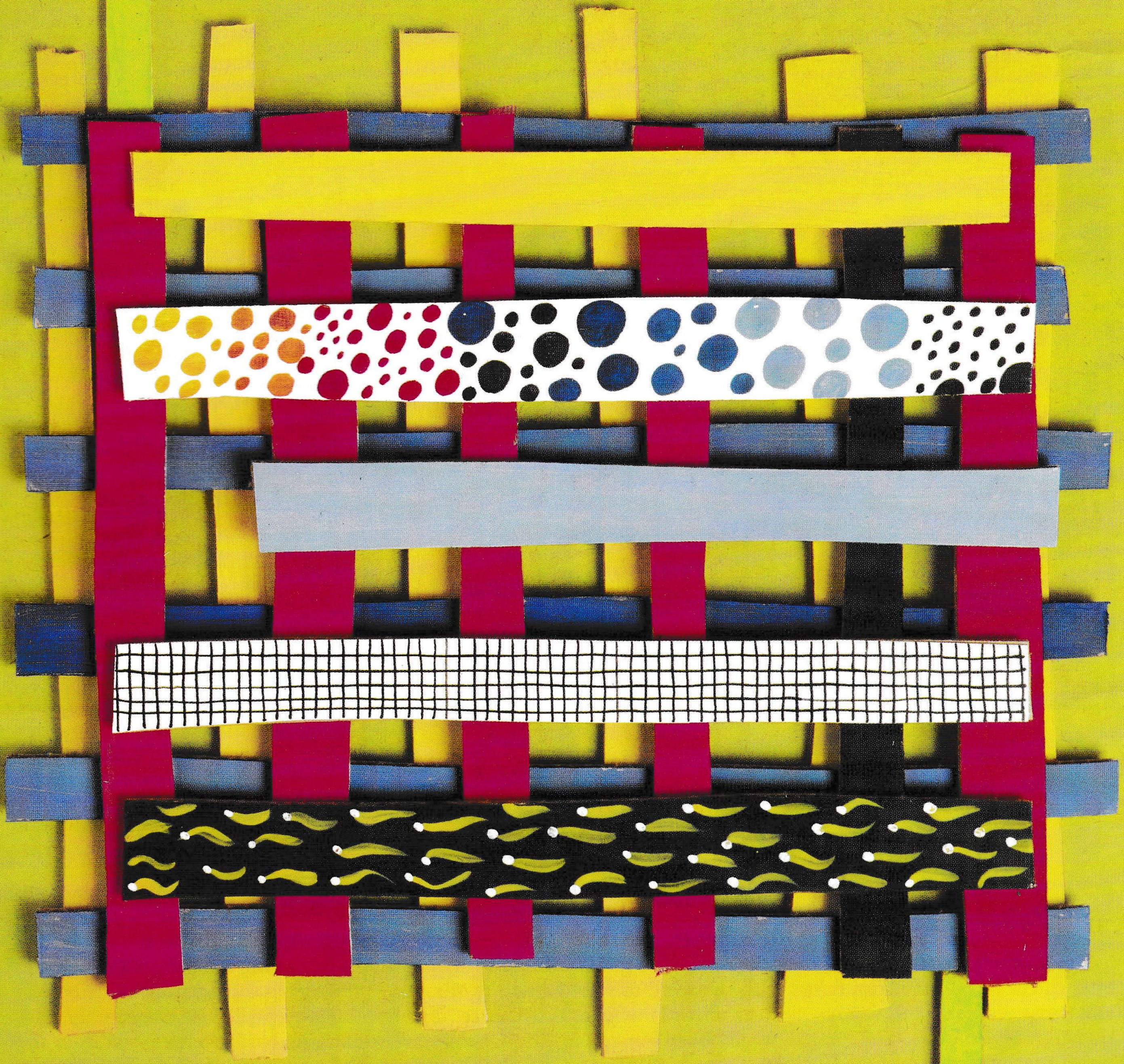
QUANTUM

JULY/AUGUST 1997

US $7.50/CAN $10.50



Springer

GALLERY Q

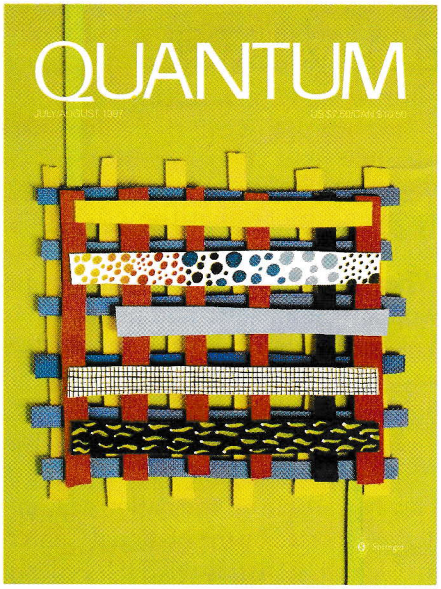
The Lighthouse at Honfleur (1886) by Georges Seurat

LIKE THE IMPRESSIONISTS IN GENERAL, GEORGES Seurat (1859–1891) was a keen observer of light and its effects. But Seurat seemed to take his interest in the subject to greater technical depths than the rest. Early in his brief career he became acquainted with the work of several Swiss aestheticians, one of whom examined the relationship between lines and images, while the other combined mathematics and musicology. Later he met the chemist Michel-Eugène Chevreul, who was 100 years old at the time, and investigated Chevreul's theories about light. In particular, Seurat experimented with the effects that can be achieved with the three primary colors—yellow, red, and blue—and their complements. Eventually he developed the technique (called Pointillism) of applying tiny spots of contrasting colors that could not be distinguished from a normal viewing distance, but which created a shimmering vibrancy in the painting as a whole.

The fact that the human eye will blend tiny spots of color into a completely different color when viewed from the proper distance is exploited in a "work of art" found in virtually every home: the color television set. The article "Can You See the Magnetic Field?" investigates this phenomenon, exposes the inner workings of the color television, and poses some interesting questions. It all begins on page 18.

QUANTUM

JULY/AUGUST 1997 VOLUME 7, NUMBER 6



Cover art by Vasily Vlasov

The colorful grillwork on our cover is a fanciful representation of a something you probably have in your own living room. It is the aperture mask in the picture tube of a color television set. And, as if to indicate the impermanence of a television image, the "same" grid is reproduced on page 19. The impression is quite different, and that is, of course, what allows television to show "moving" images. It actually shows a rapidly changing series of still images (much as a movie does).

Everyone knows how to watch TV. But do you know how to look at a magnetic field? You can use your TV set to do it. Turn to page 18 to learn how.

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

FEATURES

DEPARTMENTS

The demoflush figure

Algebra where you least expect it—the daily news!

TIME FOR TRUE CONFESSIONS. I love mathematics. No one had to tell me that I would need it or that I should study it. It was always exciting to me to explore new ideas with mathematics.

I have always found in the depth and structure of mathematics a haven; a predictability that comforts, an unexpectedness that delights, and a depth that goes far beyond polynomials or derivatives. And I have always been engaged in mathematics—having the benefit of learning by doing mathematics and the opportunity to construct my own mathematical understanding.

Another confession: I see mathematics everywhere. Most notably in science, technology, and in architecture. But I also appreciate the driving force of mathematics in dance, music, and the arts.

Although I'm immersed in mathematics, I'm still always thrilled when I see mathematics working for us in everyday—but ingenious—ways. In other words, finding mathematics where I least expect it.

Over the Memorial Day weekend I was reading the *Washington Post* when I came across the article, "In Ocean City, Md., Visitor Count Theory Holds Water." In a lighthearted manner, the *Post* pointed out that while many vacation resorts count cars and hotel occupancy rates to measure their popularity, Ocean City records every shower, every load of laundry, and every toilet flush. Town locals refer to it as the "demoflush figure" (as in "demographics" and "flush").

According to the *Post*, there's a complicated algebraic formula that over the years has become simplified this way:

> To calculate any day's demoflush figure, start with the number of gallons of sewage that rush into the town's 64th Street wastewater treatment plant by midnight. Divide that number by 36.04, a factor based loosely on the average number of gallons of water a beachgoer uses in 24 hours. Voilà: the official estimated one-day population of Ocean City.

The story included a graphic with the following equation for the total recorded sewage output s:

$$s = a + bx + cy,$$

where a is the infiltration (gallons of groundwater seeping into the sewage pipes), b is the average sewage output per resident or overnight visitor, x is the number of residents and overnight visitors, c is the average sewage output per day visitor, and y is the number of day visitors. This equation gave readers a chance to see what factors lay behind the magic number 36.04.

When I read about the use of algebra to translate sewage into a population count, I was excited—partly because of the way it was being used, but partly because the *Washington Post* saw fit to publish the article, showing the public how mathematics is used in our daily lives.

We see opportunities for mathematical reasoning in absorbing and acting upon the enormous amounts of information that we encounter daily. Take, for instance, the newspaper. Pie charts and bar graphs are the most common, but we are also seeing a variety of arithmetic equations being used to illustrate a point. When you describe mathematics in words, it is opaque and ponderous. Yet, when you take the same idea and put it into mathematical symbols, you simplify it.

For example, in the October 16, 1996, issue of the *Chicago Sun-Times*, I read an article on the nation's fat index as reported by the National Center for Health Statistics (NCHS). Federal guidelines suggest that people keep their body mass indexes under 25.

The *Sun-Times*, using the NCHS formula as a guide, advised readers to "do the math" and calculate their body mass index as follows:

> Multiply your weight in pounds by 0.45 to get kilograms. Next, convert your height to inches. Multiply this number by 0.0254 to get meters. Multiply that number by itself. Then divide this into your weight in kilograms.

I had to chuckle. Such complicated rhetoric could easily be simplified in an algebraic equation as follows:

$$\mathrm{BMI} = \frac{0.45W_\mathrm{p}}{(0.0254H_\mathrm{i})^2},$$

where W_p is weight in pounds and H_i is height in inches. The article indicates that if your body mass index (BMI) is greater than 25, you are too fat. That is, you are overweight if

$$\frac{0.45W_\mathrm{p}}{(0.0254H_\mathrm{i})^2} > 25.$$

This can be calculated to be just

about equivalent to

$$700\frac{W_p}{H_i^2} > 25,$$

which (after we divide both sides by 700) is equivalent to

$$\frac{W_p}{H_i^2} > 0.035.$$

In other words, you did not have to convert weight to kilograms or height to centimeters.[1] This would have made it easier to see the algebraic reasoning: divide your weight (in pounds) by the square of your height (in inches). If that number is greater than 0.035, you are too fat.

Isn't that much easier? Algebra, for all the mystery it holds, opens the doors to all kinds of opportunities. It provides people with a means to formulate problems and reach conclusions based on reasoning about situations. Apparently, the town of Ocean City recognizes that.

I think it's only fair to tell you, though, that there's some debate over the demoflush figure as an accurate means of population estimates. For example, what about drinking at the beach that affects people's use of water? And what about beachgoers' predicted twice-a-day showers (before and after the beach) that have not been factored in?

But as Ocean City officials point out, cars coming into town could carry a lone shopper or seven high school seniors, and condominiums could house one person or 14, so at this point their demoflush figure is still their best bet.

What do you think? If you have another way to tally the tourist population for Ocean City, or if you have seen algebraic reasoning at work in everyday situations, please write me and let me know. I'd love to see what innovative ideas you may have.

—Linda P. Rosen

Linda P. Rosen *is executive director of the National Council of Teachers of Mathematics, 1906 Association Drive, Reston VA 20191-1593.*

[1]NCTM encourages the use of SI units among the general public in the United States.

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

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

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 8 (6 issues) will be published in 1997–1998. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; electronic mail: quantum@nsta.org. **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 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Periodicals postage paid at New York, NY, and additional mailing offices. **Postmaster:** send address changes to: *Quantum*, Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485. Printed in U.S.A.

Subscription Information:

North America: Student rate: $17.70; Personal rate (non-student): $25; Institutional rate: $45; Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus NJ 07096-2485; or call 1-800-SPRINGER (777-4643) (in New York, call [201] 348-4033).

All Countries Outside North America: Personal rate is US$37 including surface delivery charges; institutional rate is US$57; airmail delivery is US$18 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Colder means slower

The elegant and wide-ranging Arrhenius equation tells us how much slower

by Henry D. Schreiber

Art by Sergey Ivanov

RELAXING ON A COOL SUMmer night in the country, you notice that the crickets aren't as noisy and the fireflies aren't flashing as brightly as the night before. The only apparent difference is that the previous night was somewhat warmer. After thinking about it, you realize your observation may not be that surprising—common sense tells you that things tend to slow down when they get colder. You keep food from spoiling by storing it in a refrigerator, and you know that bears slow down so much they often take a long nap in the winter. So perhaps it's understandable that crickets chirp less and fireflies flash less when they are cold. But then you wonder whether you can quantify these observations. Is there some factor by which the activities of these critters decrease with temperature? Can you model this temperature dependence by a mathematical equation? Would such modeling aid in understanding how crickets chirp and fireflies flash? Are theories about such behavior or the effect of temperature already available? These and numerous other questions pop into your mind. In your search for answers, you realize the importance of mathematics for interpreting the happenings in both the natural and physical worlds.

Mathematical models

Central to the scientific method are the activities of making observations and collecting data to either substantiate or refute a hypothesis, such as whether cricket chirping systematically decreases with temperature. You realize, however, that to measure the amount of cricket chirping every degree from, say, 10°C to 25°C, would take lots of time and effort. A better approach might be to do measurements at a few temperatures, then model the results with a mathematical equation to describe the temperature dependence of this behavior. After all, many modern scientific theories explain observed behaviors or properties within the context of an equation.

Figure 1 shows some data for the frequencies of cricket chirping and firefly flashing, in units of chirps per minute and flashes per minute, respectively, as a function of temperature. Both relationships show a similar dependence on temperature. You envision that you should be able to represent these data with mathematical equations.

Linear, parabolic, sinusoidal, exponential, and logarithmic functions are all available to mathematically describe behaviors of things ranging from electrons to crickets. You ask first, "What function is the best model of the data in figure 1?" and then, "Why does that equation work?" Often simple mathematical considerations for the laws of nature allow scientists to understand seemingly paradoxical situations. Agreement with a particular mathematical model provides a possible common underlying reason for an observed behavior. In other words, much of nature follows set rules, which in turn can be expressed by specific mathematical equations.

The Arrhenius equation

You survey references to see how others model and explain the effect of temperature on how fast things happen. Time and time again, the Arrhenius theory surfaces to rationalize thermal effects by a deceptively simple equation:

$$\text{rate} = Ae^{-E/RT},$$

where A and E are constants unique to the particular system, R is the ideal gas constant of 8.314 J/mole·K, and T is the temperature in kelvins (°C plus 273). Several scientists developed this equation by trial and error over a century ago, but the Swedish chemist Svante Arrhenius was the first to justify the equation on a theoretical basis with thermodynamics. Accordingly, this fundamental equation now bears his name.

The Arrhenius equation bridges the gap between the world of crickets and fireflies and that of molecules with mathematical images of rates of molecular reactions. In order for molecules to react, Arrhenius argued that two criteria have to be met. First, the molecules have to collide; and second, they have to possess enough energy for the reaction to take place. Thus the rate of the reaction depends directly on the frequency A of collisions and on the efficiency ($e^{-E/RT}$) of such collisions. Molecular collisions are most effective when the energy barriers E for reaction are low and the temperatures T are high. This exponential efficiency function is always a fraction ranging from near zero at low temperatures to near unity at sufficiently high temperatures. Its actual value represents the interplay of the relative magnitudes of E (the energy required before the molecules can react) and T (the measure of the energy available to the molecules).

You can manipulate the Arrhenius equation by taking natural logarithms of both sides of the equation to yield

$$\ln(\text{rate}) = \ln A + \ln e^{-E/RT},$$

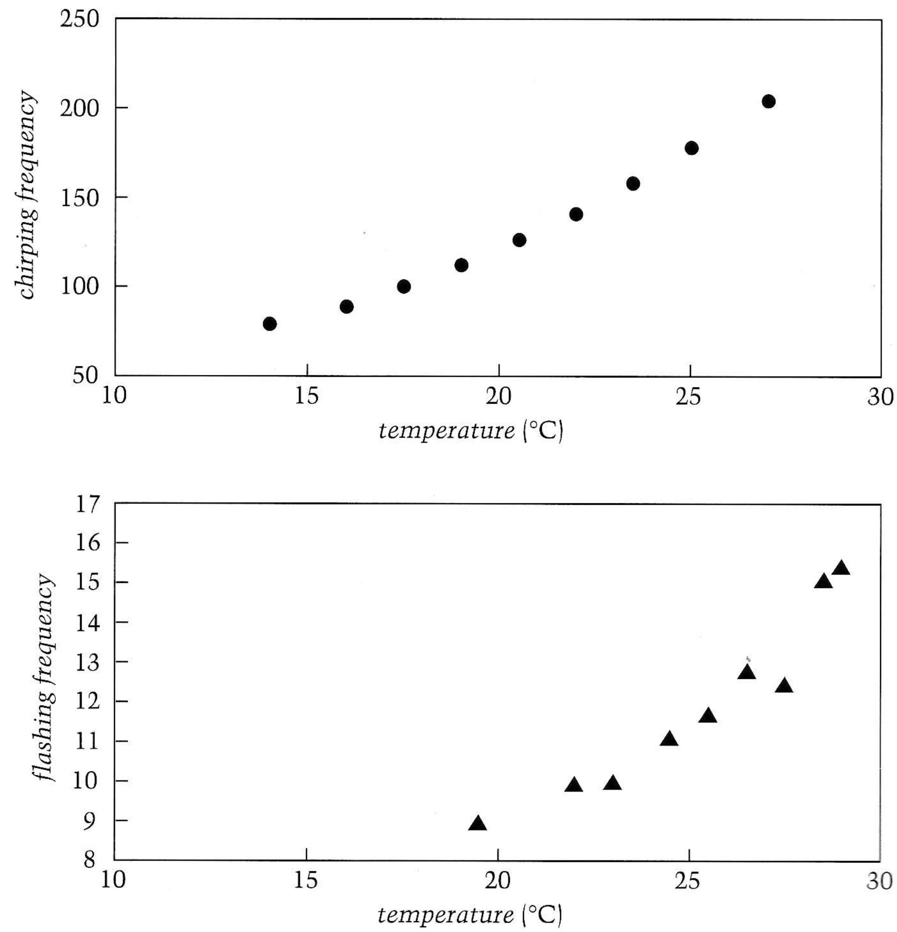


Figure 1
Temperature dependence of cricket chirping and firefly flashing.

or, upon rearrangement,

$$\ln(rate) = -\frac{E}{R}\frac{1}{T} + \ln A.$$

This is now in the form of a straight line equation: $y = mx + b$, where the y-variable is the natural logarithm of the rate and the x-variable is the reciprocal of the absolute temperature. The slope m of the line is proportional to the energy E required for reaction, while the y-intercept b depends on the frequency A of collisions.

A straight line resulting from a plot of the logarithm of how fast something happens versus the reciprocal temperature in kelvins conveys the message that that system behaves according to the Arrhenius equation. You can calculate the numerical value of E from the slope of this line. The greater the change in rate with temperature, the greater the value of E and the more energy has to be supplied in order for the molecules to react. To illustrate the collection of rate–temperature data and its subsequent correlation to the Arrhenius equation, refer to the boxed insert describing an experiment that can be done in the laboratory or at home. When a system obeys the Arrhenius equation, this implies that rates are governed by molecular reactions according to the two Arrhenius criteria—molecules colliding, but needing sufficient energy before reacting.

Chirping, flashing, and colliding

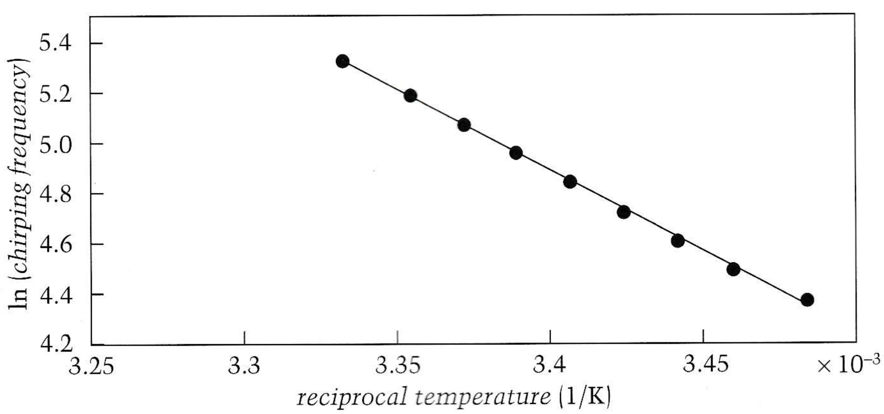
Figure 2 shows the data of figure 1 plotted as the logarithm of the chirping or flashing frequency with respect to the reciprocal temperature in kelvins—that is, as a test of the Arrhenius equation. You see the linear function for ln (rate) versus $1/T$ in figure 2 much easier than the exponential form for rate versus $e^{-E/RT}$ in figure 1. It's apparent from figure 2 that crickets chirping and fireflies flashing obey the Arrhenius relation. You can also determine the values for E and A from the slope and intercept of the straight line. The energy required to initiate chirping or flashing is about 50 kJ/mole, as calculated from the slopes of the straight lines in figure 2. This energy is consistent with your determination of the energy required for other reactions such as the bleaching of red food coloring.

Because the rates of cricket chirping and firefly flashing follow the Arrhenius equation, you conclude that these events are controlled by molecular reactions. Cricket chirping and firefly flashing are reflections of happenings in a molecular world. Thus, the underlying mechanisms involve first the collisions of molecules, then rearrangements of these molecules in reactions that result in chirps or flashes. The Arrhenius equation, although rooted in the molecular world, manifests itself by describing how fast events happen in the macroscopic world. Mathematics aid in such imaging. The temperature dependence of cricket chirping frequencies would be hard to envision directly unless you count chirps as logarithms of chirps and measure temperatures in reciprocal units.

Another advantage in determining the mathematical equation that models an activity is to use the equation to extrapolate your experimental results. For example, suppose you want to know the expected frequency of cricket chirping at 30°C, somewhat outside the temperature range of your experiments. The equation for the straight line in figure 3 is

$$\ln(chirping\ frequency) = -\frac{6420}{T} + 26.72.$$

If you now substitute the temperature of 303 K (273 + 30°C) in this equation, you solve the equation for a chirping rate of 253 chirps per minute. Other scientists may test such mathematical models, and thus your underlying hypothesis for the origin of cricket chirping, by comparing your predictions with direct measurements.



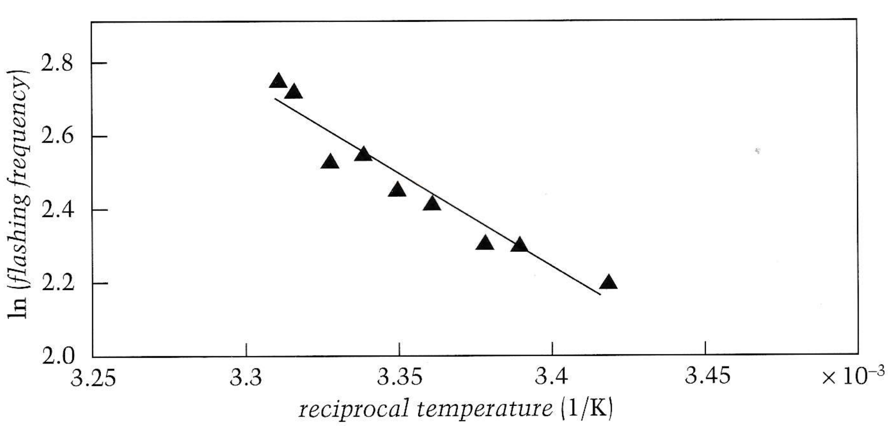


Figure 2

Arrhenius plot for frequencies of cricket chirping and firefly flashing.

Measuring how temperature affects a reaction rate

This simple experiment illustrates the collection and manipulation of data to test the Arrhenius equation. It's much easier than trying to convince crickets or fireflies to cooperate, and it can be done within the confines of a laboratory or even your kitchen. Always wear eye protection when doing such experiments and exercise good safety practices, such as adult supervision.

If you mix solutions of bleach and red food coloring, the red color eventually fades to a colorless solution. The bleach oxidizes the red dye to a colorless product, much like you add bleach to your wash to get clothes whiter. How fast this reaction occurs depends on the relative concentrations of the bleach and food coloring as well as the temperature. However, you focus on just the temperature effects by keeping the amounts of the reactants constant.

For this experiment, you measure the time that it takes for the reaction to proceed to a certain point—that is, for a defined amount of the red color to disappear. This is analogous to timing a runner in the 100-meter dash—one can then measure the speed of the runner by the time taken to run that distance. The shorter the time, the faster the runner. The rate of a reaction is similarly the distance that a reaction travels (or reacts) per unit time.

Reference solutions

Prepare a stock solution of the red food coloring by mixing 4 drops of this food coloring with 100 mL of water. Likewise, prepare a stock solution of bleach by mixing 25 drops of any brand of household bleach with 100 mL of water. Use beakers or plastic cups for these solutions.

Obtain two 13 × 100 mm test tubes (or alternatively 3-oz. plastic cups). Label one test tube I; mark a line at 3 mm and one at 6 mm from the bottom. Pour the stock solution of the red food coloring to the 3-mm mark, then add water to the 6-mm mark, and mix. Label the second test tube F; mark lines at 0.75 mm and at 6 mm from the bottom. Once again pour the stock solution of the red food coloring to the first mark, dilute with water to the 6 mm mark, and mix. Test tube I (initial) will represent the color of the mixtures before reaction; while test tube F (final) will represent the color, a much lighter red, after the reaction travels an arbitrary distance. This distance turns out to be 25% (0.75 is 25% of 3 mm) of the red color left, or in other words after the reaction has proceeded to 75% completion. In this experiment, you will time (to the nearest second with a watch) how long it takes to get a mixture from color I to that of color F at several temperatures.

The reaction

Obtain two more 13 × 100 mm test tubes (or 3-oz plastic cups). Mark each one 3 mm from the bottom. Pour the stock solution of the red food coloring to the 3-mm mark in the one test tube. In the other, do the same with the stock bleach solution. Pour the bleach into the red food coloring and start timing as you mix. Record the time that it takes for the color that is initially equivalent to that of I to fade to that represented by test tube F. Measure the temperature of the mixture with a thermometer—it should be room temperature. Afterwards the reaction mixture can be poured down the drain, and the two test tubes rinsed with water for re-use.

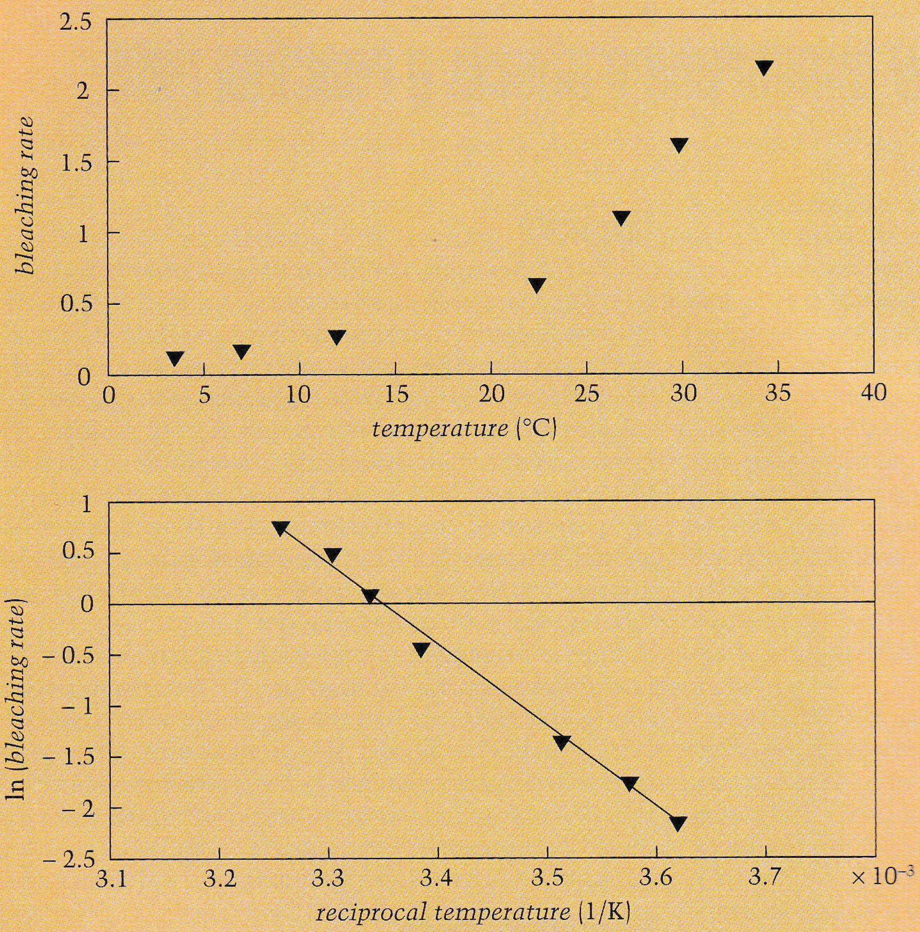
Repeat the experiment several times, but at higher and lower temperatures. For example, let the two test tubes containing 3 mm each of the respective stock solutions of red food coloring and bleach sit in the refrigerator, freezer, warm-water bath, etc., for several minutes before mixing, timing, and recording the temperature. Time the transition from color I to color F by comparing colors of the reaction mixture and the reference F solution.

Typical data and results

You will collect data similar to that displayed in the table below. However, your exact times may depend on the strengths of the solutions (due to variations in different brands of food coloring and bleach). The data should confirm your expectation that it takes less time for the reaction to proceed a certain distance at higher temperatures, and takes longer at lower temperatures. Because the reaction travels 75% of its way to completion in the times measured, you calculate relative rates by dividing 75% by the time. This provides the average rate in terms of percent progress per second. In other words, the reaction goes fast at high temperature, and slow at low temperature. If

T (°C)	t (s)	$1/T$ (K^{-1})	rate (%/s)	ln (rate)
3.5	690	0.00362	0.11	−2.22
7.0	474	0.00357	0.16	−1.84
12.0	292	0.00351	0.26	−1.36
22.5	121	0.00338	0.62	−0.48
27.0	69	0.00333	1.09	+0.08
30.0	47	0.00330	1.60	+0.47
34.5	35	0.00325	2.14	+0.76

Experimental results for the reaction of red food coloring with bleach.



Bleaching rate for red food coloring: temperature dependence.

you plot these rates, with units of percent/s, as a function of the temperature, you produce a graph (like the one in the figure at the left) with the same shape as that manifested by the chirping of crickets and the flashing of fireflies in figure 1.

To test the Arrhenius equation, you now plot the natural logarithm of the rates as a function of the reciprocal temperature in kelvins. As shown in the figure, the experimental points define a straight line. Thus the Arrhenius equation provides a reasonable explanation for the data. You calculate the energy E required for this reaction by multiplying the slope of the line by the numerical value of R and changing the sign, because the slope equals $-E/R$ in this straight line equation. In this case, E is about 70 kJ/mole. For the reaction of bleach and red food coloring to occur, first the bleach and dye molecules must collide, then only a certain number of these collisions must be effective—those in which the colliding molecules possess at least 70 kJ/mole. The reason the reaction is slower at the lower temperatures is that a smaller percentage of the colliding molecules possess energy above this critical value of 70 kJ/mole.

Flowing, counting, and more colliding

You have determined that the rates of crickets chirping and fireflies flashing, as well as the bleaching rate of red food coloring, can all be explained by the Arrhenius equation and thus have as their basis the collisions and rearrangements of molecules. What other activities follow the Arrhenius equation? The temperature dependence of the rates of most chemical reactions—for example, the decay of hydrogen iodide (HI) to hydrogen (H_2) and iodine (I_2)—likewise obey this equation. After all, to start this reaction, two HI molecules must first collide, but before they rearrange they must possess a certain amount of energy for their activation to a reactive state. A wide range of activities from diverse scientific fields similarly follow the Arrhenius equation: material diffusing through a liquid, the beating of a terrapin heart, the setting of cement, alpha brain wave frequency in humans, the creep rate of ants—the list goes on. The Arrhenius equation describes the effect of temperature on each of these phenomena.

Figure 3 shows that the Arrhenius equation is equally valid for the viscosity of a liquid and for a person's counting rate. In the former case, you might argue that the Arrhenius plot appears reversed. But remember that the viscosity of a liquid is the resistance to flow. The rate of flow or the fluidity of the liquid is the reciprocal of the viscosity, which would then correct this apparent discrepancy. Even the effect of a person's temperature on how fast that individual counts from one to ten, and thus one's time perception, follows the Arrhenius equation. This mathematical model is widely applicable! Why? Because all involve the common feature of colliding molecules and the breaking and making of bonds within the molecules to initiate the activity.

You can extend your thinking to other experiments that may serve as tests for the Arrhenius equation. For example, you can easily measure the rate at which fruit rots as a function of the prevailing temperature. Get a bunch of bananas. Separate them, assuming each to be at the same initial level of ripeness. Keep one at room temperature, another in a refrigerator, another next to a radiator. Time how long it takes each banana to achieve the same level of rottenness—that is, for three quarters of each banana's surface to turn black—as a function of the temperature. Would you expect this process to be governed by the Arrhenius equation? Why? Here's

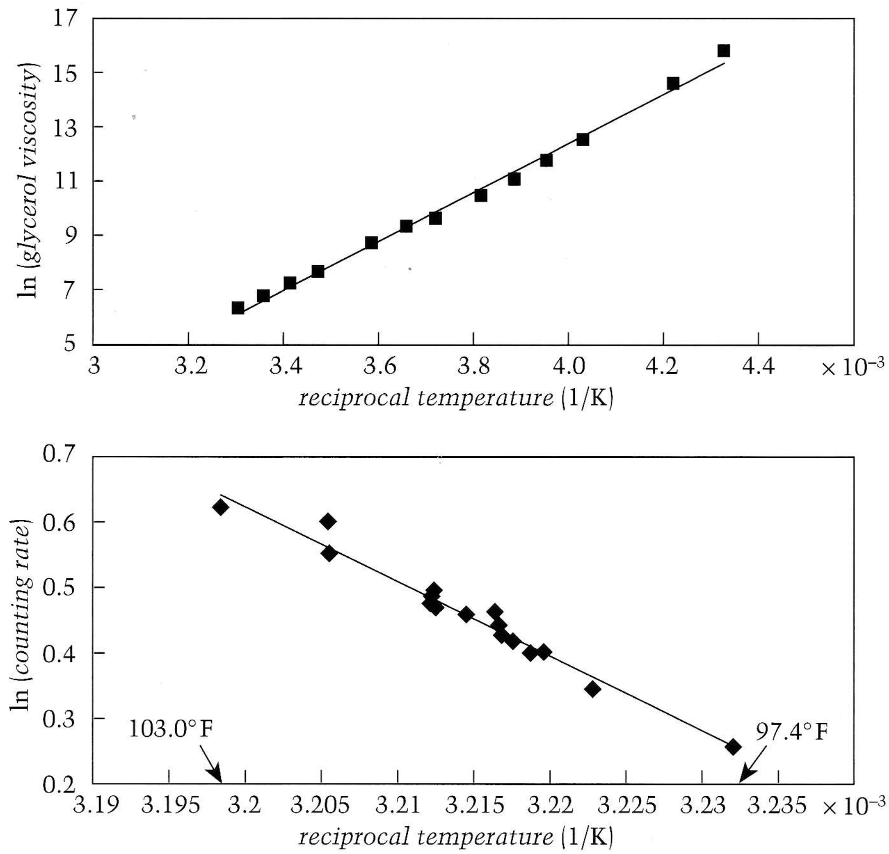


Figure 3
Arrhenius plot of glycerol viscosity and counting rate.

your chance to participate in the scientific process of posing of questions, proposing a certain hypothesis or model, then testing the models.

Interestingly, the form of equations that describe the temperature dependence of other properties is strikingly similar to the Arrhenius equation. The Van't Hoff equation details the temperature dependence of the solubility s for a solid dissolving in a liquid:

$$\ln s = -\frac{E_s}{RT} + I_s,$$

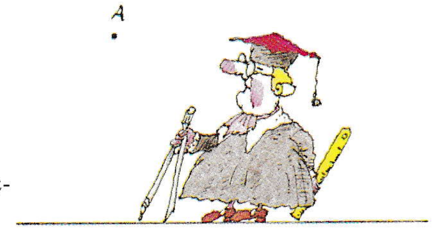
where E_s is the energy associated with the dissolution process. And the Clausius–Clapeyron equation gives the temperature dependence of the vapor pressure p of a volatile liquid:

$$\ln p = -\frac{E_v}{RT} + I_v,$$

where E_v is the energy required to vaporize the liquid. The logarithm of that property versus the reciprocal of the temperature (in kelvins) defines a straight line whose slope depends on the energy required to initiate that event, whether it involves two molecules reacting, a water molecule hitting a solid's molecule to send it into solution, or a molecule breaking free of other molecules to escape into the gas phase. All rely on the collision of molecules and a critical energy to be surpassed before anything happens. These universal mathematical equations modeling the laws of nature allow scientists to understand seemingly different situations.

One approach in science is for you to make an observation, to model that behavior with a mathematical equation, to explain the event by comparing it to similar equations, and to make predictions. The underlying theme in your understanding is that the molecular world controls events in your real world. You use mathematics—an amphibian living in two worlds—to explain both of them.

Just for the fun of it!



B206

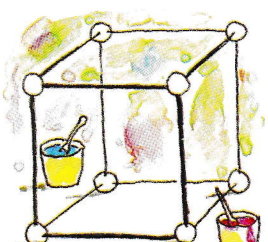
Perplexing perpendicular. A straight line and a point A outside it are given. Using a compass and a straightedge, draw the perpendicular from A to this line such that the total number of lines or circles drawn during the construction does not exceed three (the third line must be the perpendicular itself).

B207

Sunlit windows. A town is nestled in the eastern slope of a mountain. In the morning a weary traveler who rested at the foot of the mountain observed the sunlight reflected in the windowpanes. He noticed that as time passed the "illuminated" windows shifted: in some houses the light was "turned off," in others it was "turned on." In which direction do the illuminated windows shift: up or down? To the left or to the right? Explain this phenomenon.

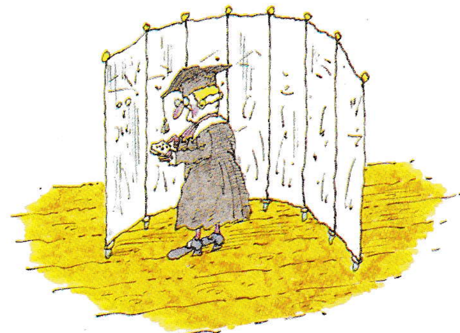
B208

Pedestrian banker. Mr. R. A. Scall, president of the Pyramid Bank, lives in a suburb rather far from his office. Every weekday a car from the bank comes to his house, always at the same time, so that he arrives at work precisely when the bank opens. One morning his driver called very early to tell him he would probably be late because of mechanical problems. So Mr. Scall left home one hour early and started walking to his office. The driver managed to fix the car quickly, however, and left the garage on time. He met the banker on the road and brought him to the bank. They arrived 20 minutes earlier than usual. How much time did Mr. Scall walk? (The car's speed is constant, and the time needed to turn around is zero.) (I. Sharygin)



B209

Color the cube. Color the eight vertices of a cube in two colors (red and blue) such that any plane containing three points of one color contains one point of the other color. (N. Vasilyev)



B210

Strange calculator. Suppose you have a hand-held calculator that can perform only two operations: for a given integer a, it can compute either $2a + 1$ or $(a - 1)/3$. (The second operation is possible only if $a - 1$ is divisible by 3.) Can you produce an 8 on this calculator starting with a 1?

ANSWERS, HINTS & SOLUTIONS ON PAGE 54

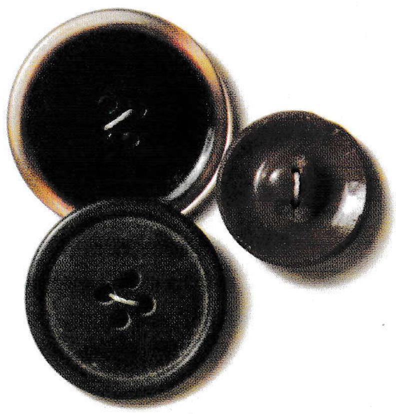
Art by V. Ivanyuk, P. Chernusky

4

2

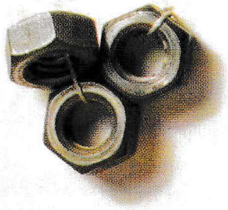
7

1



29

9



16

15

12

All sorts of sorting

Sometimes it's a big job, even for computers!

by P. Blekher and M. Kelbert

IN THIS ARTICLE WE'LL TALK about one of the branches a of a comparatively young science: pattern recognition theory. Specifically, we'll discuss classification methods. *Classification* is the division of a certain set of objects into groups of objects that are close to one another in some sense. You may want to classify, say, the factories in a given industry, the earthquakes observed in a certain region, the weather in September over the course of many years, species of dinosaur in a particular geologic period, and so on. What is important here is that the objects classified be characterized by a certain set of numbers and attributes. For simplicity, we'll consider the case where each object is described by a certain set of numbers.

Classifying students

Suppose we want to divide all the students in the ninth grade of a given school according to how they spend their free time after school. We'll question each student and write down what portion of this time the student spends doing homework and how much time is spent on recreation. The remaining time (for meals, sleep, transportation, and so on) doesn't interest us at this point. Thus, each student will be characterized by two numbers: homework time and recreation time. So a student can be represented by a point in the coordinate plane, and we arrive at a purely mathematical problem: how can a given set of points in the plane be split into groups of "close points"?

The same question arises in many serious applied problems except that the number of coordinates in these problems is much larger (usually several dozen).

For instance, there are many companies in the textile industry. Some of them are specialized, manufacturing a rather narrow range of products; others are huge enterprises that produce a wide variety of items. There are giant plants, medium-sized factories, and small, local factories. To compare the work of different enterprises and plan their operations,[1] one has to classify them—that is, divide them into groups containing each type of enterprise. The enterprises differ in their economic indices. Let's choose the most important of them (for instance, gross output, production costs, wages) to base our classification on. Denote the number of these indices by n.

We assign a set of numbers $(x_1, \ldots, x_n)$ to each enterprise, where x_1 is the value of the first index, x_2 the value of the second index, and so on.

A pair of numbers (x_1, x_2) defines a point in space, and an n-tuple $(x_1, x_2, \ldots, x_n)$, by definition, is considered a point in n-dimensional space.

Thus, our problem boils down to a problem of classifying points in n-dimensional space.

It's hard without computers

To solve classification problems, various algorithms for computer processing have been developed. The need for computers arises for two reasons. First, the number of objects to be classified is usually very large, and it's simply impossible to process all the data "by hand." Second, the objects are usually multidimensional.

If there are only two coordinates, as in our example with students, and the number of objects isn't too large, a human being can compete with a computer in solving the classification problem. The "human" approach is visual: we can simply look at the picture with the points representing our objects and isolate the regions where the points are more dense. Experiments show that the results of such partitioning are more or less the same

[1]Although this article originally appeared during the Soviet era of centralized planning for entire industries, a similar kind of planning occurs in large US corporations, especially those that have acquired a number of smaller companies in a given industry.—*Ed.*

Art by Dmitry Krymov

for all persons tested. The reason is that people involuntarily want to bunch the close points together in one group and spread out the different groups fairly far from one another. But if the number of parameters is three or more, a visual representation is practically useless, and the classification becomes difficult for a human being.

Psychologists conducted the following experiment. The persons being tested were given sets of cards with number triples representing space coordinates of a number of points. They were required to split the points into two natural groups. There was in fact a plane tilted with respect to the coordinate axes that divided the points into two groups such that the distance from each point to the plane was large compared to distances between points in either group. Most people divided points by the value of one of the coordinates, which led to unnatural partitioning. Almost no one managed to construct the correct partitioning.

So special algorithms were developed for classifying points in multidimensional space. One of the criteria for correct operation of these algorithms is the requirement that in classifying points on the plane they produce a "natural" partitioning—that is, the kind that people come up with.

The minimum length tree

Before describing one of these classification algorithms, we'll now describe an algorithm for constructing a connected system of segments joining given points with minimal total length. (A system of segments is called connected if, starting from any of their endpoints and moving along the segments of this system, you can reach any other endpoint.) To make it easier to visualize, let's consider a construction in the plane. The algorithm is the same for multidimensional space, but we can't draw the system.

Imagine we're given N points A_1, …, A_N in the plane. Suppose for simplicity that all the distances between these points are different. Write down all the pairs (A_1, A_2), (A_1, A_3), … , (A_{N-1}, A_N) and arrange them in increasing order of distance between points in a pair. Join the first pair, then the second pair, and so on. If a given segment completes a *cycle* (that is, if it's possible to isolate from the segments already drawn several segments that form a closed polygonal path), we erase it and pass to the next one. And so we continue, up to the longest segment. The system of segments thus obtained has no cycles. (A connected system of segments without cycles is called a *tree*.)

Problem 1. Prove that the system of segments constructed above has the smallest total length among all connected systems of segments joining the given points.

Problem 2. Prove that the same system of segments can be obtained by the following *dual construction*: draw all the segments and arrange them by length in descending order. Erase the longest segment, then the second longest, and so on. If deleting of the next segment destroys the connectedness of the entire system, leave it in place and pass to the next one. Keep doing this until you reach the shortest segment.

Problem 3. Suppose that some of the distances between points A_1, …, A_N are equal. Arrange each group of segments of the same length in arbitrary order and apply the construction described in problem 2. Prove that the system of segments obtained will have the minimum total length regardless of the order of the equal segments. (The minimal system of segments is not necessarily unique in this case.)

The algorithm described above for constructing the tree of minimal length is relatively simple but requires an extensive search and thus considerable computer time. Faster algorithms exist, but they're also more complicated.

Partition into groups

Now that we know how to construct the minimal-length tree, we will use it to describe an algorithm for partitioning. Essentially, after we construct the minimal-length tree Γ, the set A_1, …, A_N is partitioned into groups by deleting some of the segments from this tree. It's natural to delete the longer segments, but in such a way that the points in any of the groups obtained are arranged as densely as possible. This intuitive idea can be formalized by introducing the following values.

Suppose we want to partition the set $\{A_1, …, A_N\}$ into $k + 1$ groups. Choose an arbitrary k segments of the tree Γ and delete them (fig. 1). We get $k + 1$ connected groups of points $\Gamma_1, …, \Gamma_{k+1}$. For each of the groups calculate the ratio of the total length of the segments in the group to the number of segments—that is, the average length of a segment in this group. If a group consists of one point only, the corresponding average length is zero by definition. Denote these average lengths by $l_1, l_2, …, l_{k+1}$ and the lengths of the deleted segments by $b_1, b_2, …, b_k$.

Consider the value

$$F = l_1 + l_2 + \ldots l_{k+1} - b_1 - b_2 - \ldots - b_k.$$

Clearly the denser the points are in each group and the more distant the groups are from one another, the smaller F is. Therefore, one possible algorithm for our sorting task could be as follows: take the minimal-length tree, delete k segments from it in every possible way, calculate the corresponding values F, and choose the partition with the smallest F. If there are several "minimal" partitions, take any one of them.

In algorithms that are actually used, the value F is usually defined in a more complicated way.

Computational experience has shown that algorithms like the one

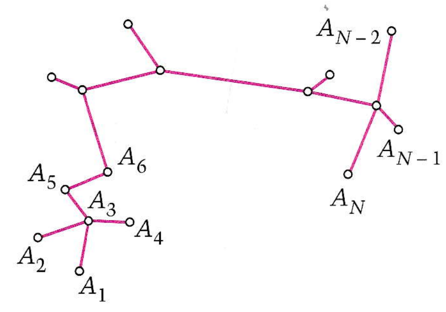


Figure 1

described above produce sufficiently reasonable partitions. The main drawback of these algorithms is they require too much search time. As the number of points grows, the problem becomes inaccessible even for modern computers.

To cope with this difficulty, we can use the following idea. We try to include points separated by a distance less than a certain a in one group. The entire construction is performed in two steps. In the first step, the set $A_1, \ldots, A_N$ is divided into smaller groups $G_1, \ldots, G_m$ such that each group G_i fits in a circle S_i of radius $R = a/2$ and the points of other groups lie outside S_i. (We'll explain later how to do this.) In the second step, only the centers $O_1, \ldots, O_m$ of the circles $S_1, \ldots, S_m$ are considered. The minimal-length tree is constructed for the set $O_1, \ldots, O_m$. Minimizing the value F corresponding to this tree, we divide the set $O_1, \ldots, O_m$ into groups as described in our first algorithm. This produces a partition of the set $A_1, \ldots, A_N$: two points A_i and A_j are put in one group if they belong to the same small group G_l, or if the centers of the circles corresponding to their respective small groups belong to the same group in the partition of the centers $O_1, \ldots, O_m$. This two-step process decreases the number of points to which we must apply our first algorithm: the number m is usually much smaller than N, and the calculation can now be done by a computer.

The problem of the moving circle

It remains to discuss how the set of points $A_1, \ldots, A_N$ can be divided into the smaller groups $G_1, \ldots, G_m$. One algorithm for solving this problem is called the "trout" (it is said to resemble a method of catching trout). Again, we'll describe this algorithm for the plane, although it works equally well in multidimensional space, too.

Suppose that N points $A_1, \ldots, A_N$ are given in the plane. Place small balls of unit mass at these points and draw an arbitrary circle S_0 of radius R containing at least one ball.

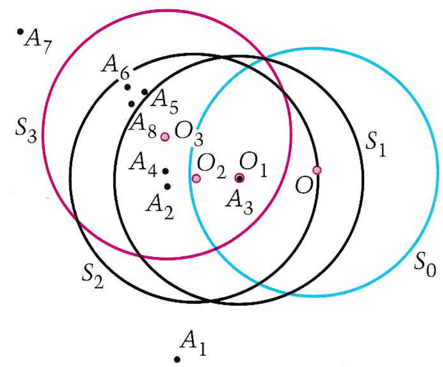


Figure 2

Let O_1 be the center of mass[2] of the balls lying in S_0, and let S_1 be a circle with center O_1 and radius R. Further, denote by O_2 the center of mass of the balls lying in S_1 and by S_2 the circle with center O_2 and radius R, and so on.

For instance, in figure 2 you see eight points $A_1, \ldots, A_8$. The initial circle S_0 covers only one point, A_3, which becomes the center O_1 of circle S_1. Circle S_1 covers points A_2, A_3, A_4, and the center of mass of these three points coincides with the center of circle S_2. We continue the construction similarly, and when we come to circle S_3, we find that it contains the points A_2, A_3, A_4, A_5, A_6, A_8, and its center coincides with their center of mass, so all the subsequent circles coincide with it. Later we'll show that this is the general situation—that is, for any points $A_1, \ldots, A_N$ and any initial circle S_0 containing at least one of these points, all the circles will coincide after a certain step. In other words, movement of the circle along the route $S_0 \to S_1 \to S_2 \to \ldots$ can't go on forever. The points covered by the last circle will be assigned to the first group. Then the same procedure is applied to the remaining points—that is, we "launch" a new circle, whose final position determines the second group. The process is repeated until all the points are assigned to one of the groups.

[2]In any coordinate system, the *center of mass* for a set of unit masses located at the points (x_1, y_1), (x_2, y_2), $\ldots$, (x_n, y_n) is, by definition, a mass of n units located at the point $([x_1 + x_2 + \ldots + x_n]/n, [y_1 + y_2 + \ldots + y_n]/n)$.

Why does the circle stop?

In the remainder of this article we'll prove that our circle indeed "stops." Looking at figure 2 we notice that the circles S_0, S_1, S_2, ... cover an ever growing number of points, and that their final position corresponds to the greatest accumulation of points. It would be natural to suppose that for any points $A_1, \ldots, A_N$ and circle S_0 the number of points covered by S_0, S_1, S_2, ... at least does not decrease. However, figure 3 shows that this is not true. But in a certain sense the density of points in the circles does indeed grow—rather than looking at the number of points falling in the circle, we need to take into account how close to the circle's center they lie.

To be specific, let the points $A_{i_1}, \ldots, A_{i_k}$ lying in a circle S with center O. We'll define the value

$$F(S) = \left[R^2 - \left(OA_{i_1}\right)^2\right] + \cdots + \left[R^2 - \left(OA_{i_k}\right)^2\right]$$

The closer a point is to the center, the greater its contribution to the value $F(S)$. Let's prove that when we construct the circles S_1, S_2, S_3, ..., the sequence of values $F(S_0)$, $F(S_1)$, $F(S_2)$, ... increases until there are no new circles. To this end we'll need the notion of rotational inertia and the Steiner Theorem, which is useful in many problems of geometry and mechanics.[3]

The *rotational inertia* of point masses $A_1, \ldots, A_m$ relative to a point

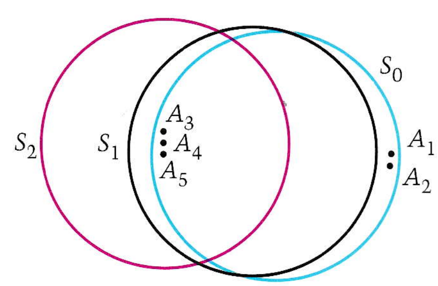


Figure 3

[3]Actually, in mechanics the rotational inertia is defined with respect to an axis rather than a point.—*Ed.*

A is defined by the formula

$$I(A) = (AA_1)^2 + \ldots + (AA_m)^2.$$

The Steiner Theorem allows one to calculate $I(A)$ if the rotational inertia $I(C)$ of this system of points relative to their center of mass C is known. The theorem says that

$$I(A) = I(C) + m(CA)^2.$$

Let's prove this theorem. Place the origin of a coordinate system at the center of mass of points $A_1, \ldots, A_m$—that is, at point C. Then $I(C) = x_1^2 + x_2^2 + \ldots + x_m^2 + y_1^2 + y_2^2 + \ldots + y_m^2$. Let (x_i, y_i) be the coordinates of the points A_i. Then the coordinates of their center of mass C are $(x_1 + \ldots + x_m)/m$, $(y_1 + \ldots + y_m)/m)$. So, because of our choice of origin, we have $x_1 + \ldots + x_m = y_1 + \ldots + y_m = 0$. Denote the coordinates of A by (x, y). Then

$$\begin{aligned} I(A) &= [(x_1 - x)^2 + (y_1 - y)^2] \\ &= \ldots + [(x_m - x)^2 + (y_m - y)^2] \\ &= (x_1^2 + y_1^2) + \ldots + (x_m^2 + y_m^2) \\ &\quad + m(x^2 + y^2) - 2x(x_1 + \ldots + x_m) \\ &\quad - 2y(y_1 + \ldots + y_m) \\ &= I(C) + m(CA)^2, \end{aligned}$$

because $2x(x_1 + \ldots + x_m) = 2y(y_1 + \ldots + y_m) = 0$. This completes our proof of the Steiner Theorem.

Let's prove now that if the circle S_1 does not coincide with S_0, then $F(S_1) > F(S_0)$. Renumber the points $A_1, \ldots, A_m$ so that the points covered by S_0 but not by S_1 get the numbers from 1 to p; the points inside both S_0 and S_1 get the numbers from $p + 1$ to q; and the points in S_1 but outside S_0 get the numbers from $q + 1$ to r. The circle S_0 now contains the points $A_1, \ldots, A_q$, and S_1 contains the points $A_{p+1}, \ldots, A_r$. Clearly $F(S_0)$ can be expressed in terms of the rotational inertia $I(O)$ of the points $A_1, \ldots, A_q$ relative to the center O of circle S_0:

$$\begin{aligned} F(S_0) &= [R^2 - (OA_1)^2] + \ldots \\ &\quad + [R^2 - (OA_q)^2] \\ &= qR^2 - I(O). \end{aligned}$$

The center O_1 of the circle S_1 is the center of mass of the points $A_1, \ldots, A_q$, so by the Steiner Theorem

$$I(O) = I(O_1) + q(OO_1)^2$$

—that is,

$$\begin{aligned} F(S_0) &= qR^2 - I(O_1) - q(OO_1)^2 \\ &= [R^2 - (O_1A_1)^2] + \ldots \\ &\quad + [R^2 - (O_1A_q)^2] - q(OO_1)^2. \end{aligned}$$

Compare the last formula with

$$\begin{aligned} F(S_1) &= [R^2 - (O_1A_{p+1})^2] + \ldots \\ &\quad + [R^2 - (O_1A_r)^2]. \end{aligned}$$

The right side of the last equation lacks the terms

$$[R^2 - (O_1A_1)^2], \ldots, [R^2 - (O_1A_p)^2], \quad (1)$$

but contains the terms

$$[R^2 - (O_1A_{q+1})^2], \ldots, [R^2 - (O_1A_r)^2] \quad (2)$$

that do not enter into the expression for $F(S_0)$. Now we notice that the points $A_1, \ldots, A_p$ lie outside the circle S_1. Therefore, the values in expression (1) are all negative. On the other hand, the points $A_{q+1}, \ldots, A_r$ lie inside the circle S_1, so the values in expression (2) are all nonnegative. It follows that if we remove the terms of the first group from the expression for $F(S_0)$ and add those of the second group, we can only increase the total—that is, $F(S_0) \le F(S_1) - q(OO_1^2)$, or

$$F(S_1) > F(S_0)$$

if the points O_1 and O do not coincide.

In the same way we can prove that $F(S_{k+1}) > F(S_k)$ if the points O_{k+1} and O_k do not coincide. Now it's not difficult to show that all the circles $S_n, S_{n+1}, \ldots$ coincide after a certain number n. Indeed, by the construction, each point O_m is the center of mass of the points covered by the circle S_m. Consider all the subsets of the set $A_1, \ldots, A_N$ and their centers of mass. The point O_{m+1} is one of these centers. Since there are only a finite number of these centers, some centers in the sequence O, $O_1, O_2, \ldots$—say, O_i and O_j—must coincide. However, we've proved that

$$F(S_i) \le F(S_{i+1}) \le \ldots \le F(S_j),$$

and since $S_i = S_j$, $F(S_i) = F(S_j)$ as well. Therefore, $F(S_i) = F(S_{i+1}) = \ldots = F(S_j)$.

Now recall that the equality $F(S_i) = F(S_{i+1})$ is possible only in the case $S_i = S_{i+1}$. But if two consecutive circles coincide, all the subsequent circles will coincide with them. This means that the movement of our circle $S_0 \to S_1 \to S_2 \to \ldots$ can't go on forever—or, as mathematicians say, the "trout" algorithm converges.

Two new problems

To conclude, we'll pose two more problems. The procedures they incorporate are also used to classify points in multidimensional space. As before, we'll restrict ourselves to the case of the plane.

Problem 4. Suppose that the points $A_1, \ldots, A_N$ are divided into l (nonempty) groups $G_1, \ldots, G_l$. Let $O_1, \ldots, O_l$ be the centers of mass of these groups. Construct a new partition of the set $A_1, \ldots, A_N$ by the following rule: if O_i is the closest of the centers $O_1, \ldots, O_l$ to the point A_k $(1 \le k \le N)$, then this point is assigned to the lth group; if there are several centers at the minimal distance from A_k, we choose from them the one with the smallest number. After discarding empty groups (give an example where empty groups emerge!), renumber the remaining, say, p groups $(p \le l)$. This gives a new partition $G_1^1, \ldots, G_p^1$. Find the centers of mass for these groups and repeat the process with the new centers to obtain a partition $G_1^2, \ldots, G_q^2$ $(q \le p)$, and proceed in the same way. Prove that the partitions will coincide after a certain step.

In the next problem the following notation will be used: for an arbitrary subset G of the set $A_1, \ldots, A_n$ and point A put $I(A, G) = (AA_i)^2 + \ldots + (AA_j)^2$, where the summation is taken over the points of G.

Problem 5. Suppose that an initial partition $G_1, \ldots, G_l$ of points $A_1, \ldots, A_N$ into nonempty groups is given as in the previous problem. The new partition $G_1^1, \ldots, G_p^1$ will differ from the initial one by the "location" of only one point, which is moved into a new group—namely, we move the point A_1 to the group G_i, for which the value $I(A_1, G_i)$ is minimal (if there are several such groups, choose the one with the smallest number). If A_1

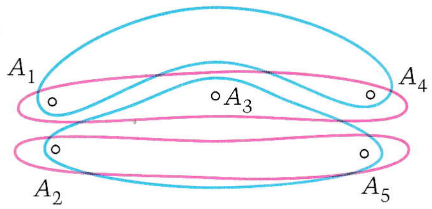


Figure 4

was the only point in its group, which thus has become empty after the move, discard this group and renumber the rest arbitrarily. The number p of new groups is either l or $l - 1$. Repeat the process for the points $A_2, \ldots, A_N$, in order; then again for A_1, for A_2, and so on. Prove that all the partitions coincide after a certain step.

The advantage of the algorithms described in these problems over the "trout" algorithm is that they converge quickly and so need less computing time. However, an unfortunate choice of the initial partition in these algorithms may result in an unnatural final partition, as illustrated in figure 4.

Corrections

May/June 1997

• p. 19, col. 3: the display equations are all off by a factor c. Thus the first equation should read

$$L_0 = G^{1/2}\hbar^{1/2}c^{-3/2}$$

(not $\ldots c^{-1/2}$ as printed). The factor c in the remaining display equations should read $c^{-5/2}$, $c^{5/2}$, and $c^{1/2}$. Our thanks to Flavis Pakianathan, a physics teacher at Methodist High School in Ipoh, Malaysia, for pointing this out.

• p. 38, col. 3: the third display equation should read $\lambda = n\lambda_{\text{film}}$. The wavelength in the film is less than that in air; therefore, we need to multiply by the index of refraction n rather than divide by it. (Thanks again to Flavis Pakianathan.)

March/April 1997

See page 47 for feedback on "A Planetary Air Brake."

LTG **Coming in September: A special issue of *Quantum* devoted to the 25th Anniversary of the publication of *The Limits to Growth*** 25

Can you see the magnetic field?

How to use your TV in a new and enlightening way

by Alexander Mitrofanov

LET'S DO A SIMPLE EXPERIMENT that's entertaining and quite colorful. All we need is a small permanent magnet—taken from an old toy or some measuring device—and a color TV. Turn on the TV, tune it to some channel, and bring the magnet up to the screen. You'll produce spectacular color changes near the magnet, which are especially striking if the original image on the screen had a large expanse of a single color (fig. 1).

In presence of a magnet, the television screen presents a splendid picture much like the colors in oily films on wet asphalt or the Northern Lights. The colored bands converge near the outline of the magnet and thus make the magnetic field "visible." Try manipulating the magnet and see what happens. Move it toward the screen and away from it, or rotate it, and watch the colors change. In this experiment the "image" of the magnetic field is even more impressive than the pattern of a magnetic field "drawn" by iron filings, needles, or nails (fig. 2), or obtained with "magnetic viewing paper" (which consists of a thin film of oil with suspended ferromagnetic particles spread on a paper backing and covered with clear plastic). Also, the TV screen "senses" rather small fields—smaller than those detected by iron filings or magnetic viewing paper.

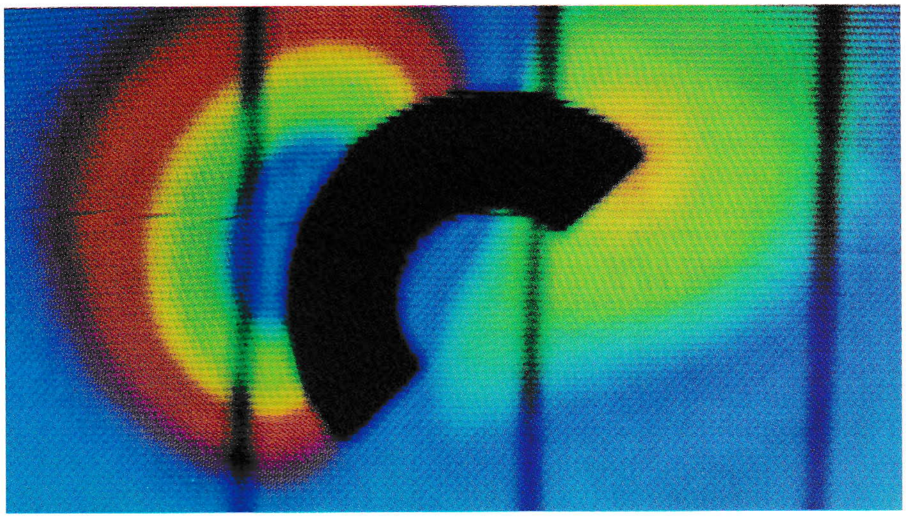


Figure 1
Photograph of the screen of a color TV with a small permanent magnet in front of it.

It's not hard to take a picture of a TV screen colorfully disturbed by a magnet. You don't even need a tripod. The luminescence of an ordinary TV screen is large enough so that, using ASA 100–200 film and opening the diaphragm completely, you need only a 1/15 s or even 1/30 s exposure. Shorter exposures won't do any good (can you say why?), while longer exposures call for a tripod.

What was going on in our experiment? No doubt many of our readers

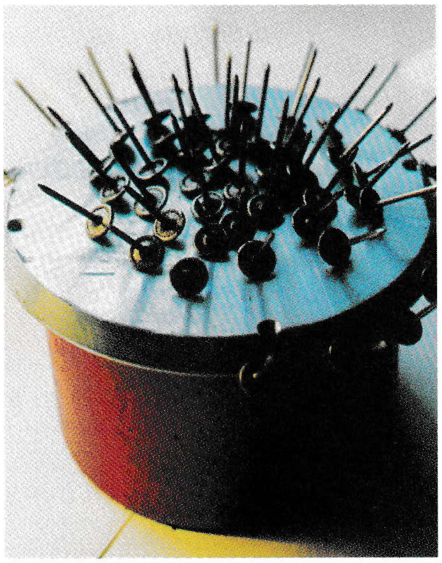
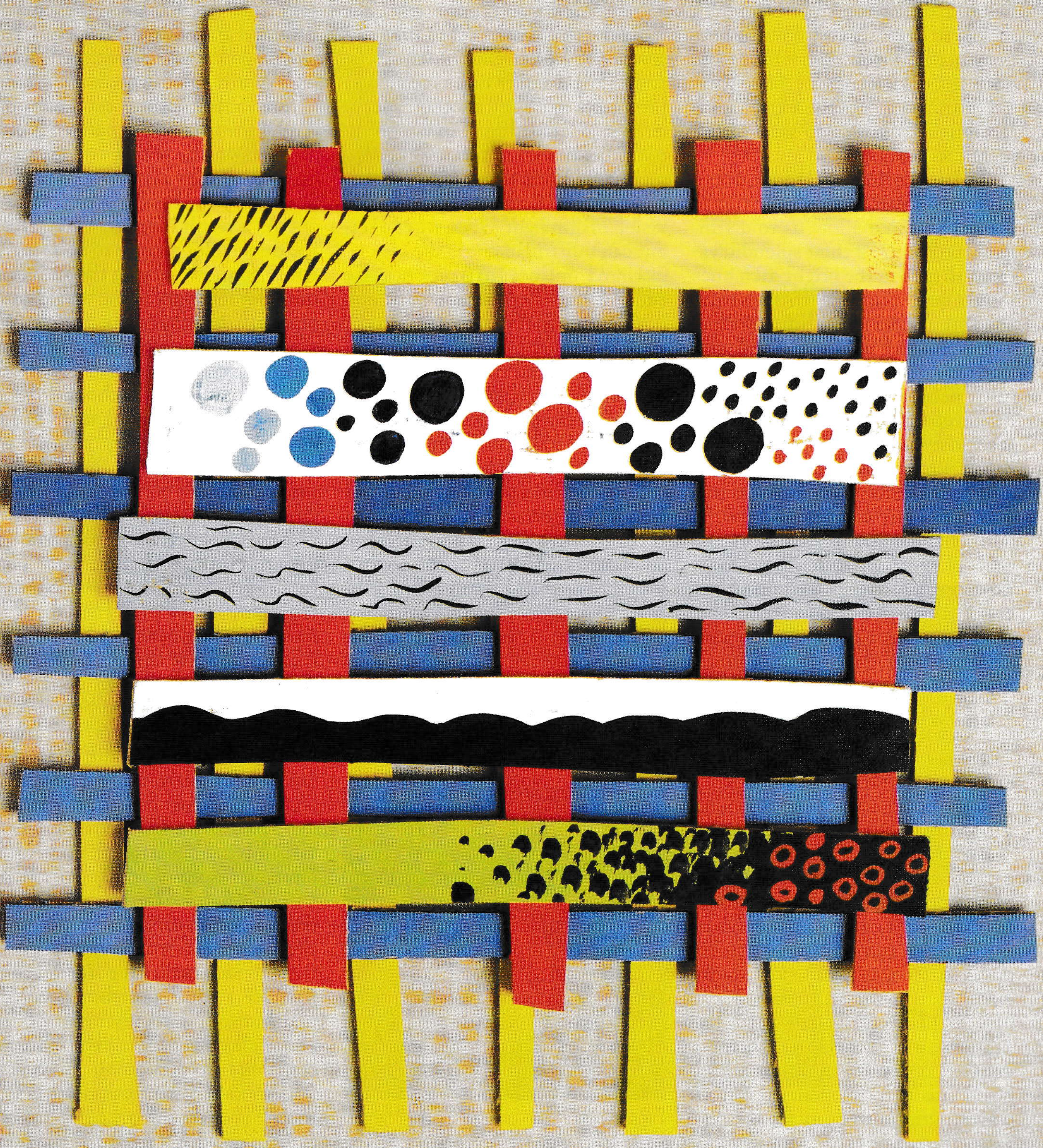


Figure 2
Nails on a magnet.

Art by Vasily Vlasov



know the answer. The explanation is really rather simple. When we bring a magnet up to the TV screen, its magnetic field penetrates the screen and enters the picture tube. The Lorentz force induced by the magnet causes an additional deflection of the electron beams, which leads to color shifts in those parts of the screen where the deflection is great enough. Deflection of an electron beam in a magnetic field is a well-known phenomenon, but what does it have to do with the color changes? This question requires a separate explanation (although we won't go into all the technical details).

We usually don't pay much attention to the wonderful properties of our eyes, which allow us to distinguish colors in all their splendid variety, to take in the bright, rich hues of our surroundings, to sense the subtlest gradations in tint and tone. The wonderful world of light sensations is a common phenomenon in most people's lives. We should be thankful that nature "supplied" us with just this type of vision!

But what do we mean by "color" and "color vision"? The light radiating from many sources, such as the Sun, a filament lamp, an illuminated sheet of white paper, or a fragment of the daytime sky—is composed of a continuous range of electromagnetic waves of different wavelengths. By definition, visible light consists of wavelengths that can be detected by the eyes of most people (leaving aside the phenomenon of color blindness). This light lies in the range of about 380 to 760 nm, which corresponds to the colors violet and dark red. Using a glass prism, diffraction grating, or a set of color filters, we can decompose a more or less homogeneous mixture of rays from any source into the narrow bands corresponding to different wavelengths that we say have "different colors," which our eyes distinguish as red, orange, yellow, green, blue, indigo, and violet—as well as innumerable shades based on these colors. The result of splitting a light beam into its components of various wavelengths or frequencies is called a *spectrum* (from the Latin word for "image"). Color vision is possible because the retina of our eye contains light detectors called cones, which absorb light differently depending on its wavelength. The pigments in the three types of cone have broad light absorption bands, but the maxima of light absorption are located in different spectral regions and correspond to the wavelengths 430, 530, and 560 nm. Cones aren't the only light receptors in the eye. When the light is weak—in a darkened room, say, or at dusk—the cones don't respond to the visible light, and another mechanism is "switched on"—one based on cells called rods. Rods contain the pigment rhodopsin, which is highly sensitive to light. Rods are responsible for vision in dim light, when we can see, but can't distinguish colors.

Using optical instruments, we can not only decompose light and obtain the spectrum of the light source, we can also perform the reversal operation: collect rays of different color together. Mixing different colors can produce some striking results, not at all obvious beforehand and quite unpredictable.

Let's look at a popular demonstration that originated in experiments Newton and Maxwell performed on color mixing. Using three projectors, we illuminate a screen with three partially overlapping beams that pass through different color filters: red, green, and blue. By varying the relative intensity of the beams, we can get white light on the screen where all three beams overlap (fig. 3). Another overlapping area shows that mixing red and green light produces yellow, and doing the same with blue and green produces cyan. We can also say that illuminating a white screen with cyan and red beams produces white light.

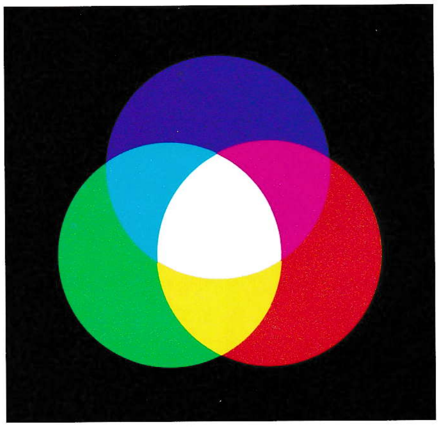


Figure 3
Mixing colors.

Now imagine that spots of red, green, and blue light are located close to one another on the screen. They don't overlap, but due to their very small angular sizes, they aren't resolved individually by the eye. Such a composite pattern will look like a white spot on the screen. This is due to diffraction: each colored spot of the three-color pattern is out of focus on the retina and slightly blurred, so the individual cones are affected by the light of all three colors. As a result, the brain considers the three-color pattern to be a single white spot. By varying the ratios of the color intensity of the beams, and possibly by changing the background, we can observe a colored spot of any color and shade. Painters of the "pointillist" school (for instance, Seurat and Signac) used a similar technique, applying small dabs of unmixed paint that blended into different colors when the canvas was viewed from a distance.

These features of color vision underlie the functioning of the picture tube in a color TV. The screen is composed of many small geometrically identical phosphorescent elements in the shape of circles or strips and collected in groups of three (fig. 4). These phosphorescent elements (cells) have different chemical compositions based on zinc, sulfur, selenium, phosphorus, and other elements, so that they radiate red, green, and blue light when struck by an electron beam. Three electron beams of controlled intensities are focused on the corresponding neighboring cells to produce these three colors in certain proportions, which makes it possible to reproduce a wide spectrum of colors and hues. The resulting image is pretty clear because the phosphorescent cells are very small. However, you can see them either with the naked eye, if you get close to the screen, or with a magnifying glass.

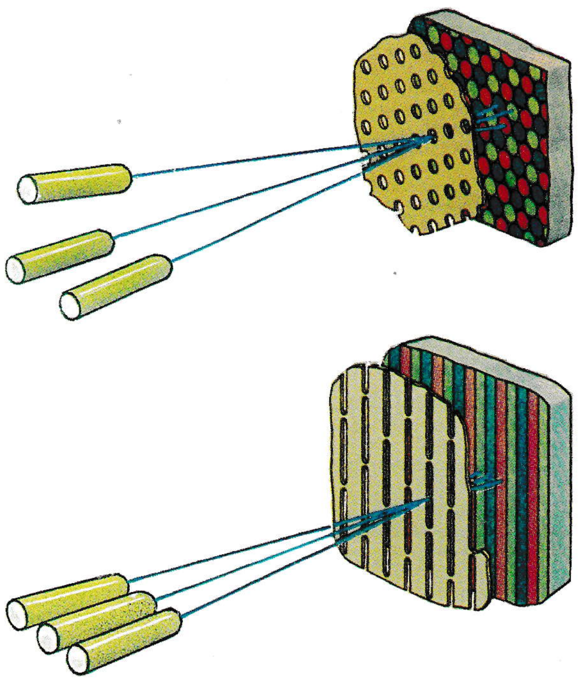


Figure 4
Screen of a picture tube with mosaic and gratelike phosphorescent elements.

They can also be seen in a magnified photo (fig. 5).

Let's look at figure 4 again. In addition to the three-color phosphorescent cells of the screens with mosaic or gratelike structures, this figure shows the corresponding shadow (or aperture) masks. These are thin metal sheets with a large number of regularly placed apertures. The shadow mask is attached behind the screen at a distance of about 1 cm from it. Every picture tube contains three electron guns (fig. 4). During TV transmission all three electron beams are focused by a common magnetic deflection yoke at a certain phosphorescent triad on the screen, but the intensity of each beam is controlled individually according to the original picture as viewed in red, green, and blue light. The mutual disposition of the electron guns, phosphorescent cells, and apertures in the shadow mask is chosen in such a way that a cell of a particular color is exposed only to "its own" electron gun, which is modulated by the video signal responsible for the corresponding color in the resulting picture. The two other electron guns bombard their own phosphorescent cells. Basically this is how a picture tube that uses a color-splitting shadow mask works.

Even this brief description of a color picture tube provides enough information for you to understand why the magnetic field of a small permanent magnet distorts the colors on the TV screen, but hardly affects the shape of the objects pictured. Indeed, when the horizontal component of the Lorentz force, acting on the electron beams from the external magnet placed near the screen, causes a deflection of the beams to distances comparable to the horizontal period of the mask apertures or to the space between the phosphorescent strips, it will not produce a visible deformation of the object's shape, but it will radically disturb the balance of the mixed colors. In the external magnetic field the electrons "go astray" and enter the wrong apertures in the shadow mask, so they bombard the phosphorescent cells of the "wrong" color. When the magnet is removed, the correct colors are restored.

Attentive readers, and in particular those who have seen the shadow mask of a color picture tube, could enrich our explanation with another possible mode of action of the external magnet. The magnet attracts the mask, which is made of soft iron, and the deformed mask begins to allow "foreign" electrons to pass through.

So, just to be safe, you should only use small magnets (1–2 cm^3) when you experiment with the TV screen. A larger magnet might break the screen, or the shadow mask might become permanently deformed, causing the electron guns and other parts of the TV to go out of tune.

To conclude, I'll leave you with an amusing story that the great Russian physicist P. L. Kapitsa used to tell his students. A certain naval captain, the commander of a battleship, arrived in Moscow from the Far East and visited the Russian

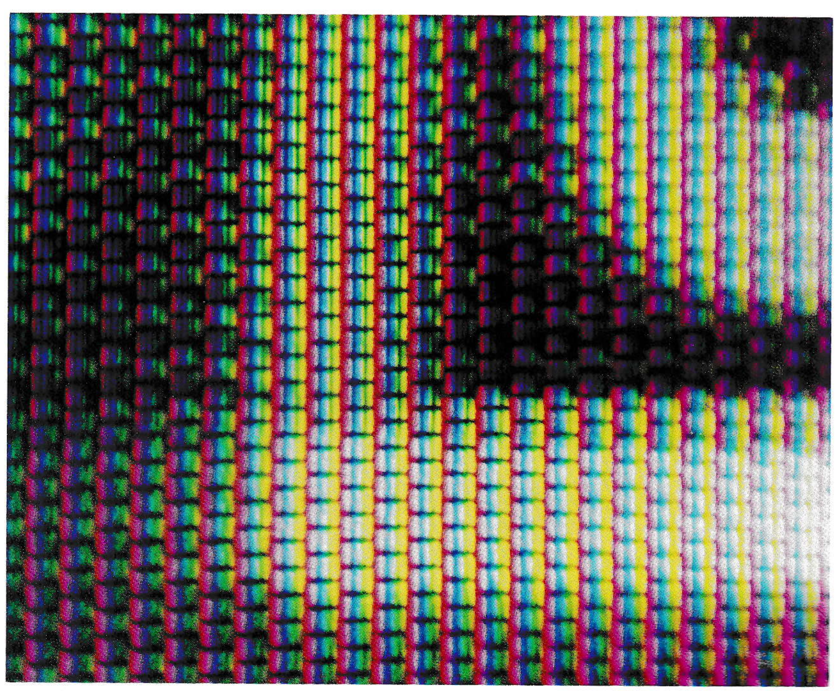


Figure 5
Photographic enlargement of a portion of a color TV picture tube.

Academy of Sciences. He claimed to have invented a new type of magnet. According to him, the magnet had only one pole—the north pole! The captain had a letter with him from his superior, an admiral, who asked the scientists to examine the invention and to give an expert opinion. Was it an important discovery?

The magnet looked absolutely ordinary: just a bar of metal of about 1 kg painted red. Lo and behold, both ends were really north-seeking! Kapitsa, to whom the captain was sent, quickly figured it out: the magnet was composed of two identical magnetized bars whose south-seeking poles were skillfully glued together. The paint covered up the joint. Kapitsa asked the captain why he pulled such a stunt. It turned out that the captain had never been to Moscow, although he dreamed of it all his life. His superior wouldn't grant him a leave of absence—this was the only way he could think of to visit the capital!

What if the admiral had a color TV set and a copy of this issue of *Quantum*? Would he have sent the captain to Moscow, or to some more remote place? Guess how a TV set could help the admiral to unravel the puzzle of the captain's magnet. While you're at it, try to answer the following questions.

1. A rainbow includes the entire spectrum of visible light. So why is the color brown absent?

2. Mixing yellow and blue paint produces green. If yellow and blue beams from projectors equipped with filters are mixed on a screen, the overlapping area is white. Why does the mixing of the same colors produce such different results?

3. How is the color black produced on the screen of a color TV? Why does the screen of the TV when turned on often seem darker than when the TV is turned off?

4. When there is a full moon, you can see many objects outside at night. Their colors, however, are far from what you see in the daytime. A similar phenomenon can be observed in the experiment with a color TV: if a balanced color image is weakened by a dark, spectrally neutral filter, red and green tints will disappear and the image will become grayish blue. Can you explain this?

5. How can one know whether an external magnet deflects the electron beams or simply attracts and deforms the shadow mask of a color TV's picture tube?

6. Are electrons deflected toward the right or left part of a TV screen due to the Earth's magnetic field? Say the TV set is located (a) in New York City, (b) in New Orleans, (c) at the equator, (d) in the south of Australia.

7. Estimate the shift of the electron beam in a TV picture tube under the influence of the Earth's magnetic field. The energy of the beam is 25 keV, and the tube is 0.2 m long.

8. Under the action of an external, variable magnetic field the image of objects on the screen of a color TV set readily change color, but their shapes seem immune to the magnetic perturbations. Why?

9. Is it possible, using a color TV set as an indicator, to find magnetic objects inside an opaque package?

10. How can one determine the charge-to-mass ratio of the electron using a TV set? ◘

Challenges in physics and math

Math

M206

Double roots. Find all a such that both of the following equations have two integer roots:

$$x^2 + ax + 1996 = 0$$

and

$$x^2 + 1996x + a = 0.$$

(V. Protasov)

M207

Points on a leg. Four points K, P, H, and M are taken on a side of a triangle. These points are the midpoint, the endpoint of the bisector of the opposite angle, the point of tangency with the inscribed circle, and the base of the corresponding altitude, respectively (fig. 1). Show that if $KP = a$ and $KM = b$, then $KH = \sqrt{ab}$. (I. Sharygin)

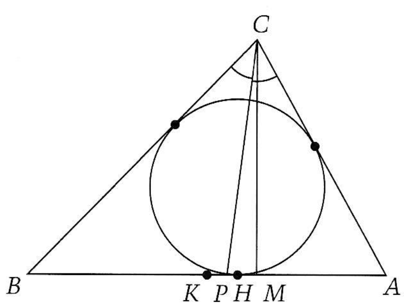


Figure 1

M208

Small as possible. Find the least possible value of the expression

$$y = \frac{x^2}{8} + x\cos x + \cos 2x.$$

(D. Averiyanov)

M209

Old-fashioned approach. With a calculator you can discover that the equation $x^3 - x - 3 = 0$ has a unique real root, and that this root is greater than $\sqrt[5]{13}$. But can you prove that this conjecture is in fact true? (V. Panfyorov)

M210

Thirty degrees less. In triangle ABC, bisectors AA_1, BB_1, and CC_1 of the interior angles are drawn (fig. 2). Prove that if $\angle ABC = 120°$, then $\angle A_1B_1C_1 = 90°$. (A. Yegorov)

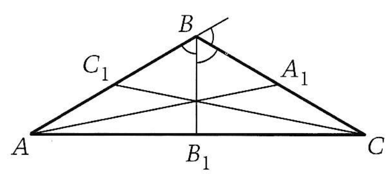


Figure 2

Physics

P206

Box with a spring. A mass m oscillates on the end of a spring hung from the top of a box of mass M placed on a table. At what amplitude of oscillation will the box jump from the table? The spring constant is k. (L. Bakanina)

P207

Gas expansion. One mole of ideal monatomic gas expands from an initial volume of 20 l to a final volume of 200 l. During this process the pressure in the gas cylinder varies according to the table in figure 3. Does this gas take in or give off heat when it expands from 40 l to 80 l? Does it get cooler or warmer when it expands from 140 l to 180 l? Find the ratio of the specific heats for these regions. (A. Zilberman)

V (l)	20	40	60	80	100	120	140	160	180	200
P (kPa)	100	35.4	19.2	12.58	9	7.8	6.95	6.3	5.75	5.3

Figure 3

P208

Two fluid films. Two fluid films are formed on the surface of a liquid and separated by a movable rod of length l (fig. 4). The coefficients of surface tension of the films are σ_1 and σ_2. What force must be applied to the rod to keep it from moving? (A. Buzdin, S. Krotov)

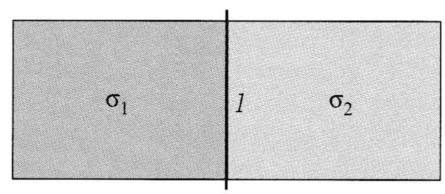


Figure 4

P209

Round capacitor. An insulated concentric spherical capacitor with internal and external spheres of radii R_1 and R_2 has a charge Q. Find the energy density of the electric field in the space between the spheres if $R_2 - R_1 \ll R_1$. (V. Mozhayev)

P210

Where's the star?. Estimate the error in measuring the angular coordinate of a star visible from the Earth at an angle of $\beta = 45°$ above the horizon. The refractive index of the air at the Earth's surface is $n = 1.0003$. (S. Gordyunin, P. Gorkov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

Seeing is believing

Visual proofs of the Pythagorean theorem

by Daniel J. Davidson and Louis H. Kauffman

THE PYTHAGOREAN THEOREM is well known to most high school students. It says that if a and b are the legs of a right triangle and c is the hypotenuse, then $a^2 + b^2 = c^2$. Elisha Scott Loomis, in his book *The Pythagorean Proposition* (Washington: NCTM, 1968), records no less than 367 proofs. Of these, 109 are algebraic and 255 involve geometric construction. We will first look in detail at Loomis's demonstration number 9. Loomis credits Henry Perigal for its publication in the journal *Messenger of Mathematics* (1873, Vol. 2, p. 104).

The Perigal proof

We begin with a square and dissect it with two cuts perpendicular to each other, meeting at the center of the square. Either cut can be regarded as the hypotenuse of a right triangle obtained by dropping a perpendicular from one end of the cut at A to the opposite side of the square at B (see figure 1a).

Problem 1. Show that any right triangle can be produced by choosing a suitable square, a suitable line AB, and dropping perpendicular AC.

Solution. The angle ABC can range from 45° to 90°, if we rotate AB about point O (with a degenerate case at 90°). This lets us construct a triangle ABC that is similar to any given right triangle. Choosing a square of the right size will allow us to make triangle ABC congruent to any given right triangle.

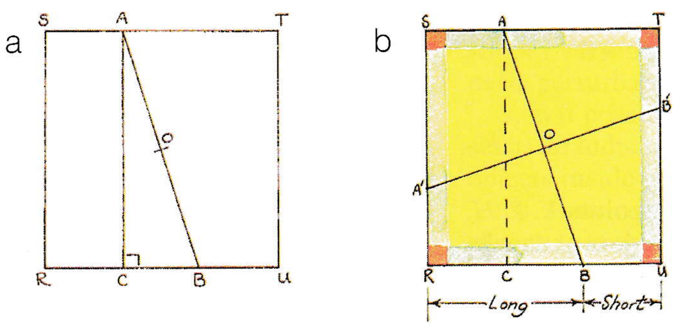


Figure 1

Our original square $STUR$ can be regarded as the square of the larger leg AC of triangle ACB. Our dissection with two cuts leaves each side of the square divided into segments $RB = l$ (for longer segment) and $BU = s$ (for shorter segment).

Problem 2. Show that all four pieces into which AB and $A'B'$ dissect the square are congruent.

In our dissected square, $l = RB = UB' = TA = SA'$, and $s = BU = B'T = AS = A'R$. Since the base of our triangle has length BC, and since $AS = RC = s$, we have $BC = l - s$. So the length of side BC of the right triangle is equal to the difference between the long and short segments of the dissection piece.

Turning to figure 2, we see that we can rearrange the four pieces of the original square to produce a new, larger square with a small square "captured" in its center. This small square has a side equal to $l - s = BC$, so it is the square on the shorter leg of the triangle. The large new square (with the hole in it) is just the square on the hypotenuse of our triangle, and the yellow area is equal to the square on the longer leg. Thus the square on the hypotenuse is equal to the sum of the squares on the two legs. This completes Perigal's proof.

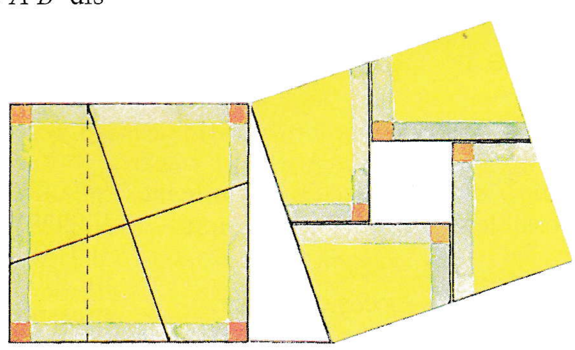


Figure 2

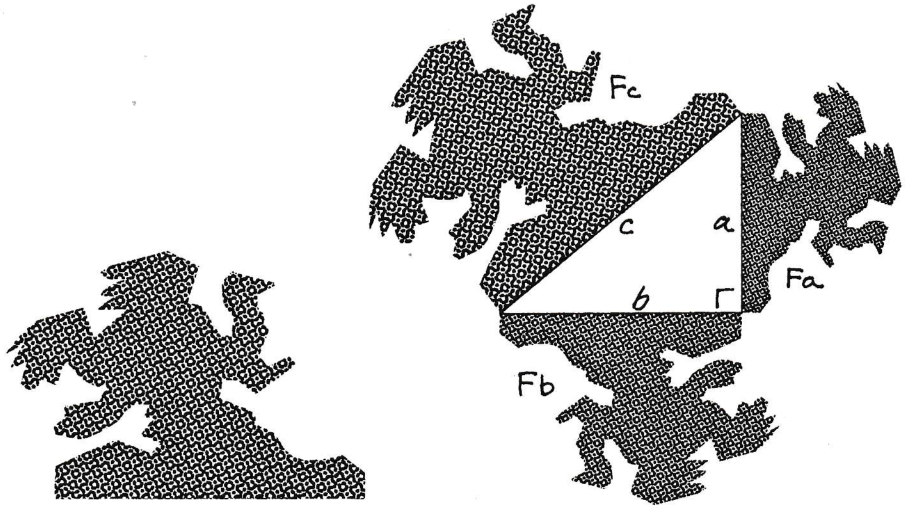


Figure 3

Symmetries and movement

Compared to other dissection proofs, the Perigal proof is both simple and unique. It's simple because it requires only two dissection lines. It's unique in that it is the only dissection proof recorded by Loomis that does not dissect the squares on both legs of the triangle. The squares on the smaller leg and on the hypotenuse are created by "rearranging" the four congruent parts of the square on the larger leg (labeled 1, 2, 3, 4 in figure 5). As it turns out, this rearrangement can be done in three different ways: by translations of the pieces across the center of the square; by rotation of each part in its own corner; or by turning each part over to its other side and translating to the opposite corner (a glide-reflection). Figure 5 shows these three types of symmetric transformations in the Perigal proof.

To highlight the differences, we have redrawn figure 5 as figure 6,

A generalization

Instead of putting a square on each side of a right triangle, suppose we choose some irregular shape with at least one straight side (for example, the shape labeled F in figure 3). Now take three scaled versions of F, labeled F_a, F_b, and F_c, with the straight side scaled up or down to match sides a, b, c of our right triangle.

THEOREM. *Area*(F_c) = *area*(F_a) + *area*(F_b).

Proof. The argument depends on the fact that the areas of similar figures are proportional to the squares of any linear measurement (a discussion of this fact can be found in any geometry text). So if the ratio of the area of F_a to a^2 is k (so that area$(F_a) = ka^2$), then area$(F_c) = kc^2$, and area$(F_b) = kb^2$. Then, since $a^2 + b^2 = c^2$, it follows that $ka^2 + kb^2 = kc^2$, which is what the theorem states.

It may seem a bit startling that this wide generalization of the Pythagorean theorem comes so easily from a simple fact about areas of similar figures. Indeed, this simple fact gives a direct proof of the Pythagorean theorem itself.

Consider right triangle RST in figure 4, with hypotenuse RT and right angle at S. Drop perpendicular SD from S to the hypotenuse. Then the triangles RSD, STD, and RST are all similar. As in the previous proof, let the ratio of the area of triangle RSD to a^2 (where a is the length of its side RS) be k, so that area$(RSD) = ka^2$. Then area$(STD) = kb^2$, and area$(RST) = kc^2$. But it is clear from the diagram that area(RSD) + area(STD) = area(RST), so $ka^2 + kb^2 = kc^2$, or $a^2 + b^2 = c^2$.

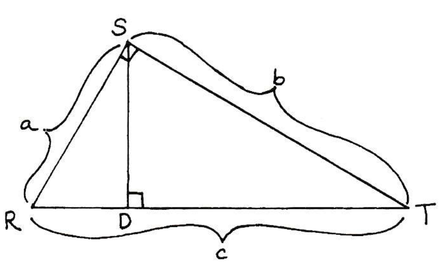


Figure 4

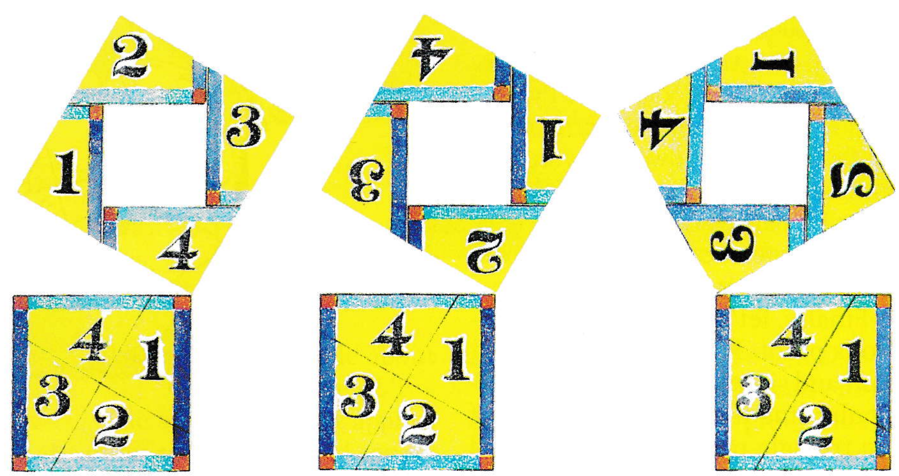


Figure 5

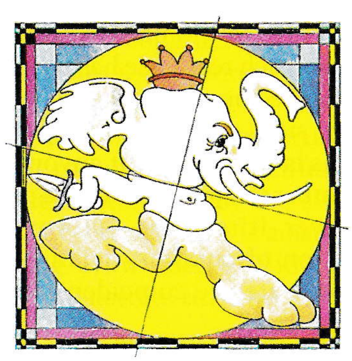


Figure 6

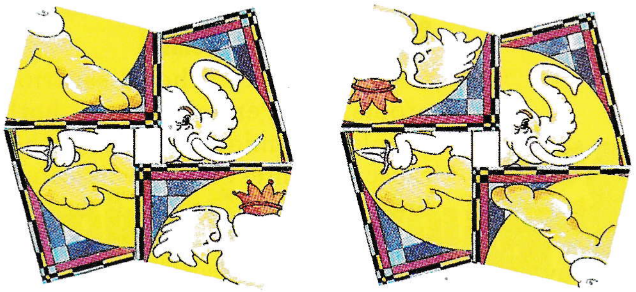


Figure 7

introducing a picture of the elephant-faced Hindu god Ganesha, striding across the square. This gives each piece of the square a unique appearance, allowing us to see the work of each of these three moves.

Figure 7 shows two ways of rearranging the pieces of the square.

Problem 3. Identify the transformation that each piece has undergone.

Problem 4. Prove that the "captured" space (the white region in the center) is a rectangle, and find its dimensions in terms of l and s.

Other proofs and the classic proof

Figures 8 through 14 illustrate other dissection proofs of the Pythagorean theorem. We leave it to the reader to explain how they were constructed and how they furnish a proof of the Pythagorean theorem. We have provided references to Loomis's book, in case a hint is needed.

Finally, let's consider Euclid's famous proof. He begins with a brilliant and simple idea: drop a perpendicular line from the right angle of the triangle to the hypotenuse and continue this line so that it bisects the hypotenuse square into two rectangles (L and R—see figure 15). Then Euclid proves the utterly fantastic fact that each rectangle has the same area as the corresponding square on the leg sides of the right triangle.

This is so simple that anyone can remember the proof up to this point. However, it requires a clever construction of auxiliary triangles to demonstrate the coincidence of areas (fig. 16).

Triangle ADE has base AD and height AB. Thus the area of triangle $ADE = {}^1/_2(AB)(AD) = {}^1/_2\text{area}(L')$. But

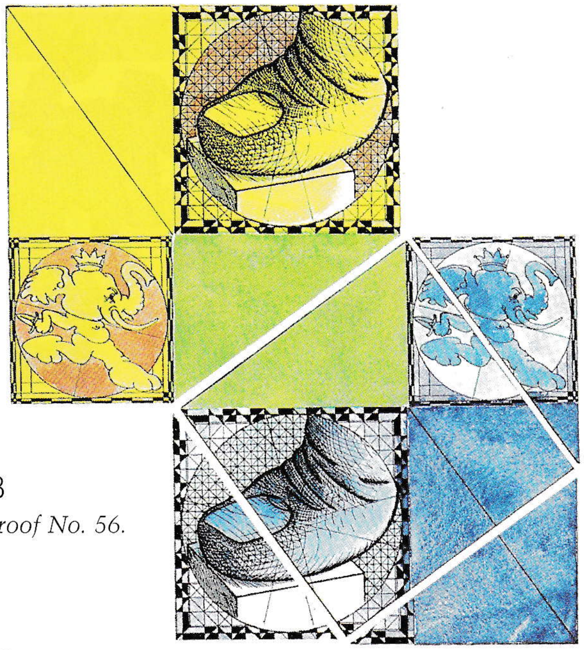


Figure 8

Loomis proof No. 56.

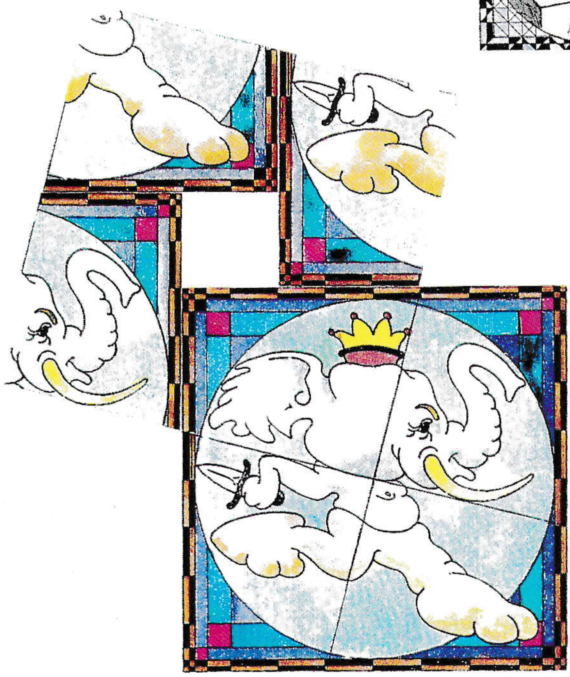


Figure 9

Loomis proof No. 205 (also credited to Henry Perigal).

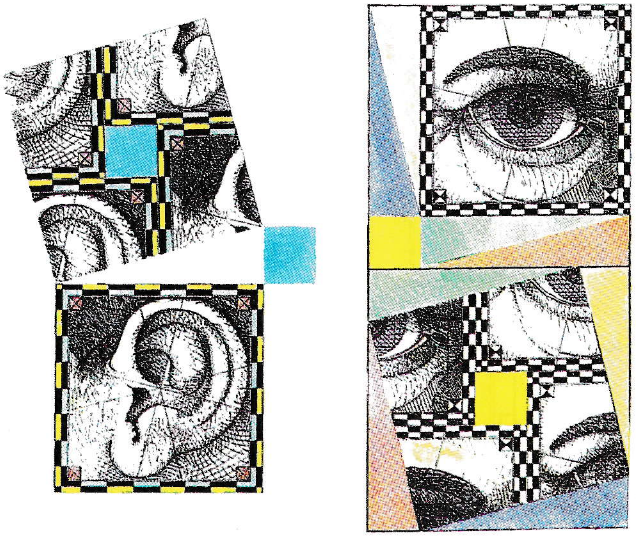


Figure 10

The Chinese and Perigal proofs placed side by side. They look simple—but who would guess that they prove the same theorem?

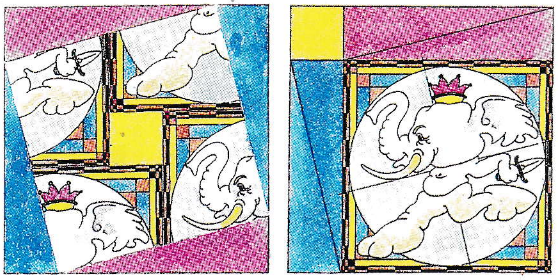


Figure 11
Loomis proof No. 91 (attributed to Chinese, Persian, and Indian authors as well as to Pythagoras himself).

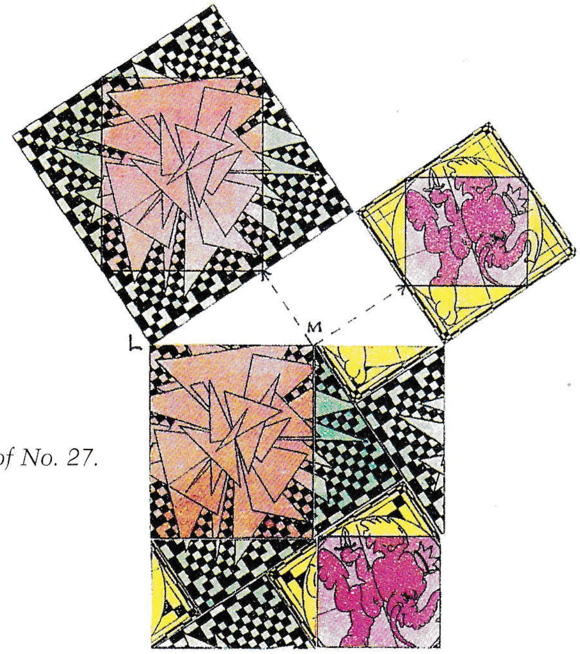


Figure 12
Loomis proof No. 27.

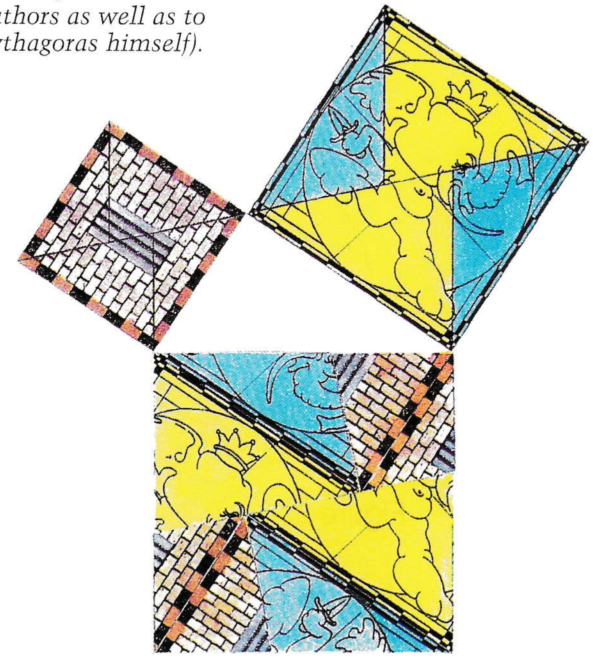


Figure 13
Loomis proof No. 19.

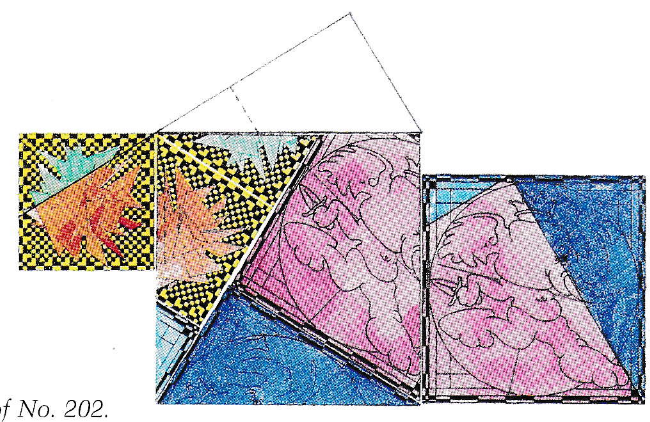


Figure 14
Loomis proof No. 202.

triangle GDH has base GD and height DE. Thus the area of triangle GDH = $^1/_2$area(L). *But* triangle GDH is congruent to triangle EDA! (Check this by rotation in either triangle about vertex D.) Therefore, area(L') = area(L). The same argument applies to the other rectangle R' and square R. This gives us Euclid's *Q.E.D.*

Euclid's proof is technically more complex than Perigal's proof and the other dissection proofs shown here. But it turns on a very simple idea in the argument (drop a perpendicular) and another neat idea (rotate a triangle to find congruence).

Problem 5. A theorem in geometry states that either leg of a right triangle is the mean proportion between the whole hypotenuse and the segment cut off by the altitude to the hypotenuse that is adjacent to that leg. How does Euclid's proof also demonstrate this theorem?

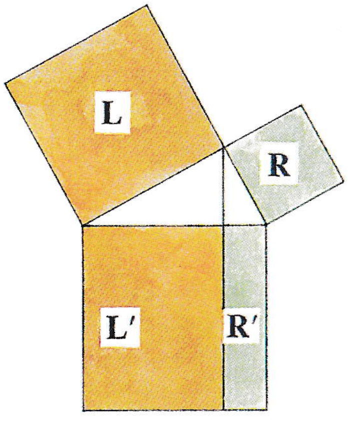


Figure 15

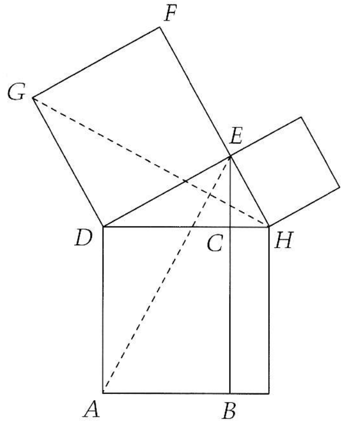


Figure 16

Daniel J. Davidson *is an artist living in Chicago, Illinois.* **Louis H. Kauffman** *is a professor of mathematics at the University of Illinois–Chicago.*

The complete qu

How did you manage with incom

by I. F. Shary

A "COMPLETE QUADRILATERAL" is a figure formed by four straight lines on the plane none of which are parallel and no three of which pass through the same point (see figure 1). The points where these lines meet are called the vertices of the complete quadrilateral. Thus any complete quadrilateral has six vertices. Segments that connect two vertices of a complete quadrilateral and do not belong to any given lines are called diagonals of the quadrilateral. Thus it has three diagonals.

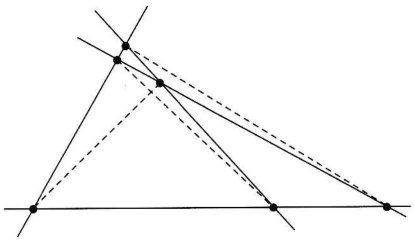


Figure 1

Complete quadrilaterals possess a number of peculiar properties. Here are some of them.

1. The midpoints of the diagonals of a complete quadrilateral belong to one straight line m (fig. 2). In Germany and Russia this line is called *Gauss's line*, while in England they call it *Newton's line*. (It seems likely that what we have here is an attribution, rather than real authorship. The light from huge stars often makes the twinkling of small stars invisible.)

2. If all the lines forming the quadrilateral touch one circle, then Gauss's (or Newton's) line contains the center of this circle (fig. 3).

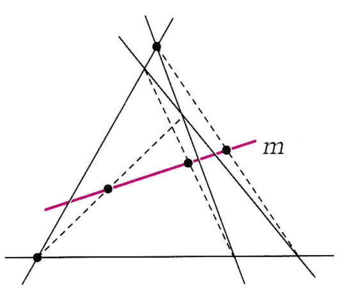


Figure 2

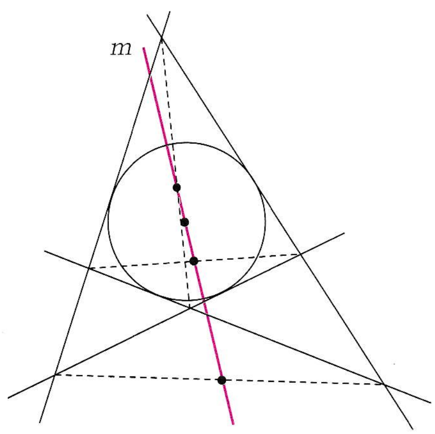


Figure 3

On this point the old geometry books are in better agreement, calling this statement *Newton's theorem*.

3. Any three lines of a complete quadrilateral form a triangle. There are four such triangles. Many properties of complete quadrilaterals are connected with wonderful points in these triangles. For example, it turns out that their orthocenters belong to one straight line p, and that this line is perpendicular to Gauss's line! (This is illustrated in figure 4, which shows two out of the three altitudes of each triangle.) In the Russian mathematical literature this line p is sometimes called *Van Oebel's line*.

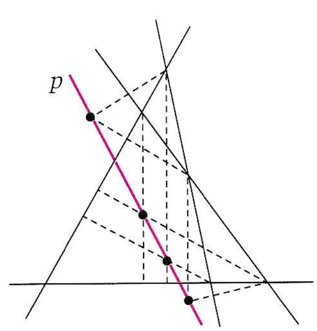


Figure 4

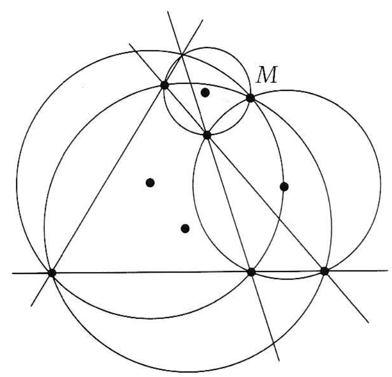


Figure 5

4. Four circles circumscribed about four triangles that make up a complete quadrilateral (see the previous paragraph) meet at one point M. This point is called *Miquel's point* (fig. 5).

5. The centers of the four circles mentioned in the previous paragraph lie on another circle (fig. 6).

6. If four vertices of a complete quadrilateral belong to one circle, then the Miquel's point lies on the

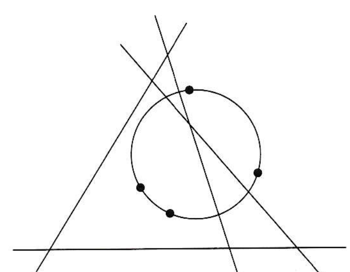


Figure 6

quadrilateral

complete ones all these years?

Sharygin

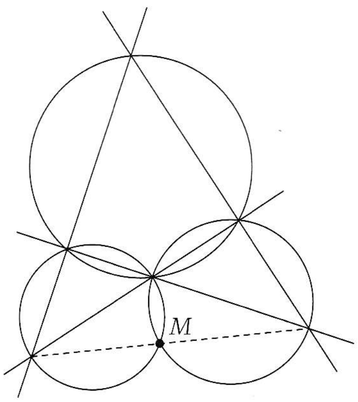


Figure 7

diagonal connecting the two remaining vertices (fig. 7).

7. If four vertices of a complete quadrilateral belong to a circle, then the center of this circle and the two remaining vertices of the quadrilateral form a triangle whose orthocenter coincides with the point where diagonals connecting the four vertices lying on the circle meet (fig. 8). Note that we might have to extend the diagonals beyond the vertices of the quadrilateral in order to find their point of intersection.

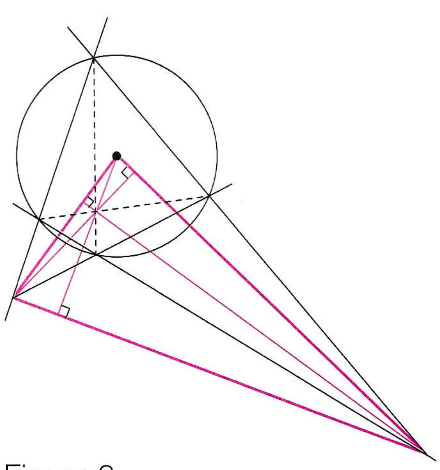


Figure 8

8. Three circles whose diameters are diagonals of a complete quadrilateral meet at two points lying on the line p (see property 3 above) (fig. 9).

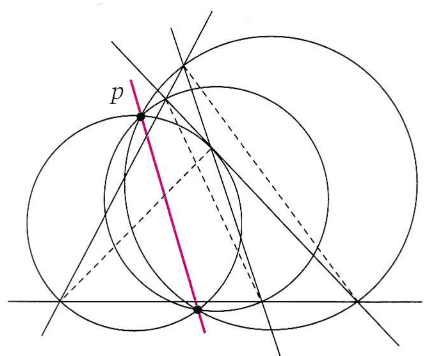


Figure 9

9. Suppose that the straight lines containing the diagonals of a complete quadrilateral form a triangle. The center of the circle circumscribed about this triangle belongs to the line p (fig.10).

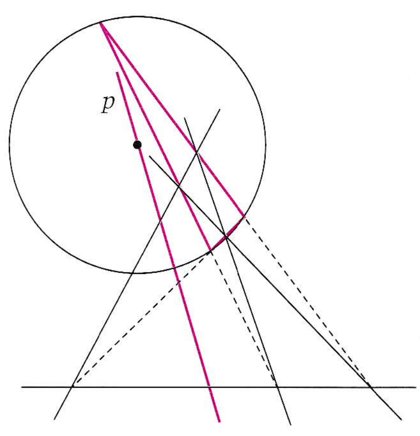


Figure 10

Before we formulate the next property of complete quadrilaterals, let's recall a certain classical statement about triangles. In an arbitrary triangle, three points—the orthocenter, the center of the circumscribed circle, and the point where its medians meet—lie on a straight line. This line is called *Euler's line* (fig. 11). The following property of a complete quadrilateral has to do with this line.

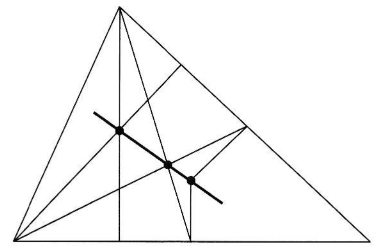


Figure 11

10. If any one of the four lines that form a complete quadrilateral is parallel to Euler's line in the triangle formed by the three other lines, then every other line of the quadrilateral possesses the same property (fig. 12).

11. Let's erect perpendiculars at the midpoints of the segments connecting the orthocenter and the center of the circumscribed circle of each of the four triangles that make up a complete quadrilateral. These perpendiculars meet in one point, called *Harvey's point*.

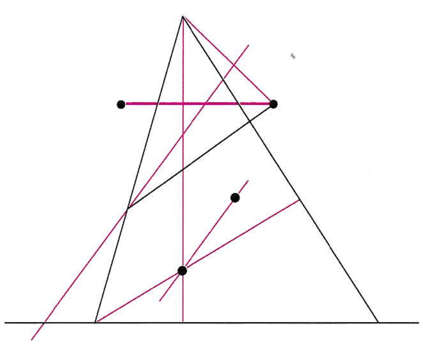


Figure 12

PHYSICS CONTEST

A physics soufflé

"Enough! or Too Much."—William Blake,
The Marriage of Heaven and Hell

by Arthur Eisenkraft and Larry D. Kirkpatrick

WHAT DISTINGUISHES THE world's great chefs from the millions of adequate cooks is an understanding of the concepts of cooking. We strive for a similar appreciation of physics concepts in our students. Most of the time the problems in physics textbooks do not require much understanding to obtain the answer in the back of the book. If the problem gives the mass m and acceleration a of an object and asks for the value of the net force F acting on the object, it is not too difficult to find a formula containing m, a, and F and plug in the numbers. To enhance such a cookbook problem, we may provide superfluous information like the velocity v of the object or its color λ. Students who simply look for an expression containing m, a, v, F, and λ will not succeed with this approach. Students must understand the concepts well enough to understand that the velocity and color are not needed. Only the capable chef can ensure that the soufflé will rise.

As an example of giving extra information, consider the following problem: "How much work does the gravitational force perform on a satellite in a circular orbit around Earth? The mass of the satellite is 4,500 kg, the color of the satellite is blue (λ = 450 nm), the circumference of its orbit is 42,000 km, the Earth's radius is 3,760 km, and the gravitational field at the Earth's surface is 9.8 N/kg." All of these numbers are extra information for students who realize that the gravitational force on the satellite is perpendicular to the satellite's displacement and, therefore, the work is zero!

Is this fair or is it a "trick" question? Recognizing what information is useful and what is extraneous is important in life and in physics problems. We have had students complain about problems in which additional, superfluous information was provided. In an attempt to generate discussion, one of us (LDK) gave the students a problem and deliberately omitted one vital piece of information. He then asked, "What do you need to be able to solve this problem?" The problem was not well received, but asking such problems should be commonplace. In the real world, physicists and engineers often have to figure out what measurements are required to obtain the data needed to solve problems.

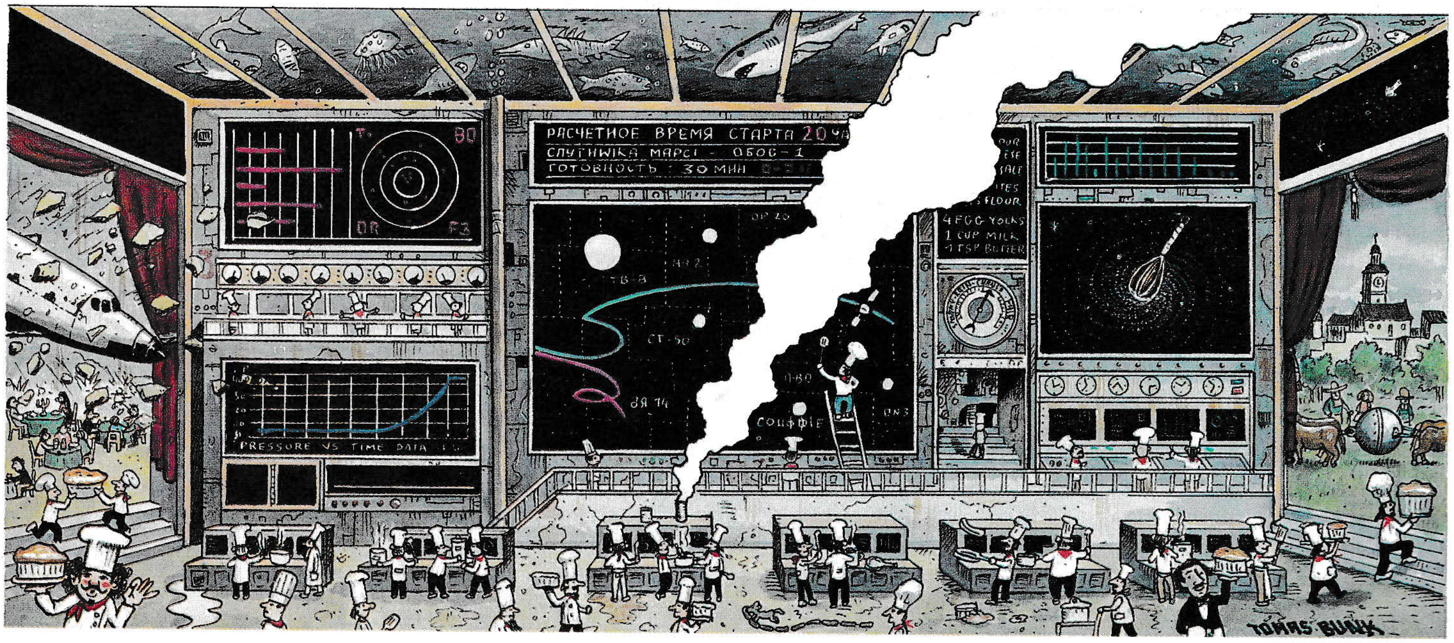
Sometimes, the physicist works through the theoretical aspects of a problem only to discover that a number is not needed, or that a calibration is not needed. This brings us to this month's problem. It is based on one of the problems on the second examination used to select members of the 1997 US Physics Team that will compete in the International Physics Olympiad that is hosted by Canada in Sudbury, Ontario, this July. The problem originally appeared in *Kvant*, our sister publication, many years ago.

Upon entering the atmosphere of a planet, a probe descended straight down to the surface. Along the way it recorded the atmospheric pressure as a function of time as shown in figure 1 (on page 32). Unfortunately the calibration of the pressure gauge has been lost and the units on the pressure axis are not known. Your mission, should you choose to accept it, is to compensate for this lack of calibration.

The atmosphere is mostly carbon dioxide with a molecular mass of 44 g/mol and can be treated locally as an ideal gas. The surface temperature T_s at the surface is 400 K, the gravitational field g at the surface is 9.9 N/kg, and the radius R of the planet is 5,000 km.

A. Apply Newton's second law to a small slab of the atmosphere of vertical thickness Δy to show that the change in pressure ΔP between the top and bottom of the slab is given by

Art by Tomas Bunk



РАСЧЕТНОЕ ВРЕМЯ СТАРТА 20
СЛУТНWIKA МАРС1 - ОБОС-1
ГОТОВНОСТЬ : 30 МИН
PRESSURE VS TIME DATA
4 EGG YOLKS
1 CUP MILK
1 TSP BUTTER
TOMAS BUNK

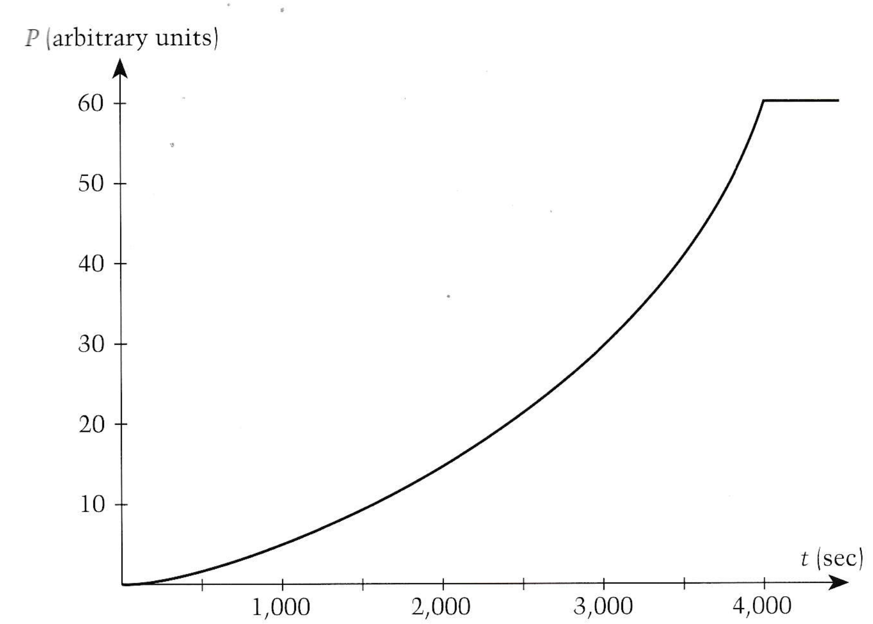


Figure 1

$$\Delta P = \rho g \Delta y,$$

where ρ is the atmosphere's density and g is the local gravitational field.

B. Using this formula for the change in pressure with altitude and the graph, estimate the probe's speed v_0 just before it strikes the surface. Why is the calibration data not needed?

C. Under the simplifying assumption that the probe's speed is constant during its travel through the lower atmosphere, estimate the temperature of the atmosphere at a height of 15 km above the surface.

D. Estimate the uncertainty in your determination of this temperature. How confident are you that your value for the temperature is meaningful?

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.

Do you promise not to tell?

The January/February contest problem asked readers to orient two radio antennas so that one friend living in town A receives a maximum signal and while a second friend living in town B receives no signal. A wonderful solution was submitted by our colleagues André Cury Maiali and Gualter José Biscuola (jointly) and Flavis Pakianathan.

In part A of the problem the two radio sources were in phase. To solve this problem, we must realize that the path difference between the antennas S_1 and S_2 and town A must be equal to an integral number of wavelengths. The geometry, as shown in figure 2, leads to the familiar equation

$$n_1\lambda = d \sin \theta_1.$$

The path difference between the antennas S_1 and S_2 and town B must be equal to an odd integral number of half-wavelengths. This leads to a similar equation

$$\left(n_2 + \frac{1}{2}\right)\lambda = d \sin\theta_2.$$

Subtracting these equations and solving for the distance d between the sources, we get

$$d = \frac{[n_1 - (n_2 + 1/2)]\lambda}{\sin\theta_1 - \sin\theta_2}.$$

We can see that there are an infinite number of selections for the distance between the antennas and their orientation that allows for an antinode at A and a node at B. If we take the simplest situation where town A lies on the perpendicular bisector of the line connecting the antennas, the distance d is defined by the orientation of the angle θ_2:

$$d = \frac{(n_2 + 1/2)\lambda}{\sin\theta_2}.$$

Part B of the problem asked for the parameters of the array (including the phase shift) such that the distance between the antennas is a minimum.

The total path difference when there is a phase delay δ between the sources is

$$d \sin\theta + \frac{\delta\lambda}{2\pi}.$$

The constructive and destructive interference equations thus become

$$n_1\lambda = d\sin\theta_1 + \frac{\delta\lambda}{2\pi},$$

$$\left(n_2 + \frac{1}{2}\right)\lambda = d\sin\theta_2 + \frac{\delta\lambda}{2\pi}.$$

Subtracting these equations and solving for the distance between the sources d, we get

$$d = \frac{[n_1 - (n_2 + 1/2)]\lambda}{\sin\theta_1 - \sin\theta_2}.$$

We can minimize d by minimizing the numerator and maximizing the denominator. The numerator would be a minimum if

$$n_1 = n_2 + \frac{1}{2}.$$

But since n_1 and n_2 are integers, this is impossible. The smallest value of the numerator occurs when $n_1 = n_2$.

We can find the maximum value of the denominator by taking the

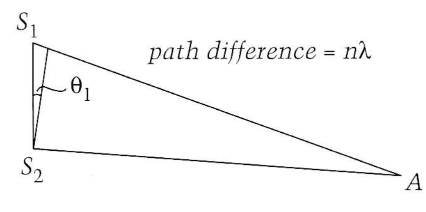


Figure 2

derivative of the denominator and setting it equal to zero:

$$f = \sin\theta_1 - \sin\theta_2.$$

Since θ_1 and θ_2 are both dependent on the orientation of the antennas, let's replace θ_2 with $\theta_1 - \phi$, where ϕ is the angle between the two locations:

$$f' = \cos\theta_1 - \cos(\theta_1 - \phi),$$
$$\cos\theta_1 = \cos(\theta_1 - \phi).$$

For the cosine of the two angles to be equal and $\phi \neq 0, 2\pi, 4\pi, \ldots$, we must have

$$\theta_1 = -(\theta_1 - \phi),$$
$$\theta_1 = \frac{\phi}{2}.$$

This makes sense, because in setting up the array we found from our analysis of the numerator that n_1 and n_2 are equal. Therefore, we expect that the perpendicular bisector of the array should pass between the two towns.

The equation for the minimum distance between the sources then becomes

$$d = \frac{\lambda}{4\sin\frac{\phi}{2}}.$$

The corresponding phase shift between the antennas can now be found:

$$n_1\lambda = d\sin\theta_1 + \frac{\delta\lambda}{2\pi},$$
$$\delta = \frac{-\pi}{2} - 2\pi n_1 = \frac{3\pi}{2} - 2\pi n_1.$$

Part C of the problem asked for a numerical solution for a broadcast frequency of 27 MHz and angles between north and the directions to the towns of 72° and 157°, respectively:

$$\lambda = \frac{c}{\nu} = 11.1\text{ m},$$
$$d = \frac{11.1\text{ m}}{4\sin 42.5°} = 4.1\text{ m}.$$

The orientation of the antenna is such that the perpendicular bisector of the line connecting the antennas makes an angle of 72° + 42.5° with the north.

A brilliant idea

You can talk of brilliant light,
Light that makes like day the night,
Or of adjectives like "bright" and nouns like "glare,"
But if you're wanting shining
That can fry your eyeball's lining,
Then you'll find that lasers are beyond compare.

See, a laser is a light
That's so very, very bright,
You can aim it up and bounce it off the Moon.
For a laser blast's duration,
It is no exaggeration,
It's more dazzling than the Sun at cloudless noon.

Yes, it's bright! bright! bright!
Whether blue or green or red—they're never white.
When a laser is a-lasing,
You had best avert your gazing
If you ever want to see another sight.

If you want to build a laser
For your wife, that will amaze her,
It's quite easy to assemble in a day.
Get a flashtube filled with xenon
(Like the strobe that you're so keen on)
And wrap it 'round a rod like DNA.

Now, that rod is made of stuff
That, when energized enough,
The atoms will emit a photon shower
Which will then leak out the end,
So you need to make it bend
Back into the rod, so it builds up more power.

It will grow! grow! grow!
If the polished ends reverse it to and fro.
Every time, it bounces back,
Hits the wall, reverses track—
All the while, the xenon tube pumps in its glow.

Well, the silver-polished ends,
On which power growth depends,
Are not polished, quite *exactly*, just the same.
So the end whose mirror's weaker
Finally gives, and there's no sleeker
Kind of light than that, which shoots with pinpoint aim.

Now, the wavelengths of this light
Are identical, all right,
And the photons all go marching, locked in sync.
It's "coherent," as they say
When describing such a ray.
So, that's how to build a laser—what'cha think?

It is cool! cool! cool!
To create a laser of one gigajoule.
You can vaporize a tree,
Rid your dog of every flea—
Just don't let your child take it in to school . . .

—David Arns

Counting problems in finite groups

This column proves finite as well . . .

by George Berzsenyi

DURING THE PAST DECADE, one of the best programs supported by the National Science Foundation (NSF) has been Research Experiences for Undergraduates (REU). Within this program, each year around twenty mathematicians receive funding to work with 6–12 students on a variety of topics in mathematics during a period of 6–9 weeks during the summer. The students work in groups and/or with their advisor and his/her associates on a suitable research problem for which they have adequate background, in the hope of finding at least a partial solution. In general, their work is highly focused and very intensive. Ideally, in addition to proving some theorems, conjecturing others, and solving related problems, they become familiar with many aspects of the research environment. They read a number of papers on related results, learn proper techniques for verifying the originality of their findings, familiarize themselves with suitable software packages, and learn how to communicate their results orally and in written form.

Of the many excellent REU programs conducted throughout the United States, in my opinion, Gary Sherman's "Computational Group Theory" was the best. He conducted it single-handedly for eight years (from 1989 to 1996 at Rose-Hulman Institute of Technology), with six students for seven weeks each year. His 48 students (14 women and 34 men) came from many different schools (Binghampton, Brown, Bowling Green, Carleton, Carnegie-Mellon, Chicago, Duke, Harvard, Haverford, Hendrix, Wooster, Illinois, Michigan, Mills, Nebraska, New Mexico State, Pomona, St. Norbert, and Rose-Hulman), and most of them went on to prestigious Ph.D. programs. During their stay at Rose-Hulman, they produced a total of 37 technical reports, which led (thus far) to 16 refereed publications (with 4 more in preparation). Many of them also gave well-received presentations at various national and regional meetings of the American Mathematical Society.

Though I was heavily involved with other activities (first, chairing the Department, and then with an NSF-supported Young Scholars Summer Program), I thoroughly enjoyed my limited interactions with Gary's students, admiring his "coaching style," and watching the incredible effects thereof. The students usually worked in groups of two to four, with each of them being in several of the groups. Hence their interactions were constant and most beneficial. The experimental aspects of their research was done with the help of the powerful computer algebra system Magma (an updated version of Cayley). For the most part, they worked on a variety of counting problems in finite group theory motivated by the question: **What is the probability that two group elements commute?** (This question was originally considered by the legendary Paul Erdős and his associates.) In particular, Gary's students managed to show that **the probability that a pair of elements in a finite group generates a cyclic subgroup is either 1 or at most 5/8.** Moreover, **a finite group has either a unique centralizer or at least four centralizers, and it has four if and only if the group, modulo its center, is the Klein four-group.** One of his students (Jordan Ellenberg, who was a multiple winner of both the USA Mathematical Olympiad and the Putnam Examination) also managed to show that **the probability that a triple product is rewritable (that is, $xyz \in \{xzy, yxz, zxy, yzx, zyx\}$) is either 1 or at most 17/18.**

For a complete listing of the technical reports and the articles based on them, the reader is referred to Rose-Hulman's Web site (http://www.rose-hulman.edu/Class/ma/HTML/REU/NSF-REU.html); to obtain copies thereof, please contact the secretary of our Department of Mathematics. A more complete description of

Dr. Sherman's program can be found in an article published in the December 1992 issue of *PRIMUS* (Volume II, Number 4, pp. 289–308). For a complete listing of the 20 REU sites in the area of mathematics in 1997, please see the NSF's Web page (http://www.nsf.gov/mps/dms/reulist.htm). I strongly recommend to my readers to explore the problems addressed in the various REU programs; they should provide excellent sources for their mathematical investigations.

Another excellent source for mathematical investigations is the regular "Unsolved Problems" column of *The American Mathematical Monthly*. While in general the problems are more demanding, most of them are well within the expertise of the readers of *Quantum*. There is also a column of "Student Research Projects" in *The College Mathematics Journal*, which I also strongly recommend to my readers.

My reason for suggesting alternative sources of problems is that this is my last "Math Investigations" column. Though I thoroughly enjoyed writing these columns and the subsequent interactions with my readers, all good things must come to an end. I greatly appreciate the opportunity provided by NCTM and NSTA to serve in this capacity, and I am most thankful to Tim Weber for his excellent editorial assistance throughout the years.

Feedback

First of all, I wish to thank Brian Hutchings (Arlington, Virginia) and Raul A. Simon (Arica, Chile) for their communications about the challenges posed in my January/February 1997 ("Revisiting the N-cluster Problem) and September/October 1996 ("Embedding Triangles in Lattices") columns. I also appreciate the interest shown by Richard B. Hanlon (Independence, Missouri) in my March/April 1996 ("The Orbit of Triangles") and March/April 1997 ("The Equalizer of a Triangle") columns.

Earlier, I heard from Vladimir Golkhovoy (St. Petersburg, Russia) in response to one of the questions posed in my March/April 1991 article ("Adventures Among P_t-sets"). He showed that $\{13, 24, 45\}$ is a P_t-set for $t \in \{-296, -56, 1624\}$, $\{3, 8, 99\}$ is a P_t-set for $t \in \{-8, 232, 1912\}$, and $\{21, 32, 45\}$ is a P_t-set for $t \in \{-656, -416, 1264\}$.

Belatedly, I also wish to acknowledge the contributions of Phillip Embree (Centertown, Missouri) and Eric Zizermann (Sunbury, Pennsylvania) to my July/August 1994 ("Constructing Triangles from Three Given Parts") and September/October 1994 ("Constructing Triangles from Three Located Points") articles, along with the contributions of James F. Murphy (Bucks Harbor, Maine), Bernardo Recaman (Bogota, Colombia), Tenie Remmel (San Diego, California), Helio Waldman (Campinas, Brazil), Richard Fischer (Princeton, New Jersey), Ron Ruemmler (Edison, New Jersey), and John L. Horst and his students (Brent Siegrist, Marcelo Mast, and Ryan Raines of Eastern Mennonite College, Harrisonburg, Virginia) to my September/October 1993 ("Endless Self-description") article. I was also most thankful to Benjie Chen (Sunnyville, California) for his comments on my November/December 1993 ("Periodic Binary Sequences") article, and to Jason A. Jones (Phoenix, Maryland) and Robert J. Veith (Louisville, Kentucky) for their interest in the problems featured in my May/June 1994 ("Happy Birthday, Uncle Paul") article. Finally, I wish to call my readers' attention to a recent article by Harold Reiter (entitled "The Chinese Restaurant Approach to Integer Representation Problems"), whose preliminary version served as the basis for my July/August 1996 ("Finding the Family Resemblance") column. Harold's article appeared in the March 1997 issue of *Mathematics and Informatics Quarterly*.

George Berzsenyi *is a professor of mathematics department at Rose-Hulman Institute of Technology in Terre Haute, Indiana. His e-mail address is george.berzsenyi@rose-hulman.edu.* Quantum *is most grateful to Dr. Berzsenyi for the time and effort he has devoted to this magazine from its very inception. We hope to see his byline on occasion in future issues.*

AT THE BLACKBOARD I

Whirlwinds over the runway

"The end of the runway rushes toward us,
The call comes: 'Rotate!' and we're off . . ."
—from an old song popular at the
Moscow Institute of Physics and Technology

by Albert Stasenko

IT'S ALWAYS EXCITING TO take off in an airplane, but if you're in a small plane, you would do well to leave plenty of room between you and the jumbo jet taking off in front of you. Have you ever seen a pigeon land on dusty ground? You can see the vortices produced, because the dust particles are stirred up by the flapping wings. Similar vortices are generated by a jumbo jet when it takes off. The plane produces powerful downward streams of air, and woe to the light plane that tries to follow the big plane at this critical moment. The wings of the small plane may enter vertical air flows having opposite velocities, which simply flip the plane on its back. Because the plane is still close to the ground, the pilot doesn't stand a ghost of a chance. Unfortunately in the annals of aviation one can find numerous reports of crashes due to this phenomenon. It stands to reason that vortices on the runaway are a subject of keen interest to pilots, air traffic controllers, and aerodynamic engineers. Although it is a dauntingly complex subject, we can certainly gain something by examining this subject, "vortex kinematics," in simplified form.

What is a vortex? It can be observed, say, when water goes down the drain of a sink or bathtub. If the water contains tea leaves or other small particles, you can readily see that the nearer a particle is to the axis of rotation, the greater its linear (circumferential) velocity. In hydrodynamics there is an important concept, the *potential vortex*, in which the linear velocity is inversely proportional to the distance from the axis: $v \sim 1/r$ (fig. 1). The same notion can also be expressed like this: the linear velocity times the circumference is a constant value, called the *circulation*—that is,

$$v \cdot 2\pi r = \Gamma. \qquad (1)$$

(By the way, the same formula also describes a magnetic field **H** if we replace v with H and Γ with I (current).

It's not hard to imagine that a moving airplane is generally followed by two vortices moving in opposite directions. Indeed, to counterbalance the force of gravity, the plane's wings must send a large amount of air downward. The air particles then move sideways and then upward. As a result of the forward motion of a plane, these particles move along a helical trajectory (fig. 2).

These two vortices can be considered as mirror reflections of each other relative to the vertical plane of symmetry of the aircraft: in figure 3 its component is the OY-axis. The air streams produced by the right

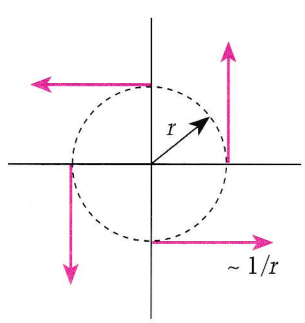


Figure 1

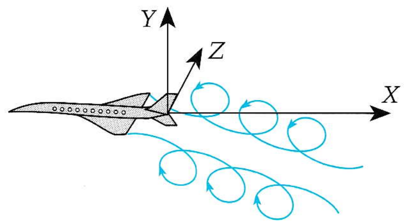


Figure 2

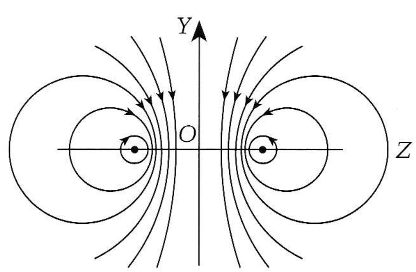
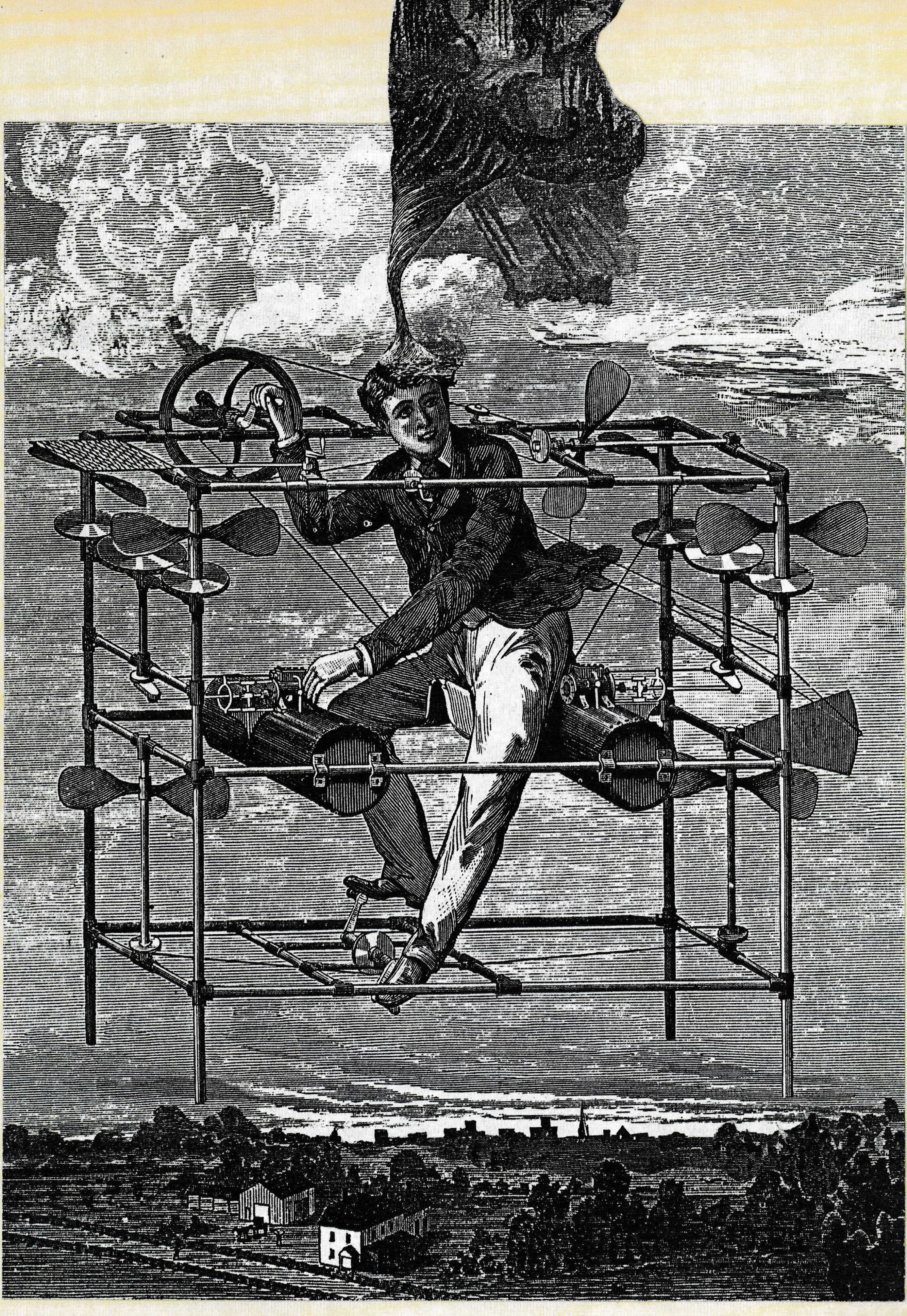


Figure 3

Collage by Vera Khlebnikova



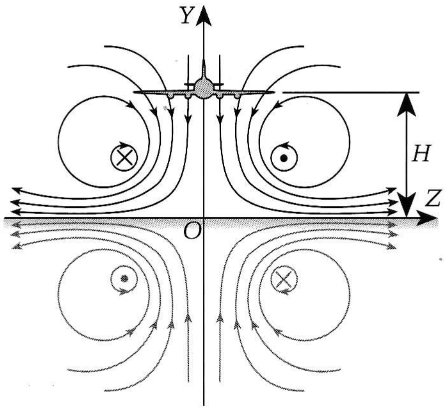


Figure 4

and left vortices move downward along the OY-axis. So this vertical plane of symmetry plays the role of an impermeable (for vortices) partition. Let's imagine a plane flying over an airfield at a low altitude H. The ground is certainly impermeable for air motion, so the air streams generated by the two real vortices will pass parallel to the ground as well (fig. 4). The pattern of the air streams will look as if there were another pair of "underground" vortices that are the mirror reflections of the two real vortices relative to the horizontal plane.

Physicists adore analogies: recall that this pattern is just like that of a magnetic field generated by four parallel wires carrying equal amounts of electrical current. In figure 4 the direction of these imaginary currents is indicated by a dot (if the current flows toward the reader) or a cross (if it flows away from the reader). Note that the magnetic field of a single wire propagates to infinity, but the composite pattern of the magnetic field generated by four wires (fig. 4) makes it look as if impermeable partitions were between the fields of each wire! This is the key point: instead of two real, complicated vortices interacting with the horizontal plane, we can work with four simple potential vortices without any boundary plane. Isn't that beautiful?

In this approach, any of the four vortices is located in the total field generated by the three others. Let's look at the motion of one of the vortices—say, number 1 (fig. 5). The position of its axis can conveniently be described in polar coordinates. A point in this system is described by the length ρ of a vector drawn to the given point from the origin and by the azimuth angle ϕ formed by the radius-vector ρ and some reference line—we'll choose OY for this purpose. In this system the velocity vector $\mathbf{v}$ has a radial component v_ρ and an azimuth component v_ϕ that is perpendicular to the radius-vector. Now we obtain the velocities generated by all three vortices (numbers 2, 3, and 4) at the axis of the first vortex. According to equation (1), the left real vortex (number 2) generates a velocity that is directed vertically downward and has an amplitude

$$v^{(2)} = \frac{\Gamma}{2\pi \cdot 2z} = \frac{\Gamma}{2\pi \cdot 2\rho \sin\phi}.$$

(Here we took into account that $z = \rho \sin\phi$.) The radial and azimuthal components of $v^{(2)}$ (fig. 5a) are

$$v_\rho^{(2)} = -v^{(2)}\cos\phi = -\frac{\Gamma\cos\phi}{2\pi \cdot 2\rho\sin\phi},$$

$$v_\phi^{(2)} = v^{(2)}\sin\phi = \frac{\Gamma}{2\pi \cdot 2\rho}.$$

Here the minus sign means that the radial component of vector $\mathbf{v}^{(2)}$ is directed opposite the radius-vector. The velocity $v^{(3)}$ generated by the left imaginary vortex (number 3) has only an azimuthal component:

$$v_\phi^{(3)} = \frac{-\Gamma}{2\pi \cdot 2\rho}.$$

Finally, the velocity $v^{(4)}$ produced by the right imaginary vortex (number 4) has the following components:

$$v_\rho^{(4)} = \frac{\Gamma\sin\phi}{2\pi \cdot 2\rho\cos\phi},$$

$$v_\phi^{(4)} = \frac{\Gamma}{2\pi \cdot 2\rho}.$$

Both the radial and the azimuthal velocity components can be expressed in terms of the corresponding time derivatives of the polar coordinates:

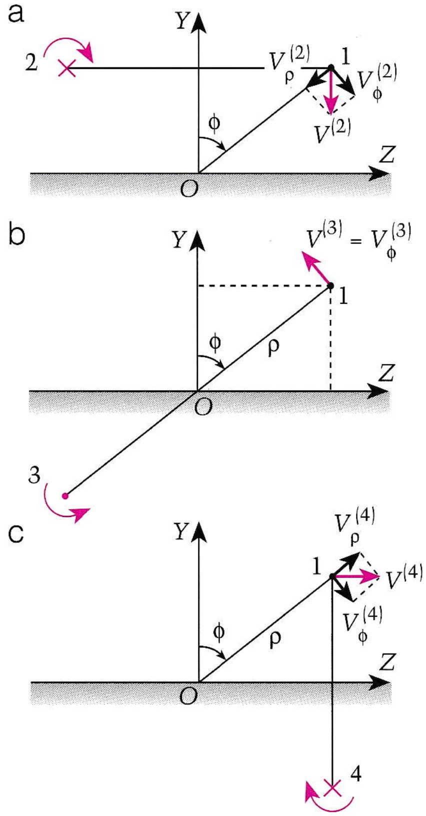


Figure 5

$$v_\rho = \frac{\Delta\rho}{\Delta t},$$

$$v_\phi = \rho\frac{\Delta\phi}{\Delta t}.$$

These are the components of the sum of the velocities formed by the three vortices. (Note: in contrast to the great Baron Munchausen, the vortex 1 can not act upon itself.) So

$$\frac{\Delta\rho}{\Delta t} = \frac{\Gamma}{2\pi \cdot 2\rho}\left(-\frac{\cos\phi}{\sin\phi} + \frac{\sin\phi}{\cos\phi}\right),$$

$$\rho\frac{\Delta\phi}{\Delta t} = \frac{\Gamma}{2\pi \cdot 2\rho}. \qquad (2)$$

Thus we have obtained a set of kinematic equations—that is, the mathematical relationships between the space and time variables. Now we'll try to solve this system—nothing is impossible for intrepid *Quantum* readers! First, let's divide one equation by the other to eliminate time. In this way we obtain a relationship between the radial and azimuth coordinates of the vortex axis:

$$\frac{\Delta\rho}{\rho\Delta\phi} = \frac{\sin^2\phi - \cos^2\phi}{\sin\phi\cos\phi} = -\frac{\cos 2\phi}{\frac{1}{2}\sin 2\phi}.$$

Here we reduced the right-hand term to a common denominator and used the trigonometric formulas for the sine and cosine of a double angle. Multiplying both sides of this equation by $\Delta\phi$, we separate the variables according to mathematical terminology: the left side depends only on $\Delta\rho$, while the right side depends only on ϕ. Thus

$$\frac{\Delta\rho}{\rho} = -\frac{\cos 2\phi \cdot \Delta(2\phi)}{\sin 2\phi} = -\frac{\Delta(\sin 2\phi)}{\sin 2\phi}.$$

(Remember, the derivative of a sine is a cosine.) Finally, simple integration (one can look it up in a table of integrals) yields

$$\ln\frac{\rho}{\rho_*} = \ln\frac{\sin 2\phi_*}{\sin 2\phi},$$

or

$$\frac{\rho}{\rho_*} = \frac{\sin 2\phi_*}{\sin 2\phi},$$

where ρ_* is the value of the radius-vector corresponding to some angle ϕ_*. Clearly the radius is minimal ρ_m when $\sin 2\phi = 1$—that is, when $\phi = \pi/4$. Thus we have

$$\frac{\rho}{\rho_m} = \frac{1}{\sin 2\phi}. \qquad (3)$$

What a nice relationship!

Now we know what the component of the vortex axis on the vertical plane looks like. First, it's symmetric relative to the bisector of the right angle YOZ. Second, when $\phi \to 0$ or $\phi \to \pi/4$, the value of the radius-vector tends to infinity. Consider, for example, the case when $\phi \to 0$, $\rho \to \infty$ (that is, when the airplane is high in the sky). The second equation of system (2) shows that the azimuth velocity tends to zero (so the vortex velocity is vertical), while the first equation of this system yields a value for the vertical velocity with which both vortices descend to the ground:

$$\frac{\Delta\rho}{\Delta t} = -\frac{\Gamma\cos 2\phi}{2\pi\cdot 2\rho\sin\phi\cos\phi},$$

or

$$-\frac{\Gamma}{2\pi l} = v_{y\infty},$$

because $\rho \sin \phi = l$, where l is the distance between the vortices.

Plugging equation (3) into the second equation of system (2), we get

$$\frac{\Delta(2\phi)}{\sin^2 2\phi} = \frac{\Gamma}{2\pi}\frac{\Delta t}{\rho_m^2}.$$

Again we have obtained an equation with separated variables: the azimuth of the vortex axis is on the left, while the time variable is on the right. Integrating this equation (again we can use a table of integrals) we have

$$\frac{\Gamma}{2\pi\rho_m^2}t = \frac{1}{\tan 2\phi_0} - \frac{1}{\tan 2\phi}, \qquad (4)$$

where the angle ϕ_0 corresponds to the initial moment $t = 0$ (when the vortices were formed behind the airplane). Figure 5a shows that $\tan\phi_0 = l/2H$, where H is the altitude of the plane.

So we see that the equations (3) and (4) completely define the positions of the vortex axes at any moment in time. For example, for $t \to \infty$ we have $\tan 2\phi \to 0$, so $\phi \to \pi/2$, and $\rho \to \infty$. Therefore, the vortex axes diverge and spread parallel to the ground. In doing so, the radial velocity $\Delta\rho/\Delta t$ becomes the horizontal velocity of the vortex axis:

$$\frac{\Delta\rho}{\Delta t} \to \frac{\Gamma}{2\pi_m} = v_{z\infty}$$

for $\phi \to \pi/2$.

Let's estimate this velocity. According to the theorem formulated by Nikolay Yegorovich Zhukovsky, the lift (which in horizontal flight is equal to the aircraft's weight W) is determined by a simple formula

$$W = \Gamma\rho_a u l,$$

where ρ_a is the density of the atmosphere and u is the velocity of the plane. For example, let's consider a flight at an altitude equal to half the distance between the vortices: $H = l/2$. In this case, $\phi_0 = \pi/4$ and $\rho_0 = \rho_m = \sqrt{H^2 + H^2} = H/\sqrt{2} = l/\sqrt{2}$, which gives us

$$v_{z\infty} = \frac{W\sqrt{2}}{2\pi\rho_a u l^2}.$$

To estimate this value, let's assume the following parameters for a jumbo jet taking off: $m = 300$ t, $W = mg = 3 \cdot 10^6$ N, $l = 50$ m, $u = 100$ m/s, $\rho_a = 1$ kg/m^3. The vortices will move to the right and to the left with a velocity

$$v_{z\infty} \cong 2.6 \text{ m/s}.$$

All our reasoning is valid when the air is calm. Now imagine that a cross wind blows from the right with the same velocity, 2.6 m/s. In this case the right vortex of a heavy plane will remain over the runaway, and it will hinder the next plane taking off until it dissipates.

And now, bon voyage—may all your flights be vortex-free!

AT THE BLACKBOARD II

Remarkable limits

(Generated by classical means)

by M. Crane and A. Nudelman

FOR ANY TWO POSITIVE NUMBERS a AND b, the number $(a+b)/2$ is called their *arithmetic mean*, and the number $\sqrt{ab}$ is called their *geometric mean*. We encounter these two mean values more often than the *harmonic mean* $2ab/(a+b)$ of a and b (for example, the average speed of a car that travels the first half of a trip at a speed a and the second half at a speed b is equal to the harmonic mean of a and b). All these mean values lie between the numbers a and b (readers are invited to check this themselves).

It's easy to verify[1] that the following inequalities hold for all positive $a \neq b$:

$$\frac{2ab}{a+b} < \sqrt{ab} < \frac{a+b}{2} \tag{1}$$

In this article we'll be using limits for a very specific purpose. Say we're given two positive numbers a and b, $a < b$. If we calculate a pair of their mean values, we obtain the numbers a_1 and b_1. Further, we can calculate the same mean values for a_1 and b_1, and thus obtain the new numbers a_2 and b_2. Then we repeat this procedure with these two numbers, and so on. As a result of this activity, we get two number sequences: (a_n) and (b_n).

For example, if we take the geometric mean and the arithmetic mean of the numbers 1 and 3, we get

$$\begin{aligned} a_1 &= \sqrt{3} \cong 1.732050808, & b_1 &\cong 2; \\ a_2 &\cong 1.861209718, & b_2 &\cong 1.86025404; \\ a_3 &\cong 1.863616006, & b_3 &\cong 1.863617561; \\ a_4 &\cong 1.863616784, & b_4 &\cong 1.863616784; \end{aligned}$$

etc.

We see that the sequences (a_n) and (b_n) in this example converge very quickly to each other. Will this always be so? It turns out that sequences of this sort always have a common limit. It's not difficult to prove this statement. But how can we find the limiting value?

[1]See, for example, *An Introduction to Inequalities* by Edwin Beckenback and Richard Bellman (Washington: Mathematical Association of America, 1961).—*Ed.*

Arithmetic-harmonic mean

We'll start with the case when the chosen pair of mean values consists of the *arithmetic* and *harmonic* means. Thus the terms of the sequences (a_n) and (b_n) are defined by the formulas

$$a_{n+1} = \frac{2a_n b_n}{a_n + b_n}, \quad b_{n+1} = \frac{a_n + b_n}{2} \tag{2}$$

$(n = 1, 2, \ldots;\ a_0 = a;\ b_0 = b)$.

It follows from inequality (1) that

$$a < a_n < a_{n+1} < b_{n+1} < b_n < b$$

—that is, the sequence (a_n) increases to "meet" the decreasing sequence (b_n).

Thus both sequences are monotonic and bounded; therefore, in accordance with Weierstrass's theorem, they tend to the limits $\alpha = \lim\limits_{n\to\infty} a_n$ and $\beta = \lim\limits_{n\to\infty} b_n$. Passing to the limit in one of the equalities (2)—for example, in the second one—we find

$$\beta = \lim_{n\to\infty} b_{n+1} = \lim_{n\to\infty} \frac{a_n + b_n}{2} = \frac{1}{2}\left(\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n\right) = \frac{1}{2}(\alpha + \beta),$$

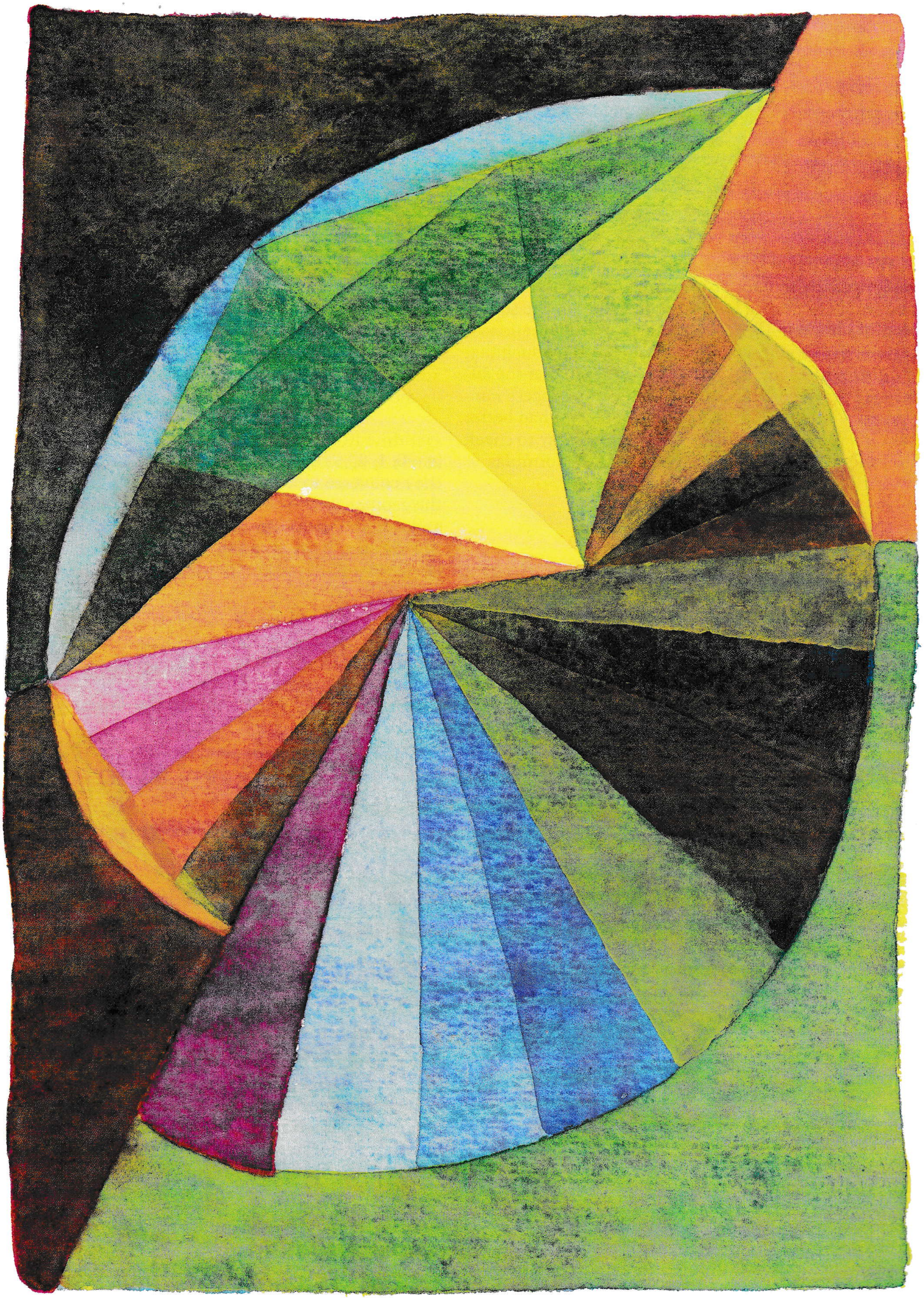
and thus $\alpha = \beta$—that is, sequences (a_n) and (b_n) have a common limit. This limit is called the *arithmetic-harmonic mean* of the numbers a and b. Let's find it algebraically. It follows from equalities (2) that

$$a_{n+1} \cdot b_{n+1} = a_n \cdot b_n = \ldots = a_1 b_1 = ab,$$

and so

$$\alpha^2 = \left(\lim_{n\to\infty} a_n\right)^2 = \lim_{n\to\infty} a_n \times \lim_{n\to\infty} a_n = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n = \lim_{n\to\infty}(a_n \times b_n) = ab.$$

Art by Vera Khlebnikova



Therefore,

$$\boxed{\alpha = \sqrt{ab} = \beta}$$

(the arithmetic-harmonic mean coincides with the geometric mean).

Exercise 1. Prove that

$$b_n - a_n < \frac{b-a}{2^n}.$$

We see that the sequences (a_n) and (b_n) converge rather quickly to $\sqrt{ab}$. So they might prove useful when one wants to find an approximate value for the square root of some number. To calculate $\sqrt{c}$, the sequences (a_n) and (b_n) should start with numbers a and b such that $c = ab$ (for instance, $a = 1$, $b = c$), and the smaller the difference between a and b, the faster this process converges. So, if we want to calculate $\sqrt{56}$, we had better take $a = 7$, $b = 8$, and not $a = 1$, $b = 56$. It's easy to check that the sequences (a_n) and (b_n) satisfy the formulas

$$b_{n+1} = \frac{1}{2}\left(b_n + \frac{c}{b_n}\right), \quad a_{n+1} = \frac{c}{b_n}.$$

In order to illustrate this reasoning, let's calculate $\sqrt{12}$ and take $b = 4$. We have

$$\begin{aligned} b_1 &= \frac{1}{2}\left(4 + \frac{12}{4}\right) = 3.5, \\ b_2 &\cong 3.464285715, \\ b_3 &\cong 3.464101620, \\ b_4 &\cong 3.464101615, \end{aligned}$$

and from this point on, all the decimal digits up to the ninth remain the same: $\sqrt{12} \cong 3.464101615$.

Arithmetic-geometric mean

When Carl Friedrich Gauss was fourteen years old, he discovered, on the basis of numerical examples, that when the sequences (a_n) and (b_n) are calculated by means of *arithmetic* and *geometric* means

$$a_{n+1} = \sqrt{a_n b_n}, \quad b_{n+1} = \frac{a_n + b_n}{2} \qquad (3)$$

$(n = 0, 1, 2, \dots;\ a_0 = a,\ b_0 = b)$, they converge very quickly.

Exercise 2. Prove that there exists a common limit of the sequences (3).

This common limit is called *arithmetic-geometric mean* of the numbers a and b and is denoted by $\mu(a, b)$. It's not at all easy to find an explicit formula that would express $\mu(a, b)$ as a function of a and b. The first person to find such a formula was Gauss himself. He obtained it by means of extremely sophisticated and ingenious reasoning that used the properties of *elliptic integrals*. We'll offer this expression without a proof:

$$\boxed{\mu(a, b) = \frac{\pi}{2\displaystyle\int_0^{\pi/2} \frac{dx}{\sqrt{a^2 \sin^2 x + b^2 \cos^2 x}}}.}$$

By the way, if we "invert" this expression, we obtain a fast way to compute the integral in the denominator: it's equal to $\pi/\mu(a, b)$, and an approximate value of $\mu(a, b)$ can be found rather quickly by means of the sequences (a_n) and (b_n).

Geometric-harmonic mean

If we construct the sequences (a_n) and (b_n) by means of the *harmonic* and *geometric* means

$$a_{n+1} = \frac{2a_n b_n}{a_n + b_n}, \quad b_{n+1} = \sqrt{a_n b_n} \qquad (4)$$

$(n = 0, 1, 2, \dots;\ a_0 = a,\ b_0 = b)$, it's not hard to prove that they converge to a common limit. Let's call this limit the *geometric-harmonic mean* of the numbers a and b and denote it by $\nu(a, b)$. However, there's nothing new in this case, if we compare it with the sequences (3), since we find immediately from formula (4) that

$$\frac{1}{a_{n+1}} = \frac{\frac{1}{a_n} + \frac{1}{b_n}}{2}, \quad \frac{1}{b_{n+1}} = \sqrt{\frac{1}{a_n} \cdot \frac{1}{b_n}}.$$

Therefore,

$$\frac{1}{\nu(a, b)} = \mu\left(\frac{1}{b}, \frac{1}{a}\right),$$

or, by Gauss's formula,

$$\boxed{\nu(a, b) = \frac{2ab}{\pi} \int_0^{\pi/2} \frac{dx}{\sqrt{a^2 \cos^2 x + b^2 \sin^2 x}}.}$$

Schwab–Schenberg mean

So only the first of three mean values—arithmetic-harmonic, arithmetic-geometric, and geometric-harmonic—can be expressed in an elementary way as a function of the original numbers a and b. Even more surprising, a slight change in the sequences (3) produces sequences whose common limit can be written as an elementary function of a and b (although the word "elementary" doesn't mean that it is easy to find this limit!). Put

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_{n+1} b_n} \qquad (5)$$

$(n = 0, 1, 2, \dots;\ a_0 = a,\ b_0 = b)$.

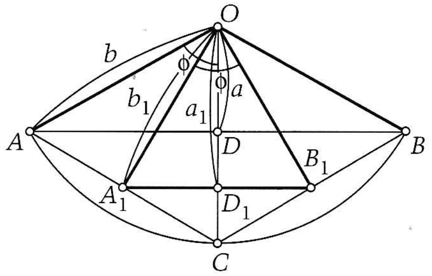


Figure 1

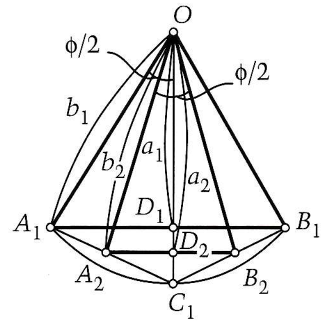


Figure 2

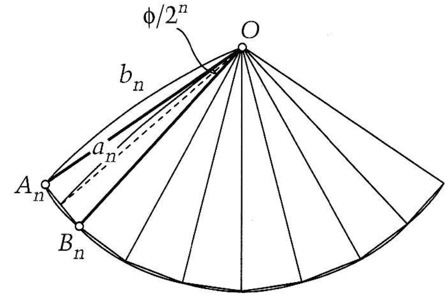


Figure 3

Exercise 2. Prove that there exists a common limit of the sequences (5).

This limit is called the *Schwab–Schenberg mean.* Let's find it. It's curious that it can be derived from elementary geometrical considerations.

Figure 1 shows an isosceles triangle AOB with legs $|OA| = |OB| = b$ and altitude $|OD| = a$ (the angle at vertex O is denoted by 2ϕ), and also shows the arc ACB of the circle with center at vertex O and radius b ($|OC| \perp |AB|$). Let $|A_1B_1|$ be the line joining the midpoints of sides $|AC|$ and $|BC|$ of triangle ABC. Then

$$|OD_1| = |OD| + |DD_1| = |OD| + \frac{|DC|}{2}$$
$$= a + \frac{b-a}{2} = \frac{a+b}{2} = a_1.$$

Since triangle OA_1C is a right triangle,

$$|OA_1|^2 = |OD_1||OC| = a_1 b,$$
$$|OA_1| = \sqrt{a_1 b} = b_1.$$

Thus we see that the numbers a_1 and b_1 are obtained from the numbers a and b after a simple geometric construction, and that the segment of length a_1 is again the altitude drawn from the vertex of an isosceles triangle with lateral side b_1. We note further that

$$\angle A_1OB_1 = \phi, \quad |A_1B_1| = \frac{1}{2}|AB|.$$

If we make similar constructions with the triangle A_1OB_1, we obtain the isosceles triangle A_2OB_2 (fig. 2), in which

$$|OD_2| = a_2,$$
$$|OA_2| = |OB_2| = b_2,$$
$$A_2OB_2 = \frac{\phi}{2},$$
$$|A_2B_2| = \frac{1}{2}|A_1B_1| = \frac{1}{2^2}|AB|.$$

Repeating this construction n times, we obtain the triangle A_nOB_n with altitude $|OD_n| = a_n$ and lateral sides

$$|OA_n| = |OB_n| = b_n,$$
$$\angle A_nOB_n = \frac{\phi}{2^{n-1}},$$
$$|A_nB_n| = \frac{1}{2^n}|AB|.$$

Now let's draw an arc of the circle with radius b_n corresponding to the central angle 2ϕ and then divide it into 2^n congruent arcs. If we sequentially connect the points into which the arc is divided by chords, we obtain a regular 2^n-link broken line inscribed in this arc. Its length is equal to $2^n|A_nB_n| = |AB|$. This broken line is circumscribed about the arc of the circle with the same center and central angle and with radius a_n (fig. 3). Since the perimeter of this broken line is confined between the lengths of the arcs, we have

$$2\phi a_n < |AB| < 2\phi b_n,$$

and thus

$$2\phi \lim_{n\to\infty} a_n \le |AB| \le 2\phi \lim_{n\to\infty} b_n.$$

Since $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$, we see that the common limit α is given by the formula

$$\alpha = \frac{\sqrt{b^2 - a^2}}{\arccos \dfrac{a}{b}}.$$

Finally, we note that if $a = 1/2$ and $b = 1/\sqrt{2}$, we have $\alpha = 2/\pi$; thus sequences (5) allow us to calculate the number π with an arbitrary degree of precision.

Internal energy and heat

Why does "Q" appear in the reference tables and not "ΔU"?

by Alexey Chernoutsan

THE GENERALIZATION OF the law of conservation of energy for thermal processes produced two new notions. First, the very concept of energy was broaden in such a way that, in addition to the widely known mechanical energy E_{mech}, it now included the internal energy U. Second, scientists realized that one need not perform mechanical work to change a system's energy—during heat exchange, energy is transferred from a warm body to a cold body at the molecular level without any macroscopic mechanical motion. The amount of energy transferred via this mechanism is called heat Q. Both of these notions are present in the new form of the law of conservation of energy:

$$Q = \Delta U + W,$$

where W is the work performed by a system against the external forces. We can write this law in a more general form that takes into account a possible change in the mechanical energy:

$$Q = \Delta(U + E_{mech}) + W.$$

This "full strength" version of the law is used in textbooks only when processes involving ideal gases are being discussed, because these gases can change their volume significantly, and this is accompanied by mechanical work. Clearly in this case all three terms of the law of conservation of energy play equally important roles. For example, during an isobaric process both the temperature and the volume of a gas vary, so we need to take into account both the change in the internal energy of the gas and the work it performs.

However, in problems where the "participants" are liquid and solid bodies, we usually apply formulas for heat transmission that describe the amount of heat transferred during the process of warming and cooling:

$$Q = cm(T_2 - T_1);$$

melting and freezing:

$$Q = \pm mL_f;$$

or evaporating and condensing:

$$Q = \pm mL_v.$$

Here c is the specific heat, m is the body's mass, T is its temperature, L_f is the latent heat of fusion, and L_v is the latent heat of vaporization. These formulas are used in the most natural and logical way to construct the thermal balance equation describing the heat exchange between bodies in a closed (that is, thermally isolated) system. In the process of attaining thermal equilibrium, these bodies exchange heat, so the formulas for heat transfer are quite natural here (although, as we'll see later, some doubts exist even in these cases).

However, in problems involving conversion of mechanical energy into thermal energy, the situation isn't so clear. Consider a simple example: the inelastic collision of two identical balls moving toward each other with equal velocities. In determining the temperature to which the balls will be heated, we usually say that in this case all the mechanical energy is converted to heat, and then we apply the formula that describes the amount of heat necessary to warm the bodies. The resulting equation

$$2\frac{mv^2}{2} = 2cm\Delta T$$

settles the issue. Still, a natural question arises: why is the formula for heat transmission applied in this case, where no heat exchange occurs at all? The answer is that the phrase "energy is converted to heat" refers not to heat transfer but to the change in the internal (thermal) energy, described by the law of conservation of energy: $\Delta U + \Delta E_{mech} = 0$. So we should apply formulas describing the changes in the internal energy rather than those describing heat exchange relationships. What are these formulas?

Let's leave melting and evaporation for a while to investigate how internal energy depends on temperature. But why only on temperature? The question should be put like this: how does internal energy depend on pressure and temperature? Indeed, the state of a system is determined by two parameters, so internal energy must depend on both of them.

Art by Pavel Chernusky

Only the energy of an ideal gas depends on a single parameter (temperature), but this isn't the case for liquid and solid bodies. However, in most problems the pressure can be considered constant (equal to, say, the atmospheric pressure). So in these problems it's sufficient to establish the dependence of internal energy on temperature at constant pressure. Note that, strictly speaking, the aforementioned formulas for heat transfer have to do with isobaric processes, so reference books give not an arbitrary value for the specific heat c, but the specific heat c_p corresponding to constant (atmospheric) pressure.

If the pressure as well as the temperature varies in a particular case, it's useful to know that changes in external pressure of a few atmospheres produce rather small variations in internal energy. For example, a 1-atm increase in the pressure applied to water at a temperature of 300 K results in a decrease in internal energy of about 10 J/kg. On the other hand, heating a body by 1 K increases the energy by 4,200 J/kg.

Let's return to the topic at hand. Say we warm a body by ΔT at constant pressure and write the law of conservation of energy for this process. The amount of heat necessary for warming is $Q = cm\Delta T$. The work performed against external forces $W = P\Delta V$, where ΔV is the increase in volume due to the thermal expansion: $\Delta V = V\beta\Delta T = (m/\rho)\beta\Delta T$ (where ρ is the body's density and β is the coefficient of thermal expansion). The law of conservation of energy $Q = \Delta U + W$ results in a formula describing the change in internal energy:

$$\Delta U = \left(c - \frac{\beta}{\rho}\right)m\Delta T,$$

which differs from the corresponding heat transfer formula by a negligible value (the correction for the heat capacity manifests itself only in the ninth decimal place). Therefore, to evaluate the change in internal energy, we can confidently apply a formula like the one for heat transmission—that is, $\Delta U = cm\Delta T$. We should keep in mind, however, that there is an important difference between these formulas: the formula describing the change in internal energy is valid not only for the process of heat exchange, but also for any other mode of changing the internal energy (for example, by collision).

Let's go further. Changes in volume are often much more pronounced in the processes of melting and crystallization than during the heating of a body. For instance, when water freezes, its volume increases by approximately 10%, which corresponds to 10 J of mechanical work performed by each kilogram of water at atmospheric pressure. This is a negligibly small value compared to the latent heat of fusion $L_f = 3.34 \cdot 10^5$ J/kg, so the correction to this large value due to the mechanical work shows up only in the fifth decimal place. Again we see that to calculate a change in internal energy we can apply formulas for heat transmission. And again, the formulas for internal energy can be used regardless of how this energy is changed.

The last process we'll examine is the evaporation of a liquid. We'll assume that evaporation takes place in a cylinder beneath a piston that maintains a constant (atmospheric) pressure in this closed system equal to the saturated vapor pressure. At a given pressure, evaporation proceeds at a particular temperature (for water at atmospheric pressure, it's 373 K). Let's estimate the mechanical work performed by the vapor, taking into account that its volume is much larger than that of the evaporated liquid:

$$W = P\left(V_{\text{vap}} - V_{\text{liq}}\right) \cong PV_{\text{vap}} = \frac{m}{M}RT,$$

where M is the molar mass of the substance. The corresponding change in the internal energy is

$$\Delta U = \left(L_{\text{v}} - \frac{RT}{M}\right)m,$$

and so the relative correction to the latent heat of vaporization is $RT/ML_{\text{v}} \cong 0.076$, or almost 8%. Clearly in this case the change in internal energy differs significantly from the latent heat of evaporation.

Well, after going through so many arguments "in favor of" internal energy, we might still wonder why the formulas for Q are taken to be fundamental, and not those for ΔU. Why do the reference books all give values for the latent heat and not for changes in internal energy? To answer this question, let's look at the heat balance equation, one of the most important practical applications of the thermodynamic formulas.

What is the correct form of the law of conservation of energy to describe heat exchange in a closed system? At first glance we might think it should look like $\Delta U = 0$, where U is the total internal energy, which can also be written as $\Delta U_1 + \Delta U_2 + \ldots = 0$. However, this is not correct. If the system includes bodies whose volumes change significantly (gases, vapors), the work performed by the system against external forces will not be zero, so the total internal energy will not be constant, even though $\Delta U + W = 0$. Let's write down the law of conservation of energy for every body in the system—$Q_1 = \Delta U_1 + W_1$, $Q_2 = \Delta U_2 + W_2$, ...—and add them all up. Since the total work performed by the bodies as they interact is zero (which follows from Newton's third law), the sum $W_1 + W_2 + \ldots$ is equal to the work W performed against the external forces only. Since $\Delta U_1 + \Delta U_2 + \ldots = \Delta U$, and $\Delta U + W = 0$, we can write the law of conservation of energy in this form:

$$Q_1 + Q_2 + \ldots = 0.$$

This means that the exact equation for thermal balance in fact deals with the amount of heat transferred between the bodies in the system, and not with the changes in their internal energies. So the reference books are doing the right thing after all. But it was fun to raise doubts about them, and even more fun to dispel these doubts.

FEEDBACK

Caution: no air brakes on this planet!

But is there another braking mechanism?

by David P. Stern

THE ARGUMENTS OF AGRAwal and Menon ("A Planetary Air Brake," March/April 1997) are flawed by the assumption that some outside force prevents the high atmosphere from rotating with the Earth. They suggest this occurs near a height of $h = 10^5$ m = 100 km, assume a viscosity η and a laminar velocity gradient, and proceed to calculate the resulting resistance to the Earth's rotation.

But there seems to exist no such force, and the atmosphere at 100 km appears to rotate with the Earth. High-altitude rocket payloads have released clouds of sodium and barium vapor at these altitudes, and those clouds are not held motionless while the Earth rotates beneath them.

In a collision-dominated gas, viscosity is practically independent of density. However, as one rises above 100 km, one soon enters the region where most atoms and molecules move on ballistic paths, with hardly any collisions, and the effective viscosity of the atmosphere becomes negligibly small, because in a perfect vacuum there is no viscosity. One does not expect any "air braking," the more so since there exists no object outside the atmosphere to which angular momentum can be transferred.

There could, however, exist an electromagnetic brake, involving the ions of the ionosphere, extending upward from around 100 km. These ions collide with atoms and molecules of the neutral atmosphere, which shares the Earth's rotation, and if that is the only factor, they will end up co-rotating as well.

The difference here is that because the ionospheric plasma has a high electrical conductivity, the motion of ions near Earth is communicated electrically along the entire length of the magnetic field line (or "line of force"). The result is that ions strung out along a field line tend to move in such a way that they continue sharing a field line at later times as well. If ions at the low end of the field line rotate with the Earth, those further away will also tend to co-rotate, even in the absence of collisions.

Does this happen? Observations in space suggest that the ionospheric plasma on field lines that close within about 5 Earth radii (32,000 km) share the Earth's rotation, forming a "plasmasphere" that rotates with the Earth below.

On field lines that close at greater distances—those of the auroral zone and the polar regions—the process is more complicated. The ionospheric ends of those field lines will still try to co-rotate, though one must realize that the rotation speed so close to the pole is rather slow. But the distant ends of the lines may be anchored in a distant medium that does not allow them to co-rotate.

In a rarefied plasma, field lines tend to channel electric currents, as if those lines were copper wires. If the Earth end of each field line rotates with Earth while the distant end is not allowed to do so, a "fluid dynamo" is created between the distant region and the ionosphere. That "dynamo" (see http://www-spof.gsfc.nasa.gov/Education/wcurrent.html on the World Wide Web) drives an electric current i around a circuit that includes parts of both regions, as well as connecting field lines, and the magnetic force $\mathbf{j} \times \mathbf{B}$ exerted on the current through the ionosphere by the Earth's magnetic field acts as a brake (the current also distorts the field lines). Such a dynamo in fact exists between Jupiter's ionosphere and the planet's moon Io, whose orbital motion lags behind the ionospheric plasma's rotation (see http://www-spof.gsfc.nasa.gov/Education/wio.html).

Starting with the "Triad" spacecraft in 1973 (http://www-spof.gsfc.nasa.gov/Education/wtriad.html), satellite observations have mapped currents between the ionosphere and distant space generated by the relative motion of the two media (ionospheric and distant). The currents turn out to be quite large—of the order of a million amperes.

The pattern of these currents, however, suggests that the relative motion to which they respond is not the one due to the Earth's rotation. Rather, it represents the motion of distant space plasmas to which those lines may be linked, especially the fast flow ($\cong$ 400 km/s) of the "solar wind" spreading out radially from the Sun.

The effect of the ionosphere's rotation seems to be negligible. Thus, not only does the Earth lack an "air brake," it does not seem to have much of a magnetic brake, either.

David P. Stern *is a physicist at the Goddard Space Flight Center in Greenbelt, Maryland. His e-mail address is u5dps@lepvax.gsfc.nasa.gov.*

Volume 7 (1996–97)

Bulletin Board

Earth & Sky radio seeks student writers

Students in grades K–12 are invited to write and produce their own science radio programs for the internationally syndicated series Earth & Sky. The Earth & Sky radio series will broadcast five winning entries in its fourth annual Young Producers Contest during the first week of May 1998. The winning teams will also win US Savings Bonds. The winning student programs will be chosen by a panel of judges from the broadcasting and scientific community. The deadline for the contest is December 15, 1997.

Earth & Sky began broadcasting on September 30, 1991, on 30 radio stations. Today it is heard by millions on over 650 public and commercial radio stations in all 50 states. Earth & Sky is heard internationally in 134 countries through the Armed Forces Radio Network, Voice of America, Canadian Broadcast News Satellite, and Radio for Peace International (short-wave), as well as independent stations in the People's Republic of China, Bulgaria, Australia, Panama, Taiwan, and the Netherlands Antilles. The program is funded by the National Science Foundation.

Educational materials and other information about Earth & Sky are available on the World Wide Web (www.earthsky.com). You can also contact Earth & Sky by e-mail (contest@earthsky.com) or fax (512 477-4441).

A CyberTeaser you can bank on

The July/August CyberTeaser (brainteaser B208 in this issue) allowed our Russian colleagues a chance to poke some fun at their new breed of financial entrepreneurs. The head of the "Pyramid Bank" is named "R. A. Scall"—or "rascal," if you ignore the periods (and drop an "l"). Luckily our contestants were not to be distracted by such silliness. And most of them got the answer right.

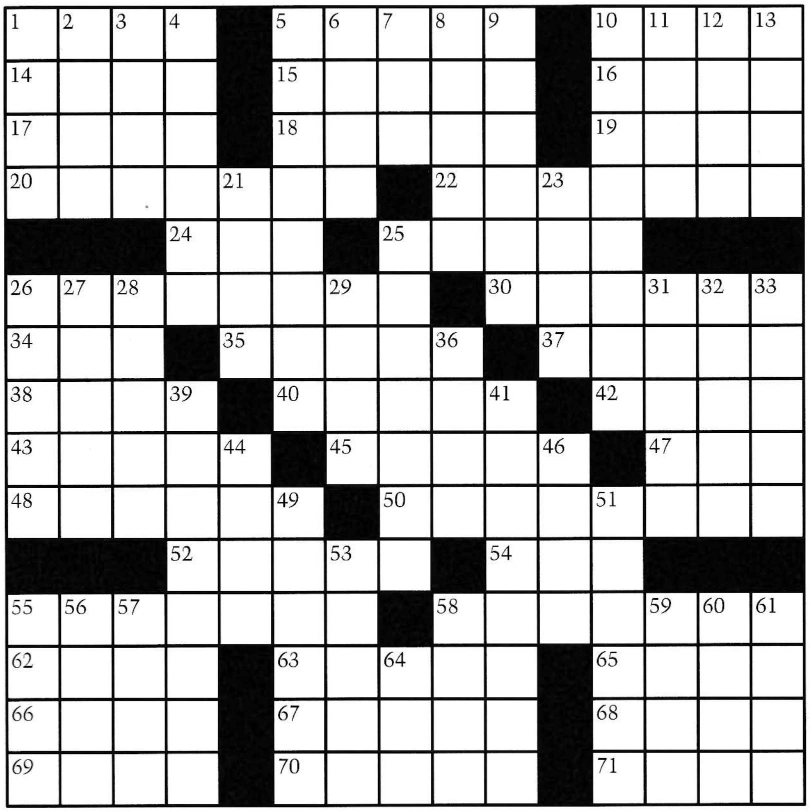
Here are the first ten persons who correctly timed Mr. Scall's perambulation:

Pasquale Nardone (Brussels, Belgium)
Johann Visser (Rotterdam, The Netherlands)
Pelle Hamberg (Lidingoe, Sweden)
Brian Platt (Woods Cross, Utah)
Keith Frikken (Dakota, Minnesota)
Oleg Shpyrko (Cambridge, Massachusetts)
Howard Brown (Idaho Falls, Idaho)
Bob Cordwell (Albuquerque, New Mexico)
How Yu Khong (Kuala Lumpur, Malaysia)
Grant Anderson (Portland, Oregon)

The newest *Quantum* CyberTeaser awaits your best efforts at www.nsta.org/quantum/contest.htm.

Criss × cross science

by David R. Martin



Across

1 Concern
5 Yellow, gummy pentosan
10 Philippine volcano
14 Plaintiff
15 Prepared
16 61,098 (in base 16)
17 Fleshy appendage
18 699,306 (in base 16)
19 Ski lift part
20 Wheels for belts
22 Molecular structure group
24 Superlative: suff.
25 Unit of magnetic flux
26 An opening
30 Artiodactyl of the camel
34 Scrap of food
35 Black
37 Carbonated beverages
38 Architect Mies van der ___
40 Microscopy dye
42 Archaic do
43 Dancer ___ Tharp
45 Issue
47 Born
48 One of a gene pair
50 Sense organ
52 Ropes or cables
54 That woman
55 C_5H_{12}
58 Element 33
62 Lab gel
63 Oil: comb. form
65 Tyndareus's wife
66 TV's Jay ___
67 Glashow–Weinberg–___ model
68 Roman poet
69 Element 26
70 Melville novel
71 Unit of heredity

Down

1 Basketball coach Adolph ___
2 Botanist Katherine ___
3 Unit of life
4 Early astronomer
5 Radiation source
6 Yes votes
7 Place for experiments
8 711,390 (in base 16)
9 African lake
10 Four-footed animal
11 Phosphatase unit
12 43,694 (in base 16)
13 ___ Onsager
21 Italian family
23 Building wings
25 Covariant and contravariant ___
26 Major blood vessel
27 Lurk
28 Type of alcohol
29 Plant part
31 ___ hinge (INV ZOO)
32 Cheese: comb. form
33 Mitosis star
36 Cry of pain
39 Fundamental particle
41 Nickel alloy
44 Healing plant
46 So-so grades
49 British math. W. ___ Brown (1866–1938)
51 1975 Chem. Nobelist Vladimir ___
53 Wait
55 Steep cliff (in Hawaii)
56 The Ohre in Germany
57 10^{-9}: pref.NANO
58 44,718 (in base 16)
59 Compacted snow
60 Chemical suff.
61 ___ oil (used in perfumes)
64 Swiss mountain

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MAY/JUNE PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M206

Let's suppose that none of the equations has a root whose absolute value is equal to 1. If x_1 and x_2 are the roots of the first equation, then we can show that $|a| = |x_1 + x_2| \le |x_1| + |x_2|$ $2 + 998 = 1000$. Indeed, $|x_1|$ and $|x_2|$ are positive whole numbers different from 1 and their product is equal to 1996. The sum of such numbers attains the greatest possible value when one of the numbers is equal to 2. Similarly we can determine from the second equation that $|a| \ge 2 \cdot 1994$. This contradiction shows that at least one of the equations has a root whose absolute value is equal to 1. Now, if we check all possibilities, we'll see that $a = -1997$.

M207

Let the points lie on side AB of triangle ABC in which $AB = x$, $BC = y$, $CA = z$. To make the situation clearer we can suppose that $y > z$. We will first calculate the distances of each point in question from P, using some common techniques for working with special points in a triangle.

Clearly, $BK = x/2$. To find the distance BP, we use the fact that an angle bisector of a triangle divides the side to which it is drawn in the ratio of the two other sides. This means that, for some number m, $BP = ym$ and $AP = zm$. Then $ym + zm = x$, so that $m = x/(y + z)$, and $BP = xy/(y + z)$.

To find BH, we locate points E and F where the triangle's inscribed circle touches sides BC and AC, respectively (fig. 1). Setting $BH = BE = p$, $AH = AF = q$, $CE = CF = r$, we find that $p + q = x$, $q + r = z$, and $r + p = y$. Adding, we obtain $p + q + r = (x + y + z)/2$, and $p = p + q + r - (q + r) = (x + y + z)/2 - 2z/2 = (x + y - z)/2$.

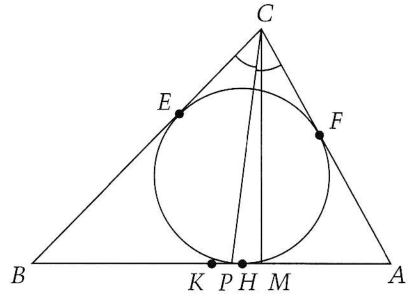


Figure 1

To find BM, we look at right triangle BCM, in which $BM = y \cos B$. The law of cosines will let us express $\cos B$ in terms of x, y, and z: $\cos B = (x^2 + y^2 - z^2)/2xy$. If follows that $BM = (x^2 + y^2 - z^2)/2x$.

Now it follows that

$$KP = BP - BK = \frac{x(y-z)}{2(y+z)},$$

$$KH = BH - BK = \frac{y-z}{2},$$

$$KM = BM - BK = \frac{y^2 - z^2}{2x},$$

and we can check that $KH^2 = KP \cdot KM$, or

$$KH = \sqrt{ab}.$$

M208

One way to find maximal or minimal values of a function is to look for the square of an algebraic expression. For real values of the variable, this square cannot be negative. In the present case, we can write

$$y = \frac{x^2}{8} + x\cos x + 2\cos^2 x - 1$$
$$= \frac{1}{8}(x - 4\cos x)^2 - 1.$$

Thus y cannot be less than -1. In fact, drawing the appropriate graphs, we can easily show that there are values of x such that $x + 4\cos x = 0$. Hence the value we seek is -1.

M209

We can verify the uniqueness of the real root of the equation $x^3 - x - 3 = 0$ by drawing the graph of its left side and finding proper extremal values. If we raise both numbers we are comparing to the fifth power and subtract one from the other, we arrive at the following problem: is the number $x^5 - 13$ greater or less than 0? (Here x is the root.) But taking into consideration that x satisfies the equation $x^3 - x - 3 = 0$, we find $x^5 = x^3x^2 = x^2(x + 3) = x^3 + 3x^2 = 3x^2 + x + 3$. Now we have to compare the number $3x^2 + x - 10$, where x satisfies the equation $x^3 - x - 3 = 0$, with 0. But the quadratic trinomial $3x^2 + x - 10$ vanishes when $x = 5/3$, while the only real root of $x^3 - x - 3 = 0$ is greater than 5/3 (the left side of the last equation is negative when $x = 6/3$). Thus $3x^2 + x - 10 > 0$ for this x—that is, for $x > \sqrt[5]{13}$.

So the calculator did not mislead us.

M210

For triangle ABB_1, the straight line BA_1 is the bisector of the angle adjacent to the angle $\angle ABB_1$. (This is a consequence of the fact that $\angle ABC = 120°$.) But, since AA_1 is the bisector of $\angle ABC$, the point of intersection of these two lines is equidistant from the lines BB_1, AB, and AC—that is, B_1A_1 is the bisector of $\angle BB_1C$. Similarly, B_1C_1 is the bisector of $\angle BB_1A$. Now it is clear that $\angle A_1B_1C_1 = 90°$.

Physics

P206

At rest the spring stretches a distance Δx_1 due to the load m (see figure 2):

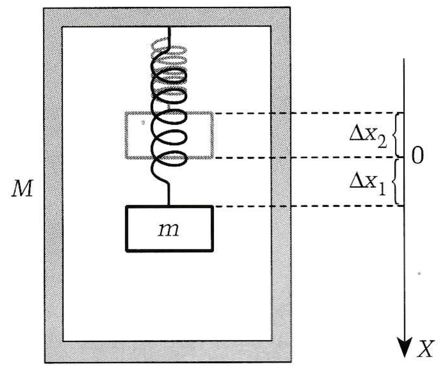


Figure 2

$$\Delta x_1 = \frac{mg}{k}.$$

The box is affected by the force of gravity $M\mathbf{g}$ and the tension $\mathbf{T}$ of the spring. The magnitude of the tension is $T = kx$. It is directed downward when the spring is stretched and upward in the opposite case. The box will start jumping when the tension T exceeds the force of gravity Mg—that is, when $T \geq Mg$. At the critical condition, $T = Mg$, which corresponds to the elongation

$$\Delta x_2 = \frac{Mg}{k}.$$

Therefore, the box will jump when the amplitude of the oscillations is equal to

$$A = \Delta x_1 + \Delta x_2 = (M + m)\frac{g}{k}.$$

P207

The best way to solve this problem is to graph the gas expansion and graphically calculate the work performed by the gas in the stages of interest to us.

For the first stage (40–80 l), this work is about 860 J; while for the second stage (140–180 l), it's approximately 250 J. The tabular data allow us to calculate the temperature of the gas at any point and thus to obtain the changes in the internal energy. In the first stage, the external energy decreases by 635 J, but in the second stage it increases by 93 J. Of course, these values are approximate, since we're calculating on the basis of a graph. Applying the law of conservation of energy, we shall find the amount of heat transferred to the gas in the first stage is $Q_1 = W_1 + \Delta U_1 = 225$ J. This means that in this stage the gas *received* heat, while its temperature dropped from 171 to 120 K. Similarly, in the second stage we have $Q_2 = 343$ J; in this stage temperature rose from 117 to 125 K, so the gas got *warmer*. In the first stage the specific heat is negative (approximately –4.5 J/(mole · K)), because the gas was heated, but its temperature dropped. In the second stage the specific heat is equal to about 45 J/(mole · K), so the ratio we seek is –10 (minus ten!).

P208

For the sake of definiteness we'll suppose that $\sigma_2 > \sigma_1$. Without external forces, the rod moves to minimize the surface energy. In doing so, the film with the larger surface tension will shrink, the other film will stretch, and the rod will shift. The force acting on the rod from the fluid on the right is $F_2 = 2\sigma_2 l$, and the force acting from the left is $F_1 = 2\sigma_1 l$. We took into account that a fluid film has two surfaces. To maintain the system at rest, an external force must be applied to the rod that is equal to

$$F = F_2 - F_1 = 2(\sigma_2 - \sigma_1)l.$$

P209

The energy stored in a capacitor does not depend on the way it has been charged. So we can choose the following mode of charging: one takes a small portion of charge from infinity and brings it to the spheres. First we charge the inner sphere. Let it have some charge q (which we've already delivered), and now additional charge Δq is added to the sphere. In doing so we perform the work

$$\Delta W = \frac{1}{4\pi\varepsilon_0}\frac{q\Delta q}{R_1}.$$

We can write this expression right off the bat, knowing the formula for the potential of a charged sphere $q/4\pi\varepsilon_0 R_1$. Or we can begin with Coulomb's law and write the expression for the work performed by moving along an elementary section of the path and integrate it for the entire trajectory of the test charge:

$$\Delta W = -\frac{1}{4\pi\varepsilon_0}\int_{\infty}^{R_1}\frac{q\Delta q}{r^2}dr = \frac{1}{4\pi\varepsilon_0}\frac{q\Delta q}{R_1}.$$

The total work required to transfer charge Q to the inner sphere is

$$W_1 = \int_0^Q \frac{1}{4\pi\varepsilon_0}\frac{q}{R_1}dq = \frac{Q^2}{8\pi\varepsilon_0 R_1}.$$

Now let's charge the outer sphere. For definiteness we assume that the charge of the inner sphere is positive, so the outer sphere must be negatively charged. Let that charge be q at some moment. When a small charge Δq is transferred, it will be affected by two fields: that of the charge $+Q$ (inner sphere) and that of $-q$ (outer sphere). So the elementary work performed during the transfer of the charge $-\Delta q$ from infinity to the outer sphere is

$$\Delta W = -\frac{1}{4\pi\varepsilon_0}\frac{Q\Delta q}{R_2} + \frac{1}{4\pi\varepsilon_0}\frac{q\Delta q}{R_2},$$

Correspondingly, the total work needed to charge the outer sphere is

$$W_2 = -\frac{Q^2}{4\pi\varepsilon_0 R_2} + \frac{Q^2}{8\pi\varepsilon_0 R_2} = -\frac{Q^2}{8\pi\varepsilon_0 R_2}.$$

Now the capacitor is charged, and its energy is

$$U = W_1 + W_2 = \frac{Q^2}{8\pi\varepsilon_0}\left(\frac{1}{R_1} - \frac{1}{R_2}\right).$$

All this work is concentrated between the spheres of the capacitor in the form of electrostatic field energy. Since $R_2 - R_1 \ll R_1$, the field can be considered homogeneous, and its strength is equal to

$$E = \frac{Q}{4\pi\varepsilon_0 R_1^2} = \frac{Q}{4\pi\varepsilon_0 R_2^2}.$$

Let's express the total energy of the electric field in terms of its strength:

$$U = \frac{Q^2}{8\pi\varepsilon_0}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \frac{Q^2}{8\pi\varepsilon_0}\frac{\Delta R}{R_1^2}$$
$$= 2\pi\varepsilon_0 E^2 R_1^2 \Delta R.$$

To find the energy density, this expression must be divided by the volume $V = 4\pi R_1^2 \Delta R$ occupied by the field:

$$u = \frac{U}{V} = \frac{\varepsilon_0 E^2}{2}.$$

We've been dealing with a normal capacitor, and strictly speaking this formula for the energy density corresponds to an electric field in a vacuum. Let's generalize this formula for a medium with dielectric constant $\kappa > 1$. In this medium the work necessary to transfer the same charge to the spherical capacitor will be less by a factor of κ. The electric field strength will be reduced proportionally, and the formula for the energy density will be

$$u = \frac{\kappa \varepsilon_0 E^2}{2}.$$

This doesn't mean that filling a charged capacitor with dielectric material results in an increase in energy density (remember, the charge is constant). Quite the opposite occurred in our case: the energy of the capacitor decreased by a factor of κ. However, if voltage rather than charge is been kept constant as the capacitor is filled with a dielectric, the electric field strength will be the same, and the energy of the capacitor will be κ times greater.

P210

The visible position of a star differs from the actual position due to the refraction of light in the atmosphere. The thickness of the atmosphere—that is, the altitude at which the air is practically absent and the refractive index is 1—equals several dozens of kilometers. It is far less than the Earth's radius, so in this problem we can assume that the atmosphere is a plane. Its refractive index gradually increases from 1 in the upper layers to $n > 1$ at the Earth's surface. Light from a star can be considered parallel rays incident upon the upper layers of the atmosphere at an angle $\pi/2 - \alpha$, where α is the true angular position of the star above the horizon. We observe the star at an angle $\beta > \alpha$ (fig. 3). According to Snell's law

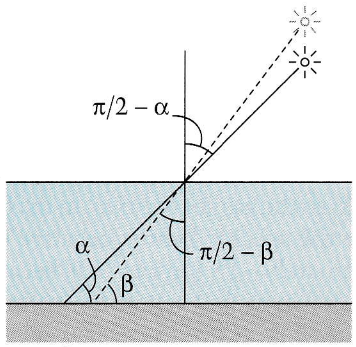


Figure 3

for a stratified medium,

$$\sin\left(\frac{\pi}{2} - \alpha\right) = n \sin\left(\frac{\pi}{2} - \beta\right),$$

or

$$\cos(\beta - (\beta - \alpha)) = (1 + (n-1)) \cos\beta.$$

Since $n - 1 \ll 1$ and so $\beta - \alpha \ll \beta$, we have the approximation

$$\cos\beta + (\beta - \alpha)\sin\beta = \cos\beta + (n-1)\cos\beta.$$

Therefore,

$$\beta - \alpha = (n-1)\cot\beta = 3 \cdot 10^{-4} \text{ rad} = 1'.$$

This is the measurement error for the angular altitude of a star at the surface of the Earth.

Brainteasers

B206

Draw two arbitrary circles passing through A with centers on the straight line (fig. 4). Denote the second point where they meet by B. Then AB is the perpendicular we seek.

207

In the situation described, the "eastern" windowpanes are arranged approximately parallel. So they can be considered fragments of

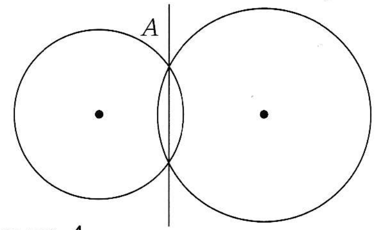


Figure 4

a single large mirror. As the Sun rises, its image in the mirror does the same. Thus the "illuminated" windows shift up the mountain slope. In addition, we note that in the Northern Hemisphere, they shift a bit to the south; but if the town is located in the Southern Hemisphere, they shift slightly to the north. At the equator, there is no shift sideways at the vernal and autumnal equinoxes.

B208

It follows from the statement of the problem that the car met the bank president when it was 20/2 = 10 minutes from the banker's house. So the total time Mr. Scall spent walking is 50 minutes (he stopped walking 10 minutes before the time the car usually arrived, and he started 1 hour before the car's usual arrival time).

B209

See figure 5.

B210

Here is one possible method: 1, 3, 7, 15, 31, 63, 127, 42, 85, 171, 343, 114, 229, 76, 25, 8.

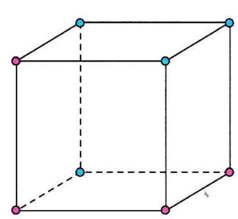


Figure 5

Index of Advertisers

Out standing in the field

The Holsteins vs. the Jerseys

by Dr. Mu

WELCOME BACK TO COWCULATIONS, THE column devoted to problems best solved with a computer algorithm.

When the spring sun warms the pasture and the green grass grows all around, a cow's thoughts naturally turn to grazing and baseball. My bovine friends love baseball. It's a nice, slow, lazy game. You can nip on the grass and shoo the flies away with your tail, and all the while be out standing in the field. And of course cows, like baseball professionals, enjoy a good chew.

Each summer two farm clubs, the Burlington Holsteins and the Waterford Jerseys, square off on the diamond out behind Paul's barn. They usually play all summer and try to get enough games so that at least one team wins fifty games. The first team to win fifty collects the prize of $1,000 put up by the BABE (Bovine Association of Baseball Enthusiasts).

Occasionally the summer weather in Wisconsin doesn't cooperate, making it difficult to get enough games played to permit a fifty-game winner. Last week, for example, one of those Midwest thunderstorms moved through the area and

Art by Mike Brenneman

turned the grass to a muddy swamp. So this year, with the game tally at 35 for the Holsteins and 41 for the Jerseys, the BABE had to call the series off and divide up the prize money. The problem, of course, is how to do this fairly.

The Holsteins, who are notoriously good at coming from behind to win the series, wanted to split the money evenly ($500 each), claiming they were robbed of their stretch run. The Jerseys favored an arrangement whereby the money would be divided based on how many games each team had won so far. Thus, the Jerseys would settle for 41/76 of the money, or $539.47.

But the BABE claimed that, according to the rules, in any series that does not go to completion the prize money must be split as follows: Beginning with the games won, cowculate the probability that one team (say the Jerseys) would have been the first team to win 50 games had the series continued. Assume that the chance that a team will win the next game they play is based on the games won so far. Thus, if the score is Holsteins: 35 and Jerseys: 41, the probability the Jerseys win the next game is 41/76. Of course this probability changes as the virtual series is played out. Once you know the probability that a team reaches fifty games first, multiply it by $1,000 to get the teams fair share.

This will take some cowculations.

COW 5. Write a program that will cowculate the winnings of each team based on BABE's rules. Report your answer for the series that ended Holsteins: 35 and Jerseys: 41. You are to assume that if the game score is currently at Holsteins: H and Jerseys: J, the probability that the next game will be won by the Holsteins is $H/(H+J)$, while the probability it is won by the Jerseys is $J/(H+J)$. Also, if P is the probability that the Jerseys will win fifty games first, then they should be awarded $P \cdot 1{,}000$ of the prize money.

Now it's your time to step up to the plate. Send your cowculation to drmu@cs.uwp.edu. To view all previous ruminations, take a peek at http://usaco.uwp.edu/cowculations.

This COW is headed for the plate.
It may be early, it may be late.
You whack the ball,
Or take the call,
But never, ever procrastinate.

—Dr. Mu

Solution to COW 4

Last time we introduced the following silage problem. As he does every year, Farmer Paul put up his feed corn last October in a huge silo capable of holding up to 120,000 pounds. In late fall, he began taking out a daily feed allotment of 300 pounds. Unknown to us, our silo was being broken into during the night and a fixed proportion removed. Each day Farmer Paul took 300 pounds for feed; each night the thief stole $(1/n)$th of the silage left in the silo (portion $= n$ is the same for each night). This was repeated for a total of five days and nights. The thief always stole an integer number of pounds. After the theft was stopped, Farmer Paul still had enough silage left to feed us for 210 days but not 211. In order to determine the proper monetary settlement, Special Agent Mark needs to determine exactly how many pounds of silage were stolen.

Before you do any coding on this problem, get out an envelope and do some "back of the envelope" cowculations. For example, given the constraints of the problem, what is the smallest portion that is even worth considering?

We know there is room in the silo for at most 120,000 pounds of silage. For five days and nights we subtract 300 and reduce the supply by a factor of (portion – 1)/portion. This is accomplished by the function **`reduce[silage]=(silage-300)(portion-1)/portion`**. Now the portion cannot be so small that after five repetitions of this process the amount of silage left is below the total feed for 210 days, or $210 \cdot 300 = 63{,}000$. The **`reduce[silage]`** function can be nested five times with the NestList command in Mathematica®.

Let's try it with portion = 8:

```
reduce[silage_]:=(silage-300)(portion-1)/
portion
portion=8;
NestList[reduce,120000,5]//N
```

```
{120000., 104737.5, 91382.8, 79697.5,
694.72.8, 60526.2}
```

It appears that portion = 8 will not work because we are left with only 60526.2 after five nights, which is below 63,000. However, for portion = 9 we have more than 63,000 pounds of silage left over:

```
portion=9;
NestList[reduce,120000,5]//N
```

```
{120000., 106400., 94311.1, 83565.4,
74013.7, 65523.3}
```

So we have established portion = 9 as the lowest value to begin our search for a solution There are two directions we can go. From top to bottom or bottom to top.

Top to bottom

First create a function **`TopToBottom[silage]`** that generates the amount of silage left after the silo has been reduced for five days and nights. Note that if **`TopToBottom[silage]`** is an integer, then the amount of silage left after each night is also an integer. So you don't need to check the silage is an integer after every night, only after all five nights. I'll leave that for you to ruminate on. Here's the function **`TopToBottom[silage]`**:

```
Clear[silage]
portion=9;
TopToBottom[silage_]=Nest[reduce,silage,5]
```

$$\frac{8\left(-300+\frac{8\left(-300+\frac{8\left(-300+\frac{8\left(-300+\frac{8(-300+\text{silage})}{9}\right)}{9}\right)}{9}\right)}{9}\right)}{9}$$

Begin with the silo filled with silage = 120,000 pounds of corn. As long as the **TopToBottom[silage]** function is not an integer (**!IntegerQ**), take a pound away (**silage--**) and try again. Once **TopToBottom[silage]** is an integer, you have a solution:

```
silage=120000;
While[!IntegerQ[TopToBottom[silage]],
silage—];
Print["Start ",silage,"  Finish
",TopToBottom[silage]];
Print["Thief stole ",silage-
TopToBottom[silage] - 1500]
Start 115698  Finish 63136
Thief stole 51062
```

Bottom to top

It's considered bad programming form to create a computer algorithm that simply tries all possibilities and looks for a solution by brute force. The silage problem presents such an opportunity. While brute force may not be able to be eliminated, it can be significantly reduced by looking at the problem differently, namely backwards. Let's start from the other direction, where we end up with at least 63,000 pounds of silage after five days and nights and go from bottom to top looking for an all integer match. First, we construct the **increase[silage]** function, which computes how much silage we had the day before some was fed to the cows and some stolen:

```
Clear[silage,increase]
portion=9;
increase[silage_]:= silage * portion/
(portion-1) + 300
```

We define the **BottomToTop[silage]** function, which uses the **Nest** command to go back five days and compute the amount of beginning silage. If **BottomToTop[silage]** is an integer, then we had an integer after every night:

```
BottomToTop[silage_]=Nest[increase,silage,5]
```

$$300+\frac{9\left(300+\frac{9\left(300+\frac{9\left(300+\frac{9\left(300+\frac{9\,\text{silage}}{8}\right)}{8}\right)}{8}\right)}{8}\right)}{8}$$

Begin with the silo depleted to silage = 63,000 pounds of corn. As long as the **BottomToTop[silage]** function is not an integer (**!IntegerQ**), add a pound (**silage++**) and try again. Once **BottomToTop[silage]** is an integer, you have a solution:

```
silage=63000;
While[!IntegerQ[BottomToTop[silage]],
silage++];
Print["Start ",BottomToTop[silage],"
Finish ",silage];
Print["Thief stole ",BottomToTop[silage]-
silage - 1500]
Start 115698  Finish 63136
Thief stole 51062
```

Going backwards, the program is 35 times faster and is essentially instantaneous.

Feedback

A group of three students—David Click, Mike Powers, and Laura Arthur—and their instructor, Thomas O'Neill , at the Shenandoah Valley Governor's School in Virginia examined the problem using Mathematica and ruled out a large number of portions greater than nine. Here is their solution, which is slightly modified to fit with the discussion given above. It tests all portions from 9 to 2560 and takes a few minutes to complete. The first time you try this program, change 2560 to 25 to get the answer quickly.

```
increase[silage_]:=silage*portion/(portion-1)+300
For[
  portion=9,portion<=2560,portion++,
  For[
     silage=63000,silage<63300,silage++,
     startSilage=Nest[increase,silage,5];
     If[IntegerQ[startSilage]&&
(startSilage<=120000),
        Print[
        "Start silage = ",startSilage,
        " End silage = ",silage,
        " Portion = ",portion,
        " Stolen = ",startSilage-silage-1500]
           ,Continue]
  ]
]
Start silage = 115698 End silage = 63136
Portion = 9 Stolen = 51062
```

From a practical standpoint, no respectable thief would bother taking less that 1/100 of the silage each night. Add this to the fact that Dr. Mu spotted a pickup truck full of silage and you must conclude that we have found the unique solution. From a purely mathematical standpoint, however, we need a proof that there is no solution for a portion $n > 2560$. Of course the insurance company would love to find one.

Vincent Béron, 18, a student at Collège de Bois-de-Boulogne in Montréal, Québec, Canada, also sent in a C program that found the correct solution.

Mathematica SILO

If you would like to learn how to cowculate with Mathematica, join me on the Internet at the Mathematica SILO (Summer Internet Learning Opportunity). During the week of July 28–31, I will ruminate on the fundamentals of Mathematica programming for all you cowhands. If you don't have a copy of Mathematica, I'll send you a trial CD with a full version. With Mathematica installed on your system, all you need to do is get on the Internet, come to our SILO (http://usaco.uwp.edu/silo), and start milkin' the COWs. I've got plenty of chores to keep you busy. Send me an e-mail at drmu@cs.uwp.edu to join in. Don't procrastinate—do it now!

Circle No. 2 on Reader Service Card