorboat that moves parallel with the first: it starts from C at the same speed, in the same direction, and at the same moment as the first motorboat. Since this third motorboat and the second motorboat move at the same speed relative to the raft (that is, to the water), both move toward the raft, and both are initially equidistant from the raft, they will always remain equidistant from the raft. When the first motorboat arrives at B, the second covers the same distance and arrives at A. So at this moment the raft will be the *same* distance from the second motorboat as from point A (that is, the imaginary third motorboat). (V. Dubrovsky)

B142

Use the ruler to draw two parallel chords AB and CD (fig. 7). Find the intersection points P and Q of the lines AC and BD and the lines AD and BC, respectively. Then PQ passes through the center O of the circle. Indeed, the reflection in the diameter perpendicular to AB and CD swaps the points A and B and the points C and D, and so it swaps the lines AC and BD. Therefore, it leaves the common point P of AC and BD intact—that is, P lies on this diameter (or its extension). This reasoning applies to Q as well, so PQ is just the extension of this diameter.

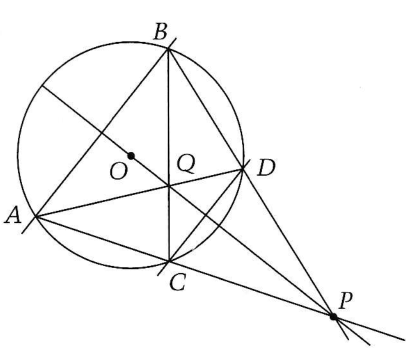


Figure 7

Now we can find the unknown center as the intersection of PQ and any other similarly constructed line.

B143

By drawing some rays we can see that the candle's image will move closer to the mirror after the sheet of glass is placed between the mirror and the candle (fig. 8). The blue lines show the paths of the rays without the glass, and the red lines correspond to the case when the glass is present.

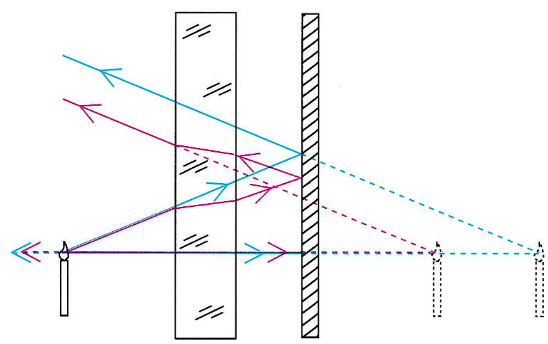


Figure 8

B144

The answer is $n = 9$. Inspecting the first digit from the right in the given equation, we find that $2 \cdot \text{O}$ is divisible by n. So either $\text{O} = 0$ or $\text{O} = n/2$. In the second case, from the third digit we derive $\text{K} > n/2$, but from the fifth digit we see that $3 \cdot \text{K} \le \text{T} \le n - 1$, so $\text{K} < n/3$. It follows that $\text{O} = 0$. Now we have these equations: $3 \cdot \text{T} = \text{K}n + \text{Y}$ (from the second and third digits), $3 \cdot \text{Y} = cn$ (where c is the number carried from the fourth to the fifth digit), and $3 \cdot \text{K} + c = \text{T}$ (the fifth digit). Multiplying the first equation by 3 and substituting the expressions for 3Y and 3K from the two other equations, we get 9T = (T – c)n + cn, and so n = 9. We must also check that there's at least one solution for this n. In fact, there are four: KYOTO = 13040, 16050, 23070, or 26080. (V. Dubrovsky)

B145

The covered loop is red. To prove this, we first note that the upper blue arc can't be connected to any of the other three blue arcs, because otherwise we'd be unable to close the red arc embracing the blue arc connected to the upper arc. If the three smaller blue arcs are parts of the same loop, the red arcs embracing them are also parts of one loop. Together with the two upper arcs this makes four partly visible loops, whereas there must be five. This leaves only one possibility: two of the smaller blue arcs belong to one loop, the third to another. Then three blue loops have uncovered parts, so the completely covered loop is red. Figure 9 shows how this is possible.

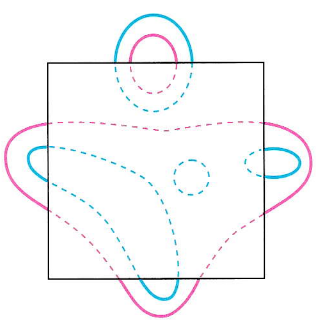


Figure 9

Kaleidoscope

1. The Earth turns 15′ on its axis every minute.

2. The Sun is at an altitude of $\phi = 45°$.

3. This will occur if the observer is at one of the Earth's geographical poles or if the star in question is situated at one of the celestial poles.

4. The visible path of the Moon in the sky practically coincides with that of the Sun, but the Moon makes a complete revolution not in a year but in a month. So if the observer is at one of the poles, the Moon will be above the horizon for two weeks and below the horizon for another two weeks.

5. The terminator on Venus is a semicircle, which is seen from the Earth at an angle as a semiellipse.

6. Yes—you can see such a reflection in equatorial regions.

7. For an observer on the Moon, the Earth neither rises nor sets.

8. Yes, one can—for example, by studying the behavior of a pendulum.

9. At the equator the ring looks like a band crossing the sky at the zenith; at the poles it can't be seen at all.

10. On the side of the Moon facing the Sun, the astronaut will see a total solar eclipse; on the other side, only the brighter stars will be seen against a black sky.

11. In summer (in the northern hemisphere) the distance between the Sun and Earth is greater than in winter, so the angular size of the Sun in summer is a bit smaller than in winter. On the other hand, the distance between the Earth and Moon on average does not depend on the season. This why the Moon covers the Sun completely more often in summer than in winter.

12. The distances from the Earth and the Moon to the Sun are practically equal. So, if the Moon and the wall had the same coefficients of reflection, their brightness would appear identical. So it can be inferred that the lunar soil is composed of dark rocks.

13. No, it doesn't. From the Moon one can see the solar corona, while it can be observed from the Earth only during a total eclipse of the Sun.

14. Light beams are deflected (refracted) in the Earth's atmosphere, so if the atmosphere were absent, the visible position of any star would shift slightly. For example, stars that can be seen near the horizon would disappear.

15. Comets participate in the diurnal rotation of the heavens.

16. Meteorites that encounter the Earth's "morning" hemisphere are traveling in the direction opposite the Earth's rotation, by and large, while those that strike the "evening" hemisphere are basically chasing after the Earth. So meteorites entering the atmosphere during the wee hours do so at a higher velocity and burn more brightly than those arriving before midnight.

Microexperiment. To verify that the difference in size at the horizon and zenith is only apparent, project an image of the disk (it's easier to do this with the Sun) on a sheet of paper using a lens with a long focal length (why?). The lens and paper must be perpendicular to the light beams. Measuring the images in both cases (the disk near the horizon and at the zenith), one can see that they have the same diameter.

Monopoles

This is the situation in which a magnetic pole is moving in an inverse-square magnetic field. It's the magnetic equivalent of an electron moving in the electric field due to a proton or, for that matter, a mass moving in the gravitational field due to another mass. The conditions are right for the moving monopole to orbit the stationary one. The total energy of the moving pole is the sum of its kinetic and potential energies:

$$E = \frac{1}{2}mv^2 - \frac{\mu_0}{4\pi}\frac{Q^*q^*}{r}.$$

The force equation for the moving monopole is

$$\frac{\mu_0}{4\pi}\frac{Q^*q^*}{r^2} = \frac{mv^2}{r}.$$

Combining these gives us

$$E = -\frac{\mu_0}{8\pi}\frac{Q^*q^*}{r}.$$

The initial energy of the moving pole is

$$E_i = \frac{1}{2}mv_0^2 - \frac{\mu_0}{4\pi}\frac{Q^*q^*}{r_0}.$$

We can solve for the radius of the

resulting orbit by equating these two energies. This gives

$$r = -\frac{\mu_0}{8\pi}\frac{Q^* q^*}{E_i}.$$

There are limits on this result. If the initial velocity v_0 is too high, the initial energy will be positive and there will be no solution for the orbital radius. The monopole will escape the pull of its neighbor. Too low an initial velocity may cause the monopoles to collide, if in fact they have any size.

What would we call our little magnetic atom? The author is partial to the name "Wylium."

Jesse James

1. The condition $D^2u(x) = u(x-1) - 2u(x) + u(x+1) = 0$ implies that

$$u(x) = \frac{u(x-1)+u(x+1)}{2}.$$

2. This rule is the same as Jesse's! Individual x gives $0.1u(x)$ to each neighbor, but receives $0.1u(x-1)$ from one neighbor and $0.1u(x+1)$ from the other, for a net change of $0.1[-2u(x) + u(x-1) + u(x+1)]$.

3. The quantity

$$[u(x-1, y) - 2u(x, y) + u(x+1, y)] + [u(x, y-1) - 2u(x, y) + u(x, y+1)]$$

now determines whether the individual located at (x, y) will get richer or poorer. In calculus it corresponds to the "Laplacian" Δu, defined as $\Delta u = u_{xx} + u_{yy}$.

4. The leftmost individual with \$130 receives a total of \$10 from two vertical neighbors and gives a total of \$10 to two horizontal neighbors. (Since that individual is at harmony with her or his neighbors, we could say that the distribution is "harmonic" at that location. In calculus, functions $u(x, y)$ satisfying $\Delta u = 0$ are called "harmonic functions.")

5. Setting the expression in problem 3 equal to zero yields the equation in the box below.

$$u(x, y) = \frac{u(x-1, y) + u(x+1, y) + u(x, y-1) + u(x, y+1)}{4}$$

Index of Advertisers

Corrections

Vol. 5, No. 2

p. 5, col. 2, first display equation: *for* $a_1 \cdot + a_2$ *read* $a_1 \cdot 10 + a_2$.

p. 5, col. 2, third display equation: *for* $9a_1\ 99a_2$ *read* $9a_1 + 99a_2$.

p. 6, col. 2, ¶1, 4 lines from bottom: *for* $n - \sigma(n)$ *read* $n - \sigma_2(n)$.

p. 7, col. 3, ¶1, l. 1: *for* a_1 *read* $a_1 = 1$.

p. 7, col. 3, ¶2, l. 2: *for* not divisible by n *read* not divisible by p.

p. 8, col. 1, second display equation: *for* $\sqrt{18}\log_4 2n$ *read* $\sqrt{2n}\log_4 2n$.

p. 8, col. 3, problem 5: *for* p^k *read* p_k.

p. 23, problem M128, l. 14: *for* points A and C *read* points A and D.

p. 32, col. 3, exercise 1, l. 4: *for* side AA_1C *read* side CA_1.

p. 35, col. 2, ¶1, line before display equation: *for* CC' *read* $C'C_1$.

p. 41, col. 1, exercise 4, l. 2: *for* in terms of the perimeter of the length *read* in terms of the length.

p. 42, col. 2, second display equation from bottom: *for*

$$\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\cos\frac{\pi}{2^{n-1}}}}$$

read

$$\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\cos\frac{\pi}{2^{n-2}}}}$$

p. 43, col. 1, problem 7, solution, l. 2: *for* $\cos(A + B)$ *read* $\cos(A + B)]$.

p. 58, problem M129, ¶1, l. 11: *for* $2 + n$ *read* $w + n$.

p. 62, problem 7(a), l. 3: *for* $2(x-1)^2$ *read* $2(x-2)^2$.

p. 62, problem 7(b), l. 1: *for* $1 \pm i\sqrt{2}$ *read* $-1 \pm i\sqrt{2}$.

p. 62, problem 7(c), l. 1: the equation should read

$$\frac{\sqrt{2}-1 \pm \sqrt{2\sqrt{2}-1}}{2}.$$

Vol. 5, No. 3

p. 9, col. 1: the display equation at the bottom of the column should read

$$h = \sqrt{\frac{\kappa t_\oplus}{c_p \rho}} \cong 4 \cdot 10^5 = 400 \text{ km}.$$

Our thanks to Prof. Paul Middents of Olympic College for bringing this to our attention.

p. 9, col. 2, ¶1, ll. 1–2: *for* Three hundred kilometers . . . a mere 4% . . . *read* Four hundred kilometers . . . a mere 5% . . . .

p. 12, fig. 4: the middle braid should be labeled "C"; the equation "$B \cdot C = D$" applies to the entire figure.

p. 35, col. 1, problem 1, last line: *for* $\ldots \times 17$ *read* $\ldots \times 17 + 1$.

Head over heels

The mechanics of an odd top

by Sergey Krivoshlykov

EDITOR'S NOTE: A LETTER arrived a while back at the editorial offices of *Kvant* (the Russian-language sister magazine of *Quantum*). A tenth-grader at the Titarevskaya Secondary School, V. Tkachov, wrote: "I bought a top in a toy store. When I set it in motion, it flips over and rotates on its handle. What physical laws underlie such a motion? What do the dimensions of the top have to be to make such paradoxical behavior possible?"

We think the answer to this letter will be of interest to many of our readers. This article was written by Sergey Krivoshlykov, a tenth-grader at Secondary School No. 45 in Kiev, and is based on the report he submitted to the Fourth Student Scientific Conference in the city of Kiev.

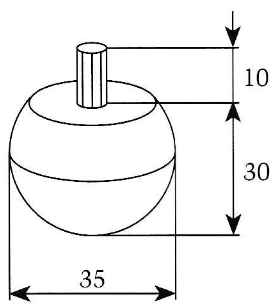


Figure 1

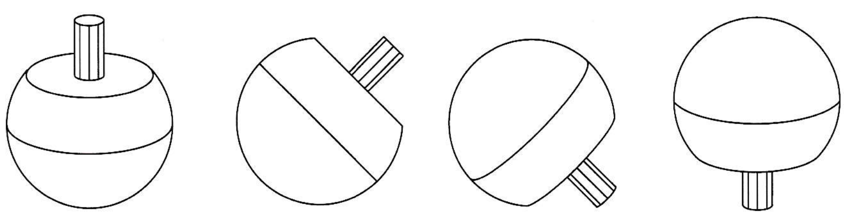


Figure 2

The top described in Tkachov's letter is often referred to as Thompson's top. It is a ball with the top part cut off. In the middle of the cut there is a handle that is used to spin the top (see figure 1, where the sizes are given in millimeters).

If the top is spun with the ball in the lower position, it flips over while rotating so that its handle touches the table top, then it pops up on the handle and continues spinning, upside down, in a stable manner (fig. 2). When the angular velocity decreases to a certain point, the top resumes its initial orientation. At first glance such behavior seems very strange. When the top flips up onto its handle, the top increases its potential energy. What makes it do this? It's widely known that any system tends to minimize its potential energy.

Let's begin by considering the rotation of a rigid body about its axis. Take a large flywheel (say, a bicycle wheel) that can freely rotate about a rigid horizontal axle. Setting the flywheel in motion, let's grab its axle with both hands and try to turn it in the vertical plane, as shown by the blue arrows in figure 3. We immediately feel a strong resistance. The spinning flywheel tries to conserve both the size and direction of its angular momentum, and hence the angular velocity of the rotation as well as the direction of the axle. Now let's apply more force. Strange as it may seem, we will not achieve the desired result. The axle turns—but not in the vertical plane, as we thought—no, it turns in the horizontal plane, as shown by the red arrows in figure 3.

This is strange only at first glance. It's known that as rigid bodies rotate, the external torque (a vector) equals the rate of change of the angular momentum (another vector):

$$\tau = \frac{\Delta(I\omega)}{\Delta t}.$$

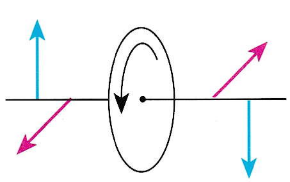
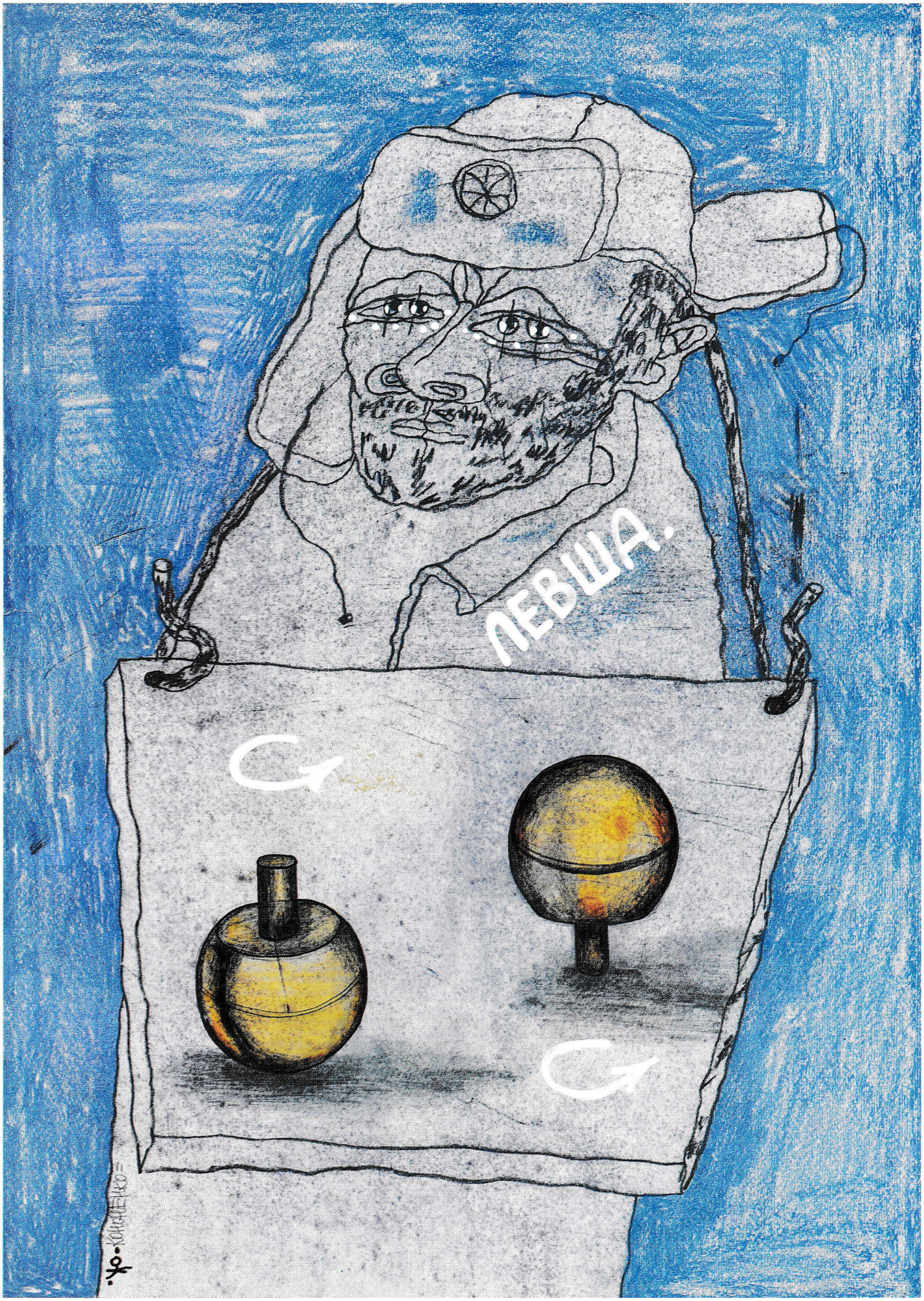


Figure 3

Art by Yury Kononenko


ЛЕВША.

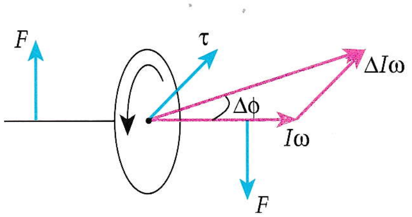


Figure 4

There is an analogous formula for linear motion: the external force acting on a body equals the rate of change of the body's momentum:

$$\mathbf{F} = \frac{\Delta(m\mathbf{v})}{\Delta t}.$$

Thus, the vector change of the angular momentum $\Delta I\omega$ of the flywheel in a small time interval Δt is parallel to the torque $\boldsymbol{\tau}$ (fig. 4)—that is, it lies in the horizontal plane. So the new vector for the angular momentum lies in the same plane. Note that because $\boldsymbol{\tau}$ is perpendicular to $I\boldsymbol{\omega}$, the change in the angular momentum $\Delta(I\boldsymbol{\omega})$ is perpendicular to the angular momentum $I\boldsymbol{\omega}$ itself. (Figure 4 shows $\Delta(I\boldsymbol{\omega})$ greatly enlarged.) So in our case the size of the angular momentum does not vary—only its direction changes. This means that the angular speed of the flywheel about its horizontal axis also does not vary.

If we try to twist a flywheel rotating in the opposite direction, its axle will also turn in the horizontal plane, but in the opposite direction. This movement of the axis of a rapidly spinning body perpendicular to the applied forces is called its "precession."

If the size of the external torque is constant, the rate of change of $I\boldsymbol{\omega}$ will be constant as well ($\Delta(I\omega)/\Delta t$ = const). In this case the flywheel's axle revolves with a certain angular velocity Ω, known as the precessional velocity. The larger the applied torque, the larger the precessional velocity of the flywheel:

$$\Omega = \frac{\Delta\phi}{\Delta t} = \frac{\Delta(I\omega)/I\omega}{\Delta t}$$

$$= \frac{\Delta(I\omega)}{\Delta t}\cdot\frac{1}{I\omega} = \frac{\tau}{I\omega}.$$

(Since the angle $\Delta\phi$ is small, $\Delta\phi = \Delta(I\omega)/I\omega$).

It's clear that the faster the flywheel rotates (that is, as ω and $I\omega$ increase), the smaller the precessional velocity for the same external torque (because the angle $\Delta\phi$ will be smaller in a time interval Δt—see figure 4).

Let's consider a common rotating disklike top. Initially, when the angular velocity is high, its axis is practically vertical. Then this velocity decreases due to friction at point A and air drag, and the top begins to precess about the vertical axis, circumscribing a conical surface whose apex is at point A (fig. 5). Why does this happen?

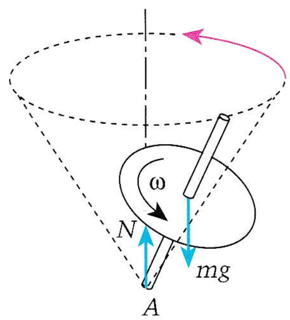


Figure 5

Let's consider the forces acting on the top. The force of gravity $m\mathbf{g}$ and the reactive force of the table $\mathbf{N}$ create a torque that tries to turn the top over. As a result, the top's axis moves perpendicular to the plane of these forces—that is, it precesses. The direction of the motion of the axis is shown in figure 5 by the red arrow. During this precession the top's axis circumscribes a conical surface whose apex is at point A.

It should be noted that the precession of the axis existed in the very beginning of the gyration due to the unavoidable push given to the top when it was "launched" (it's impossible to spin a top ideally), but the corresponding precession was small.

In addition to the torque produced by the two forces $m\mathbf{g}$ and $\mathbf{N}$, the top is affected by the torque due to the frictional force F_{fr} relative to its center of mass. Figure 6 gives an enlarged view of the top's tip. If the

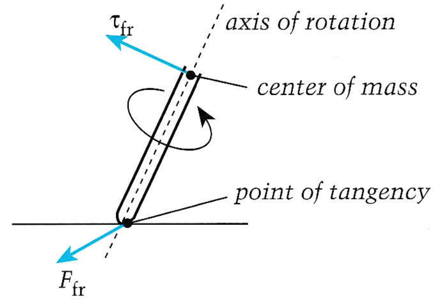


Figure 6

point of tangency of the axis with the table top doesn't lie on the rotation axis of the top (the top is tilted), the torque due to the frictional force lies in the figure's plane and points in the vertical direction. The change in the top's angular momentum (forced by friction) is also directed toward the vertical, so it is friction that causes the top's axis to assume the vertical position.

This can be clearly demonstrated by setting the top in motion at an angle with the vertical. After a while its axis assumes the vertical position. According to the right-hand rule, the torque due to the frictional forces is directed toward the vertical, so the change in the angular moment is also directed vertically and the top's axis tends to assume a position perpendicular to the plane of rotation. In sum, the tilted top is influenced by two torques: the torque due to the pair of gravity-related forces (the downward force of gravity and the upward supporting force) and the torque due to the frictional force. These two torques are always present when a top spins.

Now let's get back to Thompson's top and try to explain its behavior. Since this top consists of a truncated sphere, its center of gravity is lower than the geometric center of the sphere it's based on. When we spin the top, we unintentionally incline its axis from the vertical. Because the top is spherical, its point of support shifts as a result. Nevertheless, the axis of rotation will be vertical, so it will not coincide with the top's geometrical axis. Since the center of gravity lies lower than the geometric center of the sphere, the

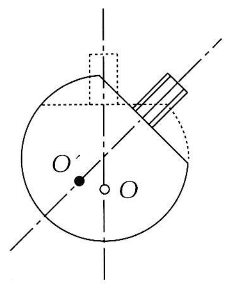


Figure 7

tilting of the top results in a displacement of the center of gravity from the axis of rotation (fig. 7). It will take the position O' and will rotate about the vertical axis. As it rotates with a high angular velocity, the top's center of gravity will rise for the same reasons a ball suspended on a thread rises when the thread is twirled, as shown in figure 8.

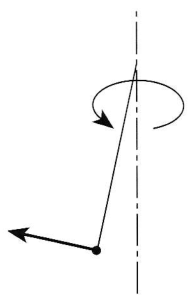


Figure 8

However, the top will not stop in the horizontal position, and after it passes through that position because of its inertia, its handle will come into contact with the table (fig. 2). As soon as this happens, the support point will "jump" from A to B (fig. 9) and the top, rotating about its axis, will precess about the axis BB'. In other words, Thompson's top will now spin just like a "standard" top. Under the influence of the frictional torque, it will bring its axis in line with the vertical axis BB' and continue to rotate,

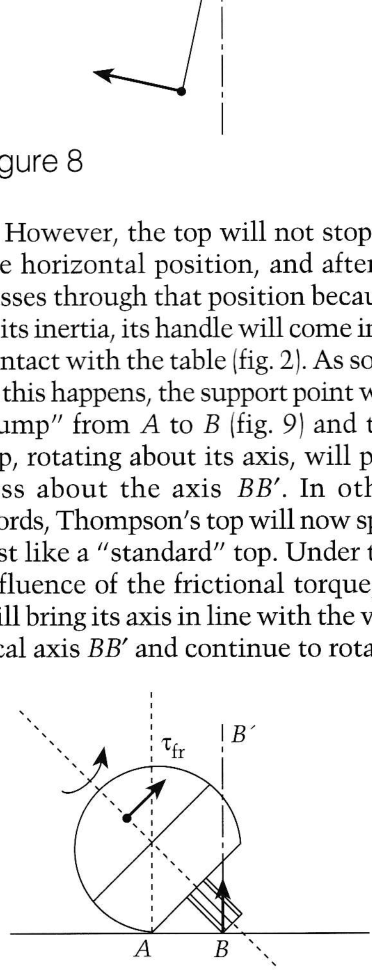


Figure 9

but now upside down.

This reasoning shows that the odd behavior of Thompson's top results from friction. Indeed, if there were no friction, after the top's handle touches the table top, the top would return to the horizontal position and continue spinning as long as its angular velocity is high enough. Then, under the action of the torque due to gravity, it would return to its initial position (fig. 2).

Here is another experiment related to Thompson's top. If this top is set in motion on a surface covered with a thin layer of powder, the powder will leave a trace on the top's surface, in effect recording the trajectory of the point of tangency of the top with the plane. This trace is shown in figure 10. On top the line

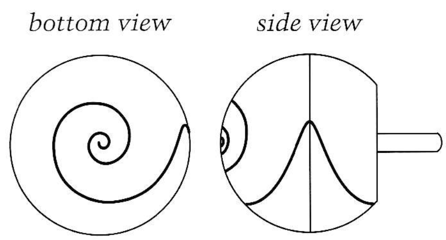
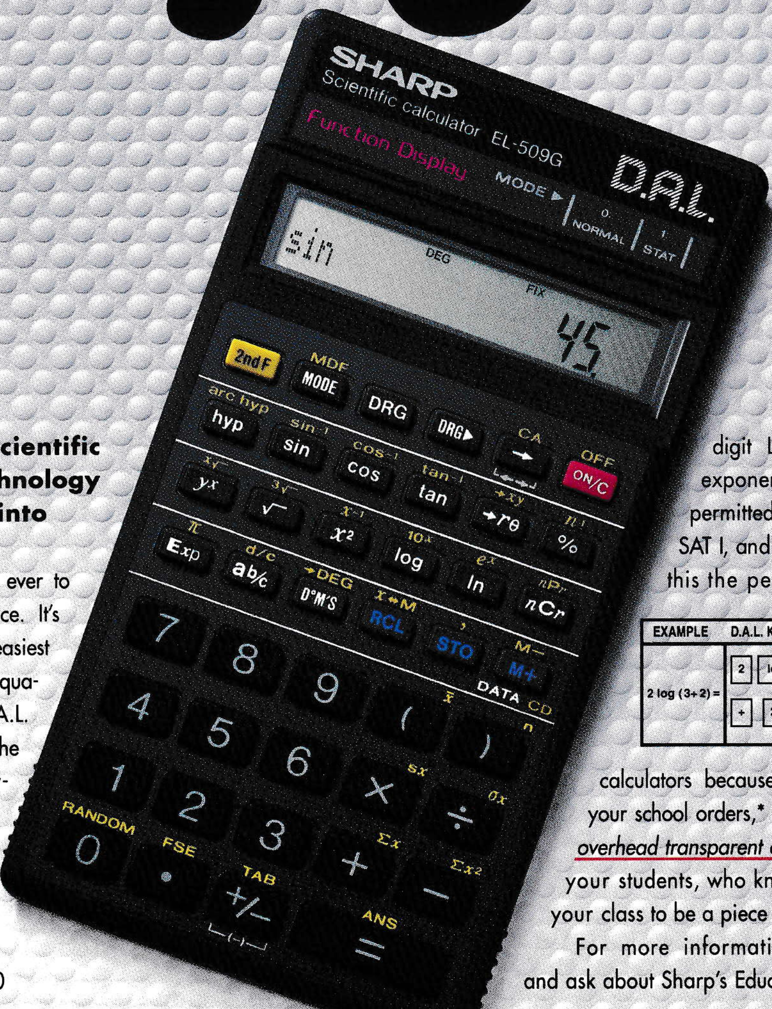


Figure 10

curls like a helix, but on the equator it changes direction and begins to uncurl backwards. Why does this happen? The law of conservation of angular momentum requires that the top rotate in the same direction in both its initial and upside-down positions. Let's spin the top in the clockwise direction (viewed from above). If the top doesn't stop rotating about its axis when turned upside-down, it would rotate counterclockwise in the new position. So, to conserve the size and direction of its angular momentum, the top must stop spinning about its geometrical axis at a certain moment in time and then begin rotating in the opposite direction. According to figure 10, this occurs when the top lies on its side.

As for the top's dimensions, they can be different, but one condition must be met: the center of gravity must not coincide with the geometric center of the truncated sphere.

Easy as π



Only Sharp's D.A.L. scientific calculators have the technology to turn algebraic logic into simple easy logic.

It's the most logical innovation ever to teach and learn math and science. It's Direct Algebraic Logic. The fastest, easiest and most error-proof way to enter equations into a scientific calculator. D.A.L. is a technology that simplifies the entry process by reducing keystrokes. Substantially less time is spent teaching calculators. More time is spent teaching coursework. Sharp's leading LCD technology is found on all 5 D.A.L. models including a large 10 digit LCD display and 2 digit exponent. What's more they're permitted for use on the PSAT, SAT I, and SAT II tests. Which makes this the perfect time to buy D.A.L.

EXAMPLE	D.A.L. KEYSTROKES	WITHOUT D.A.L.
2 log (3+2) =	2 log (3 + 2) =	2 x (3 + 2) log 2

calculators because, right now for every 80 your school orders,* Sharp will send you a *free overhead transparent calculator worth $80.* As for your students, who knows? They just might find your class to be a piece of cake.

For more information call: 1-800-BE-SHARP, and ask about Sharp's Educational Calculators.

SHARP®
FROM SHARP MINDS
COME SHARP PRODUCTS™

LCD

CALCULATORS • COLOR SCANNERS • COPIERS • CORDLESS TELEPHONES • FACSIMILE • LASER PRINTERS • NOTEBOOK COMPUTERS • PDAs • WORD PROCESSORS • WIZARD® ORGANIZERS

*Offer valid on EL-509G/EL-531G calculators purchased through authorized educational distributors.