QUANTUM

JANUARY/FEBRUARY 1993

$5.00

The student magazine of math and science

SPRINGER INTERNATIONAL

The Feast of the Gods (1514) by Giovanni Bellini and Titian

"THE FEAST OF THE GODS" HAS INTRIGUED ART LOVERS FOR centuries. It was known that the original painting, by Giovanni Bellini, was partially reworked by Titian fifteen later, but why? And what was the extent of that revision? In 1985, when the National Gallery of Art in Washington decided to clean and repair this Renaissance masterpiece, the opportunity arose to scrutinize it with advanced scientific techniques. An X-radiograph taken in 1956 had indicated that Bellini originally painted a thicket of trees in the background. But it also suggested that the painting had been reworked before Titian (perhaps by Bellini himself).

Before repairing damaged areas, the staff at the National Gallery removed microscopic samples of pigment, which were subjected to both optical and chemical microscopy. X-ray diffraction powder analysis and energy-dispersion X-ray fluorescence analysis were also conducted. X-radiographs and infrared reflectographs were made of the "Feast" as well as works by Bellini, Titian, and Dosso Dossi.

Why Dosso? It turned out that Titian was indeed the second painter to rework the "Feast." At right is head conservator David Bull's reconstruction of the work after Dosso had repainted the background (from the National Gallery monograph *The Feast of the Gods: Conservation, Examination, and Interpretation*). Dosso's lush vegetation may have overwhelmed the action in the foreground. Titian was brought in to set things right, or perhaps simply to harmonize the Bellini with three other Titians in the collection of Alfonso d'Este, Duke of Ferrara. For some reason, though, Titian left Dosso's pheasant and some of his foliage. Analyses also established that Bellini made some of the alterations discernible in the 1956 X-radiograph.

For another look at how scientific techniques expose the underlying truth in works of art and other things, see "Physics Fights Frauds" on page 10.

QUANTUM

JANUARY/FEBRUARY 1993 VOLUME 3, NUMBER 3

Cover art by Dmitry Krymov

Voting is no joking matter. But sometimes it may seem that there is an unseen jester controlling the outcome of free elections. Maybe it's because people are fundamentally unpredictable. No matter how seriously we approach a choice, armed with good reasons and the best intentions, the final result can appear whimsical rather than logical. The way we balance conflicting data is often a mystery even to ourselves.

Many commentators have noted the broad disappointment that often follows a vote. How can that be? Didn't the majority make its choice? Why should most voters still be unhappy? Well, for one thing, your favorite candidate may have been knocked out in the primary election, leaving you to vote for your second choice, at best, in the November balloting. It may even be a matter of choosing the lesser of two evils.

There's more than one way to run an election, and each approach has its flaws. Turn to page 4 for an enlightening look at the mathematics of voting.

FEATURES

4 Majority Rules
Democracy and mathematics
by Valery Pakhomov

10 Historical Sleuthing
Physics fights frauds
by I. Lalayants and A. Milovanova

18 Recreational Math
A magnificent obsession
by Michael H. Brill and Michael Stueben

36 Sensing Techniques
Tactile microscopes
by A. Volodin

DEPARTMENTS

2 Publisher's Page

17 Brainteasers

25 How Do You Figure?

26 At the Blackboard I
Glancing at the thermometer . . .

32 Kaleidoscope
What harmony means

41 Math Investigations
The Worm Problem of Leo Moser

42 Physics Contest
Row, row, row your boat

46 At the Blackboard II
The science of the jump-shot

51 Happenings
US team places second at IMO ... Bulletin Board

54 Crisscross Science

55 Answers, Hints & Solutions

64 Toy Store
The stomachion

Taking advantage

"Bad times have a scientific value."—Emerson

THE AIRPLANE IS SOMEwhere over the North Sea. I am returning to the United States from my sixth visit to Moscow in four years. This editorial need not be written in midair. But I fear that the profound impressions gained on this visit will fade too quickly.

I write with a peculiar mix of sadness and anger, hope and despair induced by the impressions of the past ten days. Russia, as well as the other republics of the former Soviet Union, is going through some of the worst aspects of conversion to a free economy. Inflation is accelerating: the ruble has gone from 30 to the dollar in the summer of 1991, to 110 in March 1992, to 360 upon my arrival in Moscow on October 19. When I left, the ruble had reached 400 to the dollar. The essentials of life are in short supply, except for those with dollars or other hard currency. Increases in wages have lagged far behind inflation. Times are hard.

Economic and social renewal

On the other hand, I saw signs of hope. Buildings are under construction and the infrastructure is being repaired. New hotels have sprung into existence, and they are busy. Many Western and Asian business people scuttle about, some genuinely trying to help, but most looking for ways to make a quick buck off the Russian people's misery.

Yet Muscovites seem more spirited. They smile more and seem more animated. For example, I attended a performance of Prokofiev's ballet "Cinderella" at the Palace of Congresses in the Kremlin. The dancing was superb, but there was no orchestra in the pit—they were dancing to a phonograph! My hosts, who were accustomed to the great Russian tradition of dance and music, were mortified. It was unheard of. But, leaving the performance, playful and laughing children could be seen running here and there, ignoring the barriers and walkways that guards had enforced so strenuously in former times. And this time no one really cared. Such perfectly acceptable and normal chaos with children is surely the first sign of a free society.

Shameless exploitation

In a more serious vein, I found two things particularly disturbing. One is the way the United States, apparently as a matter of official policy, would exploit Russian scientists in ways that are degrading and insulting. The other is the very likely wholesale theft of the intellectual property that Russian scientists, engineers, and technologists have created over the past several years. This is perpetrated by American, German, and Japanese corporations under the guise of "cooperation."

I learned of a proposal whereby distinguished Russian scientists would be employed on tasks of interest to the United States—participating, for instance, in projects to disassemble and destroy weaponry or to improve the environment. The idea is to prevent these scientists from leaving Russia and the other republics, to hold onto the intellectual resources needed to create a free economy. It would also discourage them from moving to nations like Iraq or Iran to work on military projects.

Since Russians do not want to leave their homeland, they would do it only out of desperation. They would do it to save their families. Here's the problem with the proposal, though: these distinguished scientists are being offered ridiculously low wages—apparently on the order of \$100 a month. Even though this represents some 40,000 rubles a month at the current rate of exchange—more than any of them could make in Russia—the amount is both inadequate and insulting. An offer like this was actually made by an American who didn't hesitate to pay a young, relatively uneducated guide-interpreter \$50 a day. Taxi drivers working the runs from Arbat Street to local hotels routinely take in as much as \$20 a day (each trip netting \$2 or \$3).

Now you can understand why Russian scientists are rejecting offers to work on behalf of Americans, and how our image is being tarnished by this shameless exploitation. We should take advantage of the situation, not people.

Protecting marketable ideas

As for intellectual property rights, while in Moscow I visited several institutes, including our founding editor Yuri Ossipyan's Solid-State Physics Institute at Chernogolovka, a village north of Moscow that is home to nine scientific and technological institutes. Yuri's institute is remarkable in that it's vertically integrated—that is, it includes theoreticians, applied physicists, technologists, and engineers, production as well as research facilities. An idea can be taken from theory to production, all in the same place—something unheard of in US science or technology, but common in the most successful American retail businesses (for instance, Walmart).

At the Solid-State Physics Institute I saw remarkable processes involving high-temperature superconductors and the growth of huge, single crystals of metals like tungsten with properties never observed in their normal state. I saw sapphire and ruby crystals grown with threads, sapphires grown in the shape of a bead with a hole in it, and other artificial gems with useful characteristics. Such products of Russian science are everywhere, and applications in industry are plentiful.

But new processes and ideas like these are presently unprotected. The US could help by providing, as freely given aid, legal and technical assistance in obtaining American patents. US patents would provide worldwide protection for Russians as they try to create a new economy based on their own research and development capabilities.

Forging a new partnership

It's common for Americans and others to vastly underestimate Russian science and technology because of the economic disaster created by communism. But that is a serious mistake. These scientists and engineers are among the finest in the world. They can take inferior technologies and outperform the best technologies from other countries. Consider their success with their space vehicles and aircraft. They could just as easily create the best consumer products in the world.

We must assist the republics of the former Soviet Union, not just for their benefit alone. We have much to gain as well. This bloc of nations will be the greatest trading partner in the world in only a few years. Let it be the main partner of the United States.

—Bill G. Aldridge

QUANTUM

THE STUDENT MAGAZINE OF MATH AND SCIENCE

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

Publisher
Bill G. Aldridge, Executive Director, NSTA

Associate Publisher
Sergey Krotov, Director, Quantum Bureau

Founding Editors
Yuri Ossipyan, President, Quantum Bureau
Sheldon Lee Glashow, Nobel Laureate (physics), Harvard University
William P. Thurston, Fields Medalist (mathematics), University of California, Berkeley

Field Editors for Physics
Larry D. Kirkpatrick, Professor of Physics, Montana State University, MT
Constantine Bogdanov, Professor of Physics, Moscow State University

Field Editors for Mathematics
Mark E. Saul, Computer Consultant/Coordinator, Bronxville School, NY
Vladimir Dubrovsky, Associate Professor of Mathematics, Moscow State University

Managing Editor
Timothy Weber

Art Director
Elisabeth Tobia

Staff Artist
Sergey Ivanov

Editorial Consultants
Alexander Buzdin, Professor of Physics, Moscow State University
Yuli Danilov, Senior Researcher, Kurchatov Institute
Larissa Panyushkina, Managing Editor, Quantum Bureau

International Consultant
Edward Lozansky

Advertising Managers
Paul Kuntzler (Washington office)
Thomas Heitzman (New York office)

Advisory Board
Bernard V. Khoury, Executive Officer, AAPT
James D. Gates, Executive Director, NCTM
George Berzsenyi, Professor of Mathematics, Rose-Hulman Institute of Technology, IN
Arthur Eisenkraft, Science Department Chair, Fox Lane High School, NY
Karen Johnston, Professor of Physics, North Carolina State University, NC
Margaret J. Kenney, Professor of Mathematics, Boston College, MA
Thomas D. Rossing, Professor of Physics, Northern Illinois University, IL
Alexander Soifer, Professor of Mathematics, University of Colorado—Colorado Springs, CO
Barbara I. Stott, Mathematics Teacher, Riverdale High School, LA
Carol-ann Tripp, Physics Teacher, Providence Country Day School, RI

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

Advertising:

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

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

Copyright © 1993 NSTA. Printed in U.S.A.

Subscription Information: North America

Student rate: $14; Personal rate (nonstudent): $18; Institutional rate: $30; Single Issue Price: $5. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus, NJ 07096; or call 1-800-SPRINGER (in New York, call (201) 348-4033).

All Countries Outside North America

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

Democracy and mathematics

Problems and paradoxes in free elections

by Valery Pakhomov

THEY SAY THAT POLITICS IS an art. That may be true, by and large. But it's not just an art. The role of mathematics in politics is more important than you might think. We all know the basic principle of democracy: the most important decisions should be made by the direct vote of the widest social strata. Such an approach seems to satisfy everybody. But in actual practice, the number of voters who are discontented after every election is comparable to, if not greater than, the number of those satisfied with its results. Is it just a matter of chance? Mathematics will help us examine this question in detail.

The rules of winning

I'll start with the problem of choice, in the broadest sense of the word. It may be the choice of an official, a draft constitution or law, an approach to land development, a winner in a competition, the "person of the year," and so on. What's really significant is that only one of several alternatives must be chosen, and necessarily by a vote.

For the sake of definiteness, let's talk about an election. Suppose there are n voters and m candidates. Each voter ranks the candidates in a certain order according to his or her preference. This *system of individual preference* for a particular voter x and for, say, three candidates a, b, and c is written

$$a \overset{x}{\gtrsim} b \overset{x}{\gtrsim} c$$

where a is the candidate most preferred by citizen x, b is the second best, and c is the worst of the three in x's opinion ("$a \overset{x}{\gtrsim} b$" is read "a is better (for x) than b"). Summarized in a table, the systems of individual preference for all voters constitute the *vote profile*. For instance, if there are $n = 17$ voters electing one of $m = 4$ candidates a, b, c, and d, and if 5 voters order the candidates as $a > d > c > b$, 3 other voters as $a > d > b > c$, another 5 as $b > c > d > a$, and the remaining 4 as $c > d > b > a$, we get the vote profile A shown in table 1. This profile will serve as the main "proving ground" for various voting rules. We'll see that treating this far from remarkable situation according to different rules, which all seem quite reasonable and fair, leads to unexpectedly different, sometimes opposite results.

Relative Majority Rule. Each voter casts exactly one vote for one candidate. The candidate receiving the greatest number of votes wins.

To determine the winner here we need only the first row of the vote profile; for profile A, the total number of votes for each candidate is displayed in the second column of table 2. So, according to this rule, the winner is candidate a.

Absolute Majority Rule. Again, each voter casts one vote for one candidate. But a candidate must collect

Table 1

	Number of votes				
Rank	5	3	5	4	Points
1	a	a	b	c	3
2	d	d	c	d	2
3	c	b	d	b	1
4	b	c	a	a	0

Profile A

Table 2

Candidate	Number of wins	Score	Duels against			
			a	b	c	d
a	**8**	24	–	8:9	8:9	8:9
b	5	22	**9:8**	–	8:9	5:12
c	4	27	**9:8**	**9:8**	–	**9:8**
d	0	**29**	**9:8**	**12:5**	8:9	–

Applying winning rules to Profile A. Entries in boldface are wins.

Art by Dmitry Krymov


QUANTUM
ABC

more than half of all votes to win. If there is no such candidate, a second round of voting is conducted for the two candidates who got more votes than the rest. In the second round, the winner is determined by a majority (which is necessarily more than half of all votes).

Applied to our profile A, this rule brings candidates *a* (8 votes) and *b* (5 votes—see the second column in table 2) to victory in the first round. Crossing out *c* and *d* from the profile, we get the profile A′ for the second round:

Number of votes			
5	3	5	4
a	*a*	*b*	*b*
b	*b*	*a*	*a*

=

8	9
a	*b*
b	*a*

Profile A′

Profile A′ shows that the election is won by *b*.

In this case we had to use not only the first row of profile A but also the information about the relative placement of the winners of the first round (*a* and *b*) in the voters' opinion. To use the next rule, we'll need to know the total numbers of the first, second, ... places in all systems of individual preference for each candidate.

Highest Score Rule. All voters display their entire systems of preference. A candidate gets 0 points for last place, 1 point for next-to-last place, 2 points for third-last, and so on (the last column in table 1). The highest total score wins.

A variant of this rule is often used in sports, when competitors are ranked first, second, third, and so on, by a team of judges. The victor is the competitor whose rankings have the smallest sum. For profile A, all the individual scores of the candidates are given in the third column of table 2: *d* wins, *c* comes in second, while *a* and *b*, our previous winners, lose by significant margins.

Now let's try to compare candidates in pairs. What do the voters think about, say, *d* and *a*? As profile A indicates, 8 voters prefer *a* to *d*, but 9 think *d* is better. We may say that *d* wins a "duel" with *a* by a vote of 9 to 8. Similarly, comparing *d* and *b* we conclude that *d* wins even more convincingly—by a 12–5 vote. The results of all the duels are presented in the right part of table 2: candidate *c* wins in duels with all of his rivals.

Condorcet Rule. The candidate who wins all duels wins the election.

This rule is named after the Marquis de Condorcet (1743–1794), a French philosopher, mathematician, and political figure of the Age of Enlightenment, who noticed the paradoxes that emerge when the winner of an election is determined according to the "duel rule."

Candidate *c* wins the election according to the Condorcet Rule. Sometimes it doesn't determine the winner, though: Condorcet himself discovered that some distributions of votes create "vicious circles"—when, say, *a* beats *b*, *b* beats *c*, and *c* beats *a*. But the three other rules aren't universally applicable either—they may fail when two or more candidates get the same number of votes or points (which is rare when the pool of voters is large).

Paradoxes of voting

The four rules we considered above present four different concepts of what the best choice is from the point of view of the "collective voter." Not only do they result in four different candidates winning the election, the winner according to one rule may turn out to be the worst according to another. For instance, the relative-majority winner for our vote profile A is the worst according to Condorcet, while the absolute-majority winner has the lowest total score. So, with this profile, whichever rule we care to use, the winner it establishes may not satisfy more than half of the voters (who, naturally, will think another rule is more fair)! But

Table 3

Number of votes			
5	3	5	4
d	*d*	*b*	*c*
c	*b*	*c*	*d*
b	*c*	*d*	*b*

Profile B

→

12	5
d	*b*
b	*d*

Profile B′

Candidate a *withdraws. Profile B shows the first round of an absolute-majority election; Profile B′—the second round, in which* d *wins.*

the problem of choice is fraught with even more striking twists for the contenders.

As we already know, the winner in profile A according to the Absolute Majority Rule is candidate *b*. It wouldn't be unnatural that candidate *a*, who doesn't stand a chance to win, withdraws his (or her) candidacy. Erasing *a* from table 1, we get the vote profile B (table 3). Now we see that candidates *d* and *b* (8 and 5 votes out of 17, respectively) survive to the second round, where (see profile B′ in table 3) *d* wins by 7 votes! So *a* can manipulate the result of the election even though *a* cannot win.

Problem 1. Think of a similar example of manipulation for the Highest Score Rule.

In the end, the situation verges on the absurd.

Table 4 presents two vote profiles, C and D, differing only in the last column, where *a* and *b* are switched. You can imagine that persuasive agitation for *a* won over the

Table 4

Number of votes			
6	5	4	2
a	*c*	*b*	*b*
b	*a*	*c*	*a*
c	*b*	*a*	*c*

Profile C

Number of votes			
6	5	4	2
a	*c*	*b*	*a*
b	*a*	*c*	*b*
c	*b*	*a*	*c*

Profile D

Two voters (right column) change their minds in favor of a *against* b, *causing* a *to lose the election according to the Absolute Majority Rule.*

two voters represented in the last column to a's side. According to the Relative Majority Rule, there is a tie between a and b under profile C—a tie unforseen by the rule and thus unbreakable—whereas the agitation results in a's victory (under profile D). Suppose, however, the winner is determined by the Absolute Majority Rule. Then, under profile C, a and b win the first round, and a wins the second round (check it!). But the agitation that changed the two voter's opinion in favor of a against b makes a lose: a and c win the first round, and c the second!

I could have given a lot of other voting rules and paradoxes to convince you that there are no absolutely perfect rules.[1] All these considerations suggest the discouraging (for many) conclusion that democracy—as the expression of the will of the majority—doesn't exist at all, because, as we've seen, the very notion of the "opinion of the majority" doesn't exist. Then what's the attraction of democracy for us? Perhaps the same as its danger—the possibility of manipulating and influencing the results of elections without breaking the law.

A function of collective preference

Now imagine n members of a parliament working out the priorities of how the state is to finance a number of social programs. For instance, they may put public health first, pay for an education program from the rest of the budget, finance a housing program from the remainder, and so on; or they may arrange the programs in some other way. Each member of parliament has an individual opinion about how the programs should be ordered. The problem is to create a ranking that reflects the collective opinion of the parliament as a whole.

Mathematically, there's no difference between ordering programs and ordering candidates, so we'll talk about n voters and m candidates as before. The real distinction is that now we need not only the "best" candidate (or program) but a list of all the candidates in the order of "collective preference." More formally, we want to construct a rule, called a *function of collective preference*, assigning a certain "collective" order $a \overset{P}{\gtreqless} b$ to every possible vote profile P. That is, given a profile P and any pair of candidates a and b, this rule must tell us whether a is better than b ($a \overset{P}{>} b$), a is "equal" to b ($a \overset{P}{=} b$), or a is worse than b ($a \overset{P}{<} b$). (So now we allow for a third possibility: "a and b are equally preferable for the electorate as a whole.") When it's clear that profile P is being considered, I'll omit P in the notations $\geq$, $>$, and so on.

Naturally, such a function must satisfy certain requirements. For instance, in the case of two candidates a and b, the only reasonable way to define profile P is to count the number n_a of voters that prefer a to b and the number n_b of those preferring b to a, and to set $a > b$, $a = b$, or $a < b$ if $n_a > n_b$, $n_a = n_b$, or $n_a < n_b$, respectively. By the way, it's not hard to see that all four rules of voting considered above coincide in this case, and our function can be described quite simply: "If there is a winner in the election, then the winner is better than the loser; otherwise, the candidates are equal."

The simplest way to extend this definition to more than two candidates is as follows. Let S be the set of candidates, R some "rule to win" the election. If $R(S)$ is the set of winners according to this rule (there may be several), all of them are considered equal and better than all the rest. Then we exclude these winners from the set S—formally, take the "difference" $S_1 = S\backslash R(S)$ of sets S and $R(S)$—and apply our rule to the remainder S_1 to get the set $R(S_1)$ of candidates that are awarded second place; again take the difference $S_2 = S_1\backslash R(S_1)$ and determine the "bronze medalists" $R(S_2)$; and so on.

For example, let there be 3 voters and 4 candidates a, b, c, d with the following vote profile E:

Number of votes		
1	1	1
a	d	b
b	a	c
c	b	d
d	c	a

Profile E

Let R be the Relative Majority Rule. Then each of the candidates a, b, d receives one vote and $R(S) = \{a, b, d\}$. They take first place and the remaining candidate c takes second place, so the collective preference can be written as $a = b = d > c$.

Problem 2. Apply the above "method of consecutive exclusions" to profile A (table 1), taking R to be each of the four rules considered above. Verify that the resulting collective order for the Relative Majority Rule is $a > d > c > b$, for the Absolute Majority Rule $b > c > d > a$, for the Highest Score Rule $d > c > b > a$, and for the Condorcet rule $c > d > b > a$.

This problem shows again that different rules give different results (the first two are opposite). Notice that the Relative Majority and Highest Score rules allow a direct arrangement of the candidates according to the number of votes or the score each receives. And these two methods yield still new orderings: $a > b > c > d$ and $d > c > a > b$! In addition, the Absolute Majority and Condorcet rules are inapplicable to some profiles (like profile E above), although we allowed for the equivalence of candidates.

Our earlier investigations must have prepared you for such "minor" contradictions. But now I'm coming to a theorem that will astound you. Roughly speaking, it says that a function of collective preference that complies with the conventional concept of democracy and, at the

[1] See also problem M75 in this issue and M9 in the May 1990 issue.—*Ed.*

same time, satisfies most reasonable requirements simply does not exist!

I'll present these natural requirements as a set of four *axioms of collective preference.*

COMPLETENESS AXIOM. *A collective order must be defined for every pair of candidates.*

TRANSITIVITY AXIOM. *If $a \geq b$ and $b \geq c$, then $a \geq c$.*

In other words, we want to be able to compare any two candidates, and we want an order to be an order—that is, we want to preclude the situation in which a is better than b, b is better than c, but c is better than a.

UNANIMITY AXIOM. *If all voters think a is better than b, then the collective opinion must be the same: $a > b$.*

This unassailable requirement rules out the function declaring all candidates equal regardless of the vote profile.

INDEPENDENCE AXIOM. *The final (collective) order of any two candidates depends only on their mutual order in individual preferences and does not depend on the arrangement of other candidates.*

This axiom rules out the possibility of manipulations—when, for instance, the withdrawal of a candidate influences the mutual order of the other candidates.

One example of a function that obeys all four axioms is the *Dictator Rule,* which establishes the order determined by one particular voter (the "dictator") without regard to the opinions of other voters. Of course, we can't call such a rule "democratic."

Another example is the rule for two candidates considered above (in this case, transitivity and independence are trivial). But this rule is in fact a rule of choice, not of ranking—it can't be applied to more than two candidates.

Arrow's Theorem

The unexpected result I'm going to prove is that in spite of the rationality of all our assumptions, *the only rule that satisfies them (and applies to any number of candidates) is the Dictator Rule.* This notion traces its lineage to the American economist Kenneth Arrow (b. 1921).

I'll prove Arrow's Theorem in three steps. But first, I need to introduce a couple of important definitions. Any subset C of the set V of all voters will be called a *coalition.* A coalition C will be called *decisive for candidate a against b* if $a > b$ whenever all members of the coalition prefer a to b while all other voters prefer b to a. And, finally, a coalition that is decisive for *any* pair of candidates will be called simply *decisive.*

A decisive coalition is *almost* a dictating coalition: it imposes its opinion about any candidates provided that all other voters are of the opposite opinion. I'll show that one can always find a decisive coalition consisting of only one voter, and that in fact the opinion of the other voters is irrelevant.

Step 1. *There exist two candidates a and b and a voter such that the coalition consisting of only this voter is decisive for a against b.*

There is at least one *coalition decisive for some pair of candidates*—the coalition V of all voters (it's decisive by the Unanimity Axiom). Since the collection of such coalitions contains at least one member, we can choose from it one coalition (let's call it D) that has the smallest number of members for any coalition that is decisive for some pair of candidates. This number is greater than zero, because the empty coalition can't be decisive for any a and b (if nobody prefers a to b, then everybody prefers b to a, and the Unanimity Axiom yields $a < b$, and not $a \geq b$).

On the other hand, the minimal coalition D we've chosen can't have more than one voter in it. Indeed, assume this is not true. Divide D into a one-voter coalition $\{v\}$ and the nonempty coalition E of all other voters in D. Suppose D is decisive for a against b and consider the following profile:

Coalition		
$\{v\}$	E	$V \setminus D$
a	c	b
b	a	c
c	b	a
...	...	...

(Here c is any candidate other than a or b, and the dots mean arbitrary arrangements of other candidates—they are rendered irrelevant by the Independence Axiom.) Since D is decisive for a against b, all the voters in D prefer a to b, and all other voters prefer b to a, we have $a \geq b$. If $c \geq b$, the coalition E would be decisive for c against b, which would contradict the choice of D as a coalition of minimal size (because E comprises a smaller number of voters, and D was chosen to be minimal over all pairs of voters). Therefore, $b > c$.

Now $a \geq b$, $b > c$, and, by transitivity, $a \geq c$. But this means that the coalition $\{v\}$ is decisive for a against c, in contradiction of the minimality of D! This contradiction completes step 1: D consists of only one voter. Let's call this voter k.

Step 2. *The coalition $D = \{k\}$ from step 1 is decisive.*

We know that $\{k\}$ is decisive for a against b. Take any other candidate c, and consider this profile:

$\{k\}$	$V \setminus \{k\}$
a	b
b	c
c	a
...	...

According to the choice of $\{k\}$, $a \geq b$. And, by the Unanimity Axiom, $b > c$. It follows by the Transitivity Axiom that $a \geq c$. Since this result must not depend on the position of candidate b (the Independence Axiom), our function of collective preference must declare $a \geq c$ whenever k prefers a to c and all the other voters prefer c to a. This means that

$\{k\}$ is decisive for a against c for *any* candidate c.

Problem 3. Use a similar argument to prove that $\{k\}$ is decisive for d against c for an arbitrary candidate d.

So $\{k\}$ is decisive for any d against any c, and we are ready to take the last step.

Step 3. *Voter k is a dictator.*

So far we can't be sure that the collective opinion always coincides with k's opinion: this was shown to be true only when all other individual opinions are opposite to k's. All we have to do is prove that this dependence on the opinion of the other voters is only apparent.

Suppose k prefers a to c and c to b; as to the other voters, let them believe that candidate c is the best and think whatever they want about all the other candidates, including a and b. Since $\{k\}$ is a decisive coalition, and a is better than c for k and worse than c for all the rest, $a \geq c$; by the Unanimity Axiom, $c > b$. Then, by transitivity, $a \geq b$. By the Independence Axiom, the collective opinion about a and b doesn't depend on c. Thus, we've proven that for any two candidates a and b, if k prefers a to b, then, whatever the other voters think about a and b, the collective preference puts a higher than b. And this completes the proof of Arrow's Theorem.

So what do we have in the end? Whereas dictatorship is a sufficiently clear concept satisfying simple principles, it's impossible to define democracy other than as the alternative to dictatorship. Nonetheless, democracy attracts people, since it is the natural political environment for the development of society. Only in this environment can the "strongest" and "wisest" be naturally selected, because to win an election or get others to carry out your decisions, you often need more than the ability to persuade—you need to be able to calculate.

But this is democracy's scourge as well. Calculation too often allows manipulation of the results. And maybe this is why the political winds in society continually swing from democratic to dictatorial. ◙

Until last summer **Valery Pakhomov** *was a professor of mathematics in the economics department at Moscow State University. He is now the Russian trade representative in central Africa. A mountaineer and collector of classical music recordings, Dr. Pakhomov also taught mathematics for many years at the Kolmogorov Math and Science School of Moscow University.*

ANSWERS ON PAGE 58

Back issues of
QUANTUM
are available

For more information and prices, call **1-800-SPRINGER**

Or send your order to:

Quantum Back Issues
Springer-Verlag New York, Inc.
PO Box 2485
Secaucus, NJ 07094

Why the Sky Is Not the Limit at Beloit College.

Dan Schroeder
Astronomer and Physicist
Hubble Space Telescope Research Team Member
Beloit College Graduate and Professor

Only one scientist who's not affiliated with a major research university was on NASA's team that built the $1-billion Hubble space telescope. He's Dan Schroeder, who went from Kiel, Wisconsin (population 3,087) to Beloit College (population 1,100) and the top of his profession. A world-class researcher, he's also a great teacher, three times voted Beloit College's "Teacher of the Year." He's just one reason that this historic Wisconsin school ranks among America's top 50 undergraduate colleges in producing graduate scientists and is a place young men and women learn to reach for the stars.

Beloit
College

The Results Speak For Themselves.

For more information about Beloit College, call 1·800·356·0751
(in Wisconsin call collect, 608·363·2500).

Circle No. 15 on Reader Service Card


BASYROV G.

Physics fights frauds

Carbon dating and other revealing techniques

by I. Lalayants and A. Milovanova

FOR CHRISTIANS EVERYwhere, Easter is a celebration of Christ's resurrection after his death on the cross. On this day people in Russia used to give each other Easter eggs as a sign of prosperity and fertility. In 1885 Tsar Alexander III asked Carl Fabergé, his court jeweller, to make something unique to be given as a gift. Following the lead of his active imagination, Fabergé created a miniature masterpiece: an egg decorated with laurel leaves and white and pink flowers made from precious stones.

But it could well have been a dark-green egg on a twisted golden support decorated with five pansies. Or, perhaps, a green malachite egg with an "A" monogram and golden garlands. The egg's lid opened, exposing a dark-green replica of Alexandrovsky Palace in Tsarskoye Selo. These masterpieces could be enumerated and described endlessly.

Over the course of more than 30 years, Fabergé made over fifty lavishly decorated eggs for the emperor and more than 300 simple enamel eggs. These undecorated eggs are valued highly nonetheless.

Fabergé died in 1920 in Paris. A year later, his house in Moscow was occupied by a young American businessman, Armand Hammer, who literally saved the capital of the newborn republic from hunger. In exchange for American bread, Hammer would take the tsar's jewelry, which was afterwards put on sale in the fashionable shops of New York City. Only ten Fabergé eggs can now be seen in the Kremlin museums.

Art by Garif Basyrov

The American publisher Malcolm Forbes had eleven Fabergé eggs. Forbes bought his eleventh egg, one that had belonged to the Russian emperor, in June 1985 at the stunning price of $1,760,000. The news reached another collector of Fabergé eggs and other items of Russian art, E. Ariet, who owned some 100 undecorated eggs. Not long before Forbes made his purchase, Ariet had bought his first imperial egg at a Christie's auction in Geneva. It is said to have been ordered by Tsarina Alexandra for her husband Nicholas II. Concealed inside the egg was a figure of the tsar on horseback. Ariet paid $250,000 for it.

After he heard of Forbes's purchase, Ariet got in touch with Christie's Manhattan office, asking the firm to put his own Fabergé egg on sale. Two weeks before the auction, however, the staff at Christie's announced that the famous auction house would not sell the egg: it was a fake.

Photo by Larry Stein. The FORBES Magazine Collection, New York

Fabergé's "Lillies of the Valley Egg."

Shrouded in mystery

The disappointment suffered by one individual is almost nothing compared to what happened in 1989, when millions of Catholics learned that the famous Shroud of Turin, in which Christ's body had reputedly been wrapped when it was taken down from the cross, was made 1,300 years too late. This was firmly established by accepted scientific methods. So when the results of physical research on the shroud's fabric were shown to Pope John Paul II, he said, "Make them known." Two weeks later the data were published in the scientific journal *Nature*.

In case the issue is shrouded in darkness for you, I'll provide a little background. In the northern Italian city of Turin, in a cathedral designed by Giovanni Battista (Italian for "John the Baptist," by the way), a piece of old linen is kept in a crystal ark. The cloth is 4 meters long and more than a meter wide. Its story begins in 1356, on the estate of the Count de Charnie, a crusader. The count claimed that it was the very shroud in which Christ's body had been wrapped before it was entombed. This was confirmed by a legend according to which the crusaders saw the shroud in the vicinity of Constantinople in 1203. It seemed doubtful, however, that the cloth could be kept more than 1,000 years: any fabric falls to pieces after 200–300 years.

It should be mentioned that the bishop of Trois (near Paris) called it a fake right away. The Vatican, relying on the bishop's letter, also doubted its authenticity. Nevertheless, the church on the count's estate (near Paris) overflowed with pilgrims. Nobody was puzzled by the fact that Christ was shown bearded on the shroud. Old Roman pictures depict him beardless. The same is applicable to the picture on the walls of Roman catacombs as well as to the inlay of the temple at Ravenna, erected at the end of the 5th century. A crucifix carved in ivory (approximately A.D. 420) is no exception.

In 1963 a Roman floor inlay was discovered on the Hinton St. Mary estate in southwest England. It depicted Jesus surrounded by the four evangelists. To prevent confusion, Christ was marked by the Greek letters "X" and "P" (the first letters of Χριστος) above his head. Again, he is beardless.

In spite of Pope Clement VIII's verdict (1389) rejecting the shroud's au-


UPI/Bettmann Newsphotos

A portion of the controversial Shroud of Turin.

thenticity, in 1578 the duke of Savoy managed to have the shroud transported to Turin, which became the capital of Savoy. (The shroud had been handed down to him by the Count de Charnie's granddaughter.) Coincidentally, in the 16th century an Italian master painted a picture depicting Jesus at the foot of the cross, wrapped in a white cloth. In the upper part of the painting, another piece of white cloth had been drawn. It bears the imprint of Christ's body . . .

A great controversy swirled around the Shroud of Turin and its "imprint" in the 20th century. In 1931 a reverse photograph of the man's face appeared and created quite a furor. Another photograph showed an imprint of the entire body on the shroud. In the photo the man's legs were outstretched and his arms crossed over the lower part of his abdomen. Rumor had it that the imprint on the shroud was of supernatural origin.

The body was 180 cm tall and estimated to weigh about 77 kg. A wound was visible on the left wrist—was it left by crucifixion? Lash marks could be seen on his shoulders and back. There were scratches as well. This meant that he had been carrying something heavy and rough just before his death—was it a cross? In the 1970s, when all these details came to light, there was no other method of proving the shroud's authenticity. And yet . . .

As far back as 1979, W. McCrone of the University of Chicago proved by microscopic analysis of the coloring that it was a manufactured substance called "dragon's blood," or cinnabar. People had learned to make it in the Middle Ages. In addition, the shroud contained a gelatin-based paint (distemper) that wasn't used for painting on fabric until the end of the 13th century.

Unfortunately, McCrone's research received scant attention. Even scientists wanted to believe that the shroud was authentic. Still, an American scientist, E. Jumper of Albuquerque, New Mexico, said he thought it strange that anyone would take the imprint to be anything but a drawing.

But it was all circumstantial evidence up to that point. A direct proof was needed. Why couldn't scientists carry out the proper analysis? This was the 1970s, after all!

Dated by carbon

Well, they did have the necessary tool: carbon dating, which makes it possible to assign a date to anything made by humans. This method was developed by the American scientist Willard Libby, who was awarded the 1960 Nobel Prize in chemistry for it.

Libby's idea, like all sparks of genius, is simple. Our atmosphere is known to be constantly bombarded by cosmic radiation. High-energy particles bumping into the nuclei of atoms in the atmosphere knock out neutrons—nuclear particles with a mass of 1 atomic mass unit (amu) and no electrical charge.

It's common knowledge that the atmosphere contains 80% nitrogen (which occupies the seventh position in the periodic table of the elements). This means that the nitrogen nucleus has 7 protons—nuclear particles with a mass of 1 amu and an electrical charge of +1. In addition, it has 7 neutrons. The scientific shorthand for nitrogen, $^{14}_{7}N$, means that its atomic mass is 14 (7 + 7) amu and its nuclear charge is +7. Carbon is designated as $^{12}_{6}C$ (its mass is 12 amu and its nuclear charge is +6).

A neutron knocked out of a nucleus "attacks" a nitrogen atom. A sort of atomic billiards game begins: a neutron hitting a nitrogen atom knocks out a proton, but the neutron gets stuck. This is shown schematically in the figure at right.

The newly formed nucleus has only 6 protons. We know that an element's position in the periodic table is determined by the nuclear charge of the element—that is, by the number of protons. That's how cosmic radiation turns atmospheric nitrogen into carbon.

But this is not ordinary carbon, with an equal number of protons and neutrons. The nucleus of this "new" carbon $^{14}_{6}C$ has more neutrons than protons and, therefore, the nucleus is unstable and must decay. That's why this type of carbon is called radiocarbon—that is, carbon that decays, emitting high-energy radiation (from the Latin *radius*, which means "ray").

Radiocarbon is known to decay very slowly. It takes about 5,730 years for half of the radioactive carbon 14 to decay. That is, out of 1,000 atoms, 500 will decay in 5,730 years. That doesn't mean, though, that the rest of the atoms will decay in the next 5,730 years. In fact, only half of the remaining atoms will decay. Therefore, in 11,460 years, about 500 + 250 = 750 nuclei will decay. During the next 5,730 years, half of the remaining half (about 125 nuclei) will decay, and the process continues almost endlessly. So the decay of radiocarbon serves as a clock.

As is the case with its stable natural isotope, radiocarbon is oxidized by atmospheric oxygen. The resulting compound—CO_2—is absorbed by plants during photosynthesis (the process by which plants use the energy of photons in sunlight to synthesize organic compounds—in particular, ordinary sugar). While the plant is living, the ratio of radiocarbon atoms to ordinary carbon atoms in the plant is the same as that in the atmosphere. The quantity of radiocarbon absorbed by plants is extremely small—no more than 1 radiocarbon atom per trillion (10^{12}) ordinary carbon atoms. So there's no reason to abstain from apples or other fruits and vegetables out of a fear that they're too radioactive.

Carbon is used by plants as links in the large cellulose molecules they build. Cellulose is an ingredient in paper and wood, cotton, flax, and other natural fibers. Grass is consumed by ruminant animals, from which leather, tusks, horns, hoofs, wool, and hair were obtained for making armor and sword belts, strings for musical instruments, jewelry, and so on.

Once the plant or animal ceases living, there is no longer any exchange of CO_2 with the atmosphere, and the number of radiocarbon atoms decreases steadily due to the radioactive decay process. If we now determine the radiocarbon content of all these old items, we'll get their ages (actually, the times they were made) with some small degree of error. But to date an object we need to know how much $^{14}_{6}C$ was in the original sample—that is, what the atmospheric concentration of radiocarbon atoms was at that time. Unfortunately, in most cases scientists don't have this information. Here objects dated by historical evidence come to the rescue. This is where we get our reference points.

Gradually the number of such reference points has grown. But even today the accuracy of carbon dating is ±30–80 years. When we're trying to establish the age of very old items, this error is certainly acceptable. It often happens, though, that the antiquities are too valuable to be subjected to carbon dating, since a portion of the object is burned up in the process. This is why carbon dating is often applied not to the object itself but to its surroundings—for example, cinders from a fire, soot from a hearth, and so on.

That's exactly how researchers dated the burial grounds of a Cro Magnon settlement in Moravia. It turned out to be 28,000 years old. In

Halgenberg, Austria, archeologists found a soapstone figure of a naked woman, whose beauty inspired an ancient artist some 27,000 years ago. This is the earliest known female image in European art.

Further refinements

The resolution of carbon dating increased drastically when modern computers began to be used in combination with mass-spectrometry, which makes it possible to "count" directly the number of "heavy" $^{14}_{6}C$ atoms. How? By noting the deviation in the trajectories of the dispersed carbon atoms.

Naturally, the lighter the atom, the easier it is to knock it off its "true path." You can demonstrate this easily yourself. You can change the path of snowflakes with a slight wave of your hand. You can't do that with raindrops.

Mass-spectrometry reduced the size of the sample needed by a factor of 1,000. The older technique required 1.5 g of pure carbon for the analysis, whereas mass-spectrometry sometimes needs only 0.5 mg. The required level of radioactivity in the sample also changed drastically. The radioactivity in samples older than 40,000 years doesn't differ from the background radioactivity—that is, the radioactivity caused by cosmic radiation and radioactive contamination.

Actually, the mass-spectrometer doesn't detect radioactivity as such—it determines the number of "heavy" carbon atoms, which greatly simplifies the problem. It's no wonder that, when the scientists presented their ideas at the Vatican, it was agreed that the Shroud of Turin be analyzed.

Flax and fiction

Researchers from Zurich, Oxford, and Tucson (the University of Arizona) took part. Pieces of fabric with precisely assigned dates were used as controls. These items were medieval priestly vestments from the chapel of St. Jean's Basilica in southern France and linen cloth that was used to cover the relics of one of the prophets of Islam.

Preliminary analysis confirmed historical evidence about the age of the vestments and the linen—they dated from the 11th or 12th century. Then the researchers proceeded to the central problem: determining the age of the Shroud of Turin. The results indicated that the flax used for the shroud had been grown sometime between 1262 and 1384. The shroud was more than 1,000 years too "young" to have been used to wrap the body of Christ. Upon learning the results of the analysis, Cardinal Ballestrero, the archbishop of Turin, said, "I see no reason for the Church to question the results or even doubt them." Interestingly enough, after the news was published pilgrims still lined up to see the shroud—they wanted to make sure it hadn't been damaged. And the question remained: who indeed had transferred to cloth the remarkable image that had captivated generations of believers—and perhaps unbelievers?

Art historians think Leonardo da Vinci did it. The level of skill and the style of the "portrait" point to that great master. Remarkable also is the knowledge of human anatomy displayed by the artist. Leonardo is known to have dug up tombs at night in order to dissect dead bodies by torchlight. Da Vinci died in 1519—at the beginning of the same century in which the painting mentioned above, with its rendering of the imprinted shroud, was created. So the Shroud of Turin still holds many secrets.

Future generations of scientists may finally unravel them. How they'll do it is merely a question of time, so it's hard to predict. Just as with the Shroud of Turin, even a few years ago it was hard to believe that it's true age could be established.

On the rocks

You may be thinking, "Okay, so carbon dating is useful when there is some organic carbon around. But what about, say, rock paintings with nothing else nearby? The researcher doesn't have access to remnants from a funeral pyre or anything like that. What then?" In that case, responded chemists and anthropologists at Texas A&M University, you need to use our new low-temperature, low-pressure oxygen plasma method.

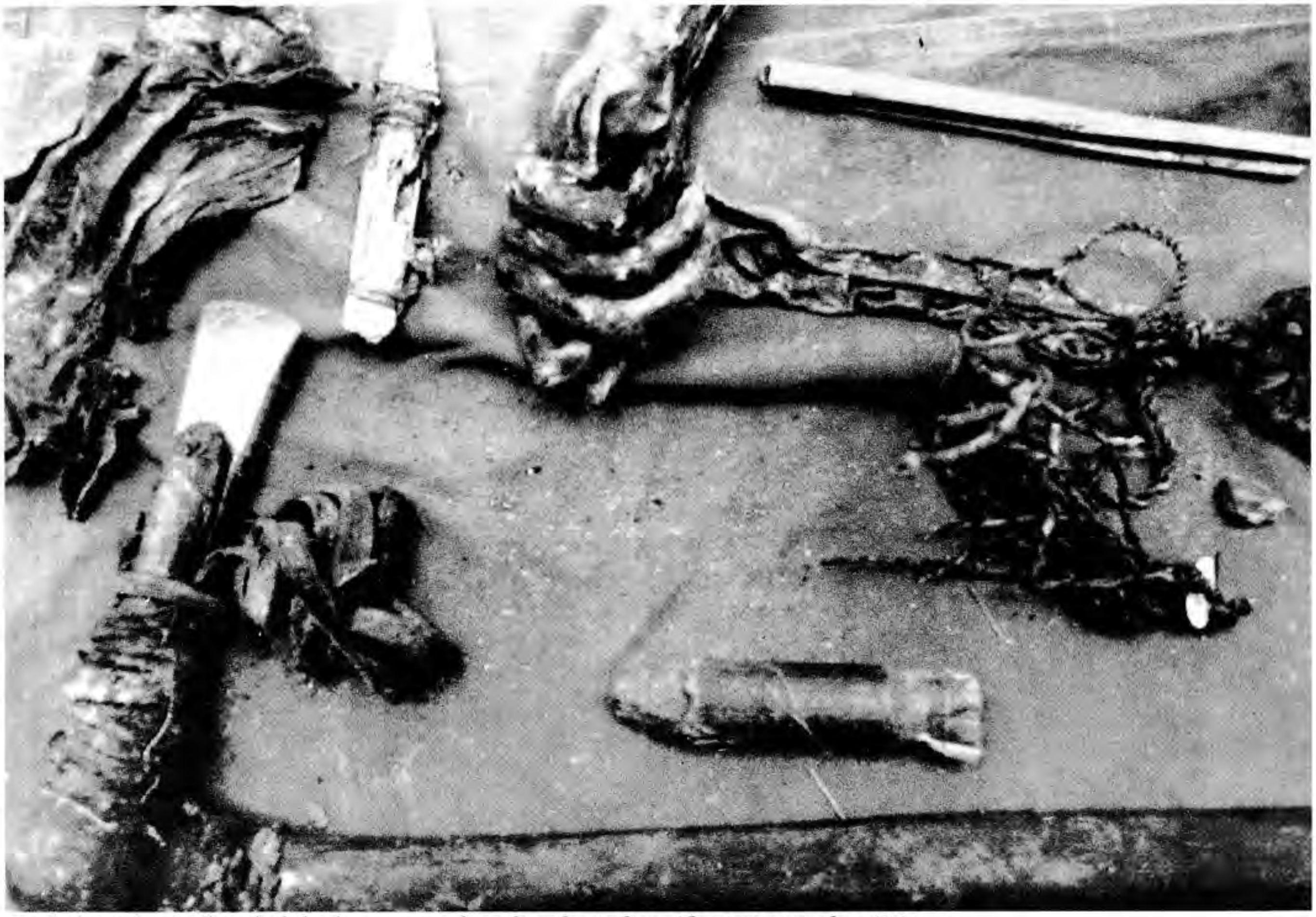
This plasma is obtained by irradiation at radio frequencies. It makes it possible to separate the organic carbon in the paints or "primer" used by ancient artists and the carbon from limestone ($CaCO_3$). At a temperature of 100°C and a pressure of 4 mm Hg, limestone remains intact, so only organic carbon is subjected to the mass-spectroscopic analysis. This was how a rock painting in southwest Texas was dated. Its age—4,000 years (3,865 ± 100)—makes it contemporaneous with the flowering of the ancient Egyptian civilization halfway around the globe.

This new technique isn't restricted to cave paintings. It can be used to study drawings on any ceramic surface—for instance, on ancient pottery. The ceramics themselves are unaffected by the process, but the colors in the painting may change due to catalytic oxidation. This can be repaired, though. For instance, the ferrous magnetite on vases with black figures may turn reddish, like ordinary rust. It may be possible to restore the original color by placing the vase in a hydrogen chamber.

Lost in the Alps

Carbon dating came in handy recently when something quite remarkable was discovered high up in the Alps. The completely preserved corpse of a hunter or herdsman (presumably prehistoric) was found partially protruding from the Similaun glacier in Tirol, just inside Italy. He had a bow, arrows, a copper ax with a long handle, an implement for lighting a fire, and other personal belongings on him or nearby. The body was found at an altitude of 3.2 km. Never before had such a well-preserved human body, untouched by predators or decay, and of such apparent antiquity, been made available for scientific analysis.

The man is thought to have been approximately 20–40 years old; he is



AP/Wide World Photos

Tools lying by the hand of the frozen man found in the Italian Alps in September 1991.

5 feet 2 inches tall. Small crosses and wavy lines are tattooed on his skin. He subsisted on meat and roughly ground food, judging from the amount of wear in his teeth. At first experts at the University of Innsbruck thought he was a medieval messenger, sent on a mission through the mountain pass, who was caught in a sudden snowstorm and perished.

Carbon dating proved, however, that he was a contemporary of the ancient civilizations in the Nile and Indus valleys (around 2000 B.C.). Here was an unexpected visitor from 4,000 years ago! He had leather shoes, a stone necklace, and a flint knife at his waist. These last two items should have immediately pointed to the great antiquity of the Tirolean man.

Forging a head

Carbon dating is a strong shield against forgers who "happen to find" ancient pieces or items of art that, because they're old, acquire great scientific importance and great pecuniary value.

There is nothing new under the sun, in this respect. A collection of famous frauds was recently exhibited at the British Museum. Egyptologists, at least, are well acquainted with counterfeit seals purportedly used by the pharaohs. The pharaohs themselves, during the Middle and New kingdoms, ordered that sculptures from the Old Kingdom be copied. Visitors to the exhibit saw a 4,000-year-old Babylonian cuneiform tablet—a copy of one that was 1,000 years older. To make it look more "authentic," the imitation bore an old Sumerian inscription (a standard warning whose wording never varied): "If you forge this document, Enki will plug all your ducts with slime." This was serious stuff—Enki was the supreme deity in the Sumerian pantheon.

The 19th century was rich in frauds, even as intense interest in things historical grew widespread. Chinese artisans organized a kind of "production line" of sculptures sold as antiquities from the Shang dynasty (1766–1112 B.C.) to wealthy European collectors. Nowadays these skillful frauds have an artistic value of their own. Fake Michelangelos and imitations of other Renaissance artists were particularly popular in the last century.

A splendid "Man with the Golden Helmet," which until recently was considered a genuine Rembrandt, was also painted at that time. It's very hard to determine whether a work by Rembrandt is an original. Rembrandt, who lived from 1606 to 1669, was extremely productive. In 44 years he produced hundreds of paintings, about 1,500 drawings, and 300 engravings. About a hundred students worked side by side with the master. They used to write "Rembrandt" on their works to make the price go up (Rembrandt let them do it). For instance, Samuel van Hoogstraten put "Rembrandt" on his "Young Woman by a Half-Open Door," and Jan Lievens did the same with his "Easter Feast" (1625). It's no wonder that, by the turn of the century, there were literally thousands of "Rembrandts" floating around—some well painted, some not.

Art historians made the first inventory of Rembrandt's work in the 1920s, and only 700 made the cut. By 1963—300 years after his death—the number of paintings by Rembrandt dropped to 420. In 1967 the Rembrandt Research Project (RRP) started up. Experts think that when the RRP is through with its work, the total number of paintings credited to Rembrandt may drop to 300.

As a result of such activities, many art museums have "lost" their Rembrandts—for instance, the Metropolitan Museum of Art in New York ("Portrait of a Man" and "Portrait of a Woman"), the Berlin State Museum ("Man with the Golden Helmet"), and the Hermitage in St. Petersburg, Russia ("David and Jonathan" and "Parable of Vine").

When a museum removes the "Rembrandt" label from a painting, its auction price goes down drastically (if not its artistic value). The two portraits at the Met might previously have sold for at least $3 million each. If they were done by Jan Lievens, Rembrandt's friend, they might bring in $600,000 to $800,000 for the pair. If they were done by a student, Govert Flinck, the pair would sell for $40,000.

While in London, we visited an exhibition at the National Gallery dedicated to the study of Rembrandt's technique and artistry. Besides original Rembrandts, there were a lot of X-ray, ultraviolet, and infrared photographs of his works. We also saw photomicrographs of the layers of paint that, one after another, went into the creation of "Saskia van Ulyenburgh," the unforgettably beautiful portrait of Rembrandt's beloved wife.

An exhibition of original Rembrandts was also held at the Altes Museum in Berlin. Visitors had the chance to compare Rembrandt with his students and followers and also to rack their brains over "Polish Dragoon" (1657). Dutch scholars aren't sure if it's a Rembrandt.

Enlisting neutrons . . .

Today art historians and museum workers have at their disposal a wide variety of techniques for analyzing every detail of paintings and other works of art. Infrared light exposes what is concealed beneath the surface layer of paint. X-ray photography is a great help in learning the secrets of the old masters. For instance, Titian's "Judith" conceals a portrait of a man (he may be King Carlos V of Spain). Another painting shows a woman with four hands: she was first painted with her arms folded, then with her arms apart.

Scholars also have recourse to neutron activation analysis, in which an object is bombarded with fast neutrons that excite the nuclei of metals.[1] The resulting energy is then dispersed, creating a specific spectrum that distinguishes between, say, lead and titanium. This is the basic difference between different types of gesso (the white coating applied to a bare canvas). Painters in previous centuries used lead gessoes, while titanium gesso came into use in the 1920s. This is one way of telling a modern copy from an authentic work painted in the 19th century.

No less famous in this regard is the color Prussian blue, which came into use at the end of the 18th century. At a London exhibition of frauds, a picture supposedly painted by Botticelli (1444?–1510) was on display. The Madonna in the painting looks like a star from the silent movies. But what really gives the forger's game away is the Prussian blue used in painting the Virgin's bright-blue garments. This "Botticelli" was produced after World War I.

. . . and protons

A new technique was developed recently. It's called proton-induced X-ray emission (PIXE). The surface of a document is irradiated by a narrow beam (0.5–1 mm in diameter) of protons with energies reaching 4.5 MeV. The beam excites the atoms of metals contained in dyes, pigments, and inks. The priceless items aren't threatened by the process, since the energy of the beam is no more than that produced by a 100-watt bulb illuminated for a few moments at a distance of 0.5 m. The excited atoms of the metals emit X rays that are recorded by detectors. Signals produced by the detectors are passed on to a computer for further analysis.

PIXE analysis recently helped scholars from California "read" a Gutenberg Bible. This particular volume, kept at Harvard University, has 42 lines per page (Gutenberg would vary the layout). The researchers learned that Gutenberg used large quantities of lead and copper in making his ink. He mixed it fresh every day, and this allowed the scholars to determine the order in which he printed the pages: after printing the first sheet on both sides, consisting of pages 1 and 119, he printed the next (pages 2 and 120). Sheets were then sewn into "signatures," and six signatures were printed simultaneously. The staff at the Louvre in Paris would like to analyze a 36-line Gutenberg Bible that was printed before the 42-line version.

The list goes on. Almost every day the need arises to verify—or debunk—a claim. In England, a huge table with elaborate decorations was thought to have accommodated the venerable Knights of the Round Table in King Arthur's time. Unfortunately, upon analysis the table turned out to be 500 years younger than the knights themselves.

[1]See also "Neutrons Seek the Murderer!" in the May/June 1992 issue.—*Ed.*

Just for the fun of it!

B71

Time machine. "You know," a friend of mine once said, "the day before yesterday I was 10 years old, and next year I'll be 13!" Can this be true? (S. Korshunov)

B72

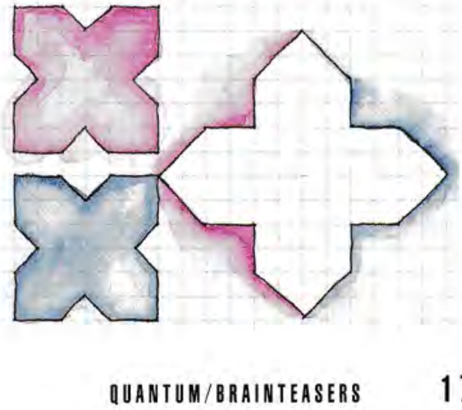
Mysterious pictographs. What's the rule for drawing this sequence of figures? Guess what figure should be drawn next. (A. Zvonkin)

B73

Escaping pools. In karstic regions—that is, areas with irregular limestone formations, including caverns and underground streams—one can come upon unusual ponds. During the rainy season they gradually fill with water, but then they suddenly become surprisingly shallow. Why does that happen? (A. Buzdin)

B74

Calculating on the way. Alice used to walk to school every morning, and it took 20 minutes for her from door to door. Once on her way she remembered she was going to show the latest issue of *Quantum* to her classmates but had forgotten it at home. She knew that if she continued walking to school at the same speed, she'd be there 8 minutes before the bell, and if she went back home for the magazine she'd arrive at school 10 minutes late. What fraction of the way to school had she walked at that moment in time? (S. Dvorianinov)



B75

Doublecross. Cut two crosses in the same way, each into four pieces, so that all eight pieces can be put together to form a similar cross with twice the area. (L. Mochalov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 58

Art by Pavel Chernusky

A magnificent obsession

The strange story of perfect (and perfectly useless) numbers

by Michael H. Brill and Michael Stueben

THERE IS A CURIOSITY IN THE HISTORY OF number theory that intermixes mathematics and religion. It traces back to the year 1644, to the French cleric and mathematician Merin Mersenne. Father Mersenne acted as an important clearinghouse of mathematical and philosophical inquiry by carrying on continuous correspondence (published in eight volumes) with the great minds of his time: Descartes, Pascal, Galileo, Huygens . . . But the only piece of mathematics for which he is remembered is a set of numbers that can be written in the form $2^p - 1$, where p is a prime number. These are *Mersenne numbers*. The name was chosen by the Cambridge scholar W. W. Rouse Ball because of a reference to them in the preface to Mersenne's *Cogita physico-mathematica* (Paris, 1644). The first ten Mersenne numbers are formed from the first ten prime numbers:

1. $M_2 = 2^2 - 1 = 3$ (prime)
2. $M_3 = 2^3 - 1 = 7$ (prime)
3. $M_5 = 2^5 - 1 = 31$ (prime)
4. $M_7 = 2^7 - 1 = 127$ (prime)
5. $M_{11} = 2^{11} - 1 = 2{,}047$ (23×89)
6. $M_{13} = 2^{13} - 1 = 8{,}191$ (prime)
7. $M_{17} = 2^{17} - 1 = 131{,}071$ (prime)
8. $M_{19} = 2^{19} - 1 = 524{,}287$ (prime)
9. $M_{23} = 2^{23} - 1 = 8{,}388{,}607$ $(47 \times 178{,}481)$
10. $M_{29} = 2^{29} - 1 = 536{,}870{,}911$ $(233 \times 1{,}103 \times 2{,}089)$

These numbers are interesting to mathematicians because some of them can be used to generate the so-called perfect numbers. And here begins a story that will lead us back to the good Father.

Curious properties

A *perfect number* is a positive integer that is equal to the sum of its positive proper divisors—for example, $6 = 1 + 2 + 3$. There are many even perfect numbers, but it is not known whether odd perfect numbers exist. The central theorem was half discovered by Euclid and half discovered by Leonhard Euler (1707–1783), so Mersenne got to see only half of the theorem.

Around 300 B.C. Euclid proved that if $2^p - 1$ is prime, then $2^{p-1}(2^p - 1)$ is an even perfect number. His exact words (translated from the Greek) are: "If as many numbers as we please beginning from a unit be set out continuously in double proportion, until the sum of all becomes prime, and if the sum multiplied into the last make some number, the product will be perfect."[1] Remember, Euclid had no algebra; he described everything in terms of geometry and arithmetic. His words will make sense when we consider the following two sequences:

(n =	2,	3,	4,	5,	6,	7,	8,	...,	n,	...)
Sequence 1:	1,	2,	4,	8,	16,	32,	64,	...,	2^{n-1},	...
Sequence 2:	1,	**3**,	**7**,	15	**31**,	63,	**127**,	...,	$2^n - 1$,	...

The upper sequence is a sequence of powers of two. Each element in the lower sequence is the sum of the upper sequence up to the number immediately above itself. Euclid said that if the lower number is prime (for example, 3, 7, 31, 127, ...), then the product of the upper and lower numbers is a perfect number ($2 \times 3 = 6$, $4 \times 7 = 28$, $16 \times 31 = 496$, and $64 \times 127 = 8{,}128$).

About two thousand years later the Swiss number wizard Leonhard Euler proved the converse: Any even perfect number must be of the form described by Euclid—$2^{p-1}(2^p - 1)$, where $2^p - 1$ is prime. Notice that if $2^p - 1$ is a prime, then p is a prime, but the converse is not true. So for every Mersenne prime $M_p = 2^p - 1$ (where p, of course, is a prime number), we can generate an even perfect number, and vice versa. (You should check the proof in the box at right.)

Problem 1. (a) Prove that the expression $2^p - 1$ must be composite if p is even.

[1] Euclid, *Elements*, Book 9, proposition 36 (Dover Publications, 1956, p. 421).

Art by Chrystine Simmons

THE EUCLID–EULER THEOREM. *The number $2^{p-1}(2^p - 1)$ is an even perfect number if and only if $2^p - 1$ is prime.*

To prove this theorem it helps to define $\sigma(p)$ as the sum of the divisors of p including p itself—for example, $p(6) = 12$, $p(7) = 8$—and to note that $\sigma(p) \times \sigma(q) = \sigma(pq)$ if p and q are relatively prime. (A proof of this can be found in any book on elementary number theory.) Then p is perfect if and only if $\sigma(p) = 2p$.

PROOF. *Part I (Euclid).* Consider the number $r = 2^{p-1}(2^p - 1)$, where $2^p - 1$ is prime. If $2^p - 1$ is prime, then it has only two divisors, itself and 1; therefore, $\sigma(2^p - 1) = 1 + 2^p - 1 = 2^p$. Since the only divisors of a power of 2 are itself and the smaller powers of 2, we have $\sigma(2^{p-1}) = 1 + 2 + 4 + 8 + \ldots + 2^{p-1} = 2^p - 1$. Now

$$\sigma[2^{p-1}(2^p - 1)] = \sigma(2^{p-1})\sigma(2^p - 1) = (2^p - 1)2^p = 2r.$$

Therefore, r is a perfect number if $2^p - 1$ is prime.

Part II (Euler). Conversely, suppose r is an even perfect number. Write r as $(2^{p-1})q$, where $p > 1$ and q is odd. Since q and 2^{p-1} are relatively prime, we have

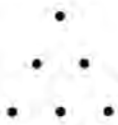
$$2^p q = 2r = \sigma(2^{p-1}q) = (2^p - 1)\sigma(q).$$

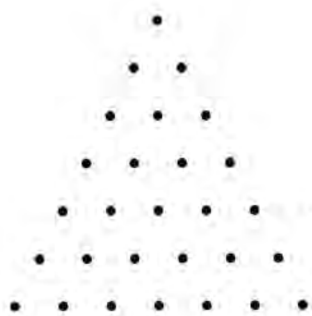
Therefore, $\sigma(q) = 2^p q / 2^{p-1} - q + q/(2^p - 1)$. This implies that $q/(2^p - 1)$ is an integer, and q and $q/(2^p - 1)$ are the only divisors of q. In other words, $q = 2^p - 1$, and q is prime. Thus, r is a perfect number only if $2^p - 1$ is prime.

(b) Prove that the expression $2^p - 1$ must be composite if p is odd and composite.

The only perfect numbers below 33,000,000 are 6, 28, 496, and 8,128. It's easy to write a computer program to search for these numbers. We need only generate prime numbers p and check to see if $2^p - 1$ is prime. But a number is prime if it has no nonunit divisors less than its square root. So, for example, if $p = 31$, we need to examine only the numbers from 2 to $\sqrt{2^{31}} - 1 = 46{,}341$ to see if any of them divides $2^{31} - 1 = 2{,}147{,}483{,}647$. A computer can check these numbers in seconds and show that none of them is a divisor. Therefore, $2^{30}(2^{31} - 1)$ is a perfect number.

Perfect numbers have some curious properties. For example, all even perfect numbers end with the numbers 6 or 28. All even perfect numbers are triangular—that is, if you had an even perfect number of marbles, you could arrange them in the form of an equilateral triangle:


= 6


= 28

Problem 2. Prove that even perfect numbers are triangular. (Even perfect numbers are of the form $2^{p-1}(2^p - 1)$; triangular numbers are of the form $n(n+1)/2$.)

The sum of the reciprocals of *all* the divisors of any perfect number equals 2:

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{6}=2$$

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{4}+\frac{1}{7}+\frac{1}{14}+\frac{1}{28}=2$$

This follows from the definition of a perfect number. Let the perfect number be n, and let its divisors be d_i (including n itself). By the definition of a perfect number n, the sum of its divisors is $2n$ (n itself is a proper divisor).

Consider $n\sum_i \frac{1}{d_i} = \sum_i \frac{n}{d_i}$, where each term in this sum is a d_i for some i. The sum is none other than $\sum_i d_i$ again, which is just $2n$. Therefore, $\sum_i \frac{1}{d_i} = 2$, and the sum of the reciprocals of *all* the divisors of any perfect number equals 2.

All even perfect numbers are equal to the sums of successive powers of 2:

$$\begin{aligned}
6 &= 2^1 + 2^2\\
28 &= 2^2 + 2^3 + 2^4\\
496 &= 2^4 + 2^5 + 2^6 + 2^7 + 2^8\\
8{,}128 &= 2^6 + 2^7 + 2^8 + 2^9 + 2^{10} + 2^{11} + 2^{12}\\
33{,}550{,}336 &= 2^{12} + 2^{13} + 2^{14} + \ldots + 2^{22} + 2^{23} + 2^{24}\\
8{,}589{,}869{,}056 &= 2^{16} + 2^{17} + 2^{18} + \ldots + 2^{30} + 2^{31} + 2^{32}\\
137{,}438{,}691{,}328 &= 2^{18} + 2^{19} + 2^{20} + \ldots + 2^{34} + 2^{35} + 2^{36}
\end{aligned}$$

Problem 3. Prove that even perfect numbers are sums of successive powers of 2. Hint: consider the well-known geometric progression of n terms

$$a + ar + ar^2 + \mathrm{L} + ar^{n-1} = \frac{a(r^n - 1)}{(r-1)}.$$

Every even perfect number $2^p(2^{p+1} - 1)$ greater than 6 is the sum of the cubes of the first $2^{p/2}$ consecutive odd numbers:

$$\begin{aligned}
28 &= 1^3 + 3^3\\
496 &= 1^3 + 3^3 + 5^3 + 7^3\\
8{,}128 &= 1^3 + 3^3 + 5^3 + \ldots + 11^3 + 13^3 + 15^3\\
33{,}550{,}336 &= 1^3 + 3^3 + 5^3 + \ldots + 123^3 + 125^3 + 127^3\\
8{,}589{,}869{,}056 &= 1^3 + 3^3 + 5^3 + \ldots + 507^3 + 509^3 + 511^3
\end{aligned}$$

Add the digits of a perfect number greater than 6. If the sum is not a single digit, add the digits of the sum. Continue until you have a single-digit answer. This is called the *digital root* of a number, and the digital root of any perfect number is 1:

$28 \to 2 + 8 = 10 \to 1 + 0 = 1$

$496 \to 4 + 9 + 6 = 19 \to 1 + 9 = 10 \to 1 + 0 = 1$

$8{,}128 \to 8 + 1 + 2 + 8 = 19 \to 1 + 9 = 10 \to 1 + 0 = 1$

$33{,}550{,}336 \to 3 + 3 + 5 + 5 + 0 + 3 + 3 + 6 = 28 \to 2 + 8 = 10 \to 1 + 0 = 1$

These properties were discovered over a period of years in an attempt to find a pattern of perfect numbers. Unfortunately, none of these properties has led to a generating formula. Perfect numbers have even been analyzed in other bases. In base two, each even perfect number is n ones followed by $n - 1$ zeros:

Base 2:	4,096	2,048	1,024	512	256	128	64	32	16	8	4	2	1
6 =											1	1	0
28 =									1	1	1	0	0
496 =					1	1	1	1	1	0	0	0	0
8,128 =	1	1	1	1	1	1	1	0	0	0	0	0	0

Problem 4. Verify that in base two each even perfect number is n ones followed by $n - 1$ zeros.

This is an interesting property, but like the other properties, it's of no use in finding more perfect numbers. There are 32 known Mersenne primes and, consequently, 32 known even perfect numbers. Table 1 is a list of the known *even* perfect numbers.

The thirteenth even perfect number is $2^{520}(2^{521}-1)$ and has 314 digits. Care to calculate it? Part of the information for this table comes from the second edition of a 479-page book called *The Book of Prime Number Records.* Books on number theory aren't usually best sellers, but the first edition of this one quickly sold out. The other source is the *Journal of Recreational Mathematics* (August 1961). The study of perfect numbers can produce methods that are very useful in applied mathematics, but the properties of perfect numbers are purely recreational.

Odd perfect numbers

All perfect numbers discovered so far are even. No one knows if there is such a thing as an odd perfect number. But if one does exist it must be a whopper, because it has been proven that there are no odd perfect numbers with fewer than 150 digits. Mathematicians have determined a large number of properties of an odd perfect number and so far have found no contradictions. For example:

1. An odd perfect number must have a remainder of 1 when divided by 12 or 4, and a remainder of 9 when divided by 36.
2. It must have at least 150 digits.
3. It must be divisible by at least eight distinct primes.
4. It is not divisible by 105—that is, 3, 5, and 7 cannot all be factors.
5. It is divisible by a prime power greater than 10^{12}.
6. If it has exactly n different prime divisors, the smallest of them must be smaller than $n+1$.

Since there may be no such thing as an odd perfect number, it's amazing that we have so many theorems about them.

Problem 5. If an odd perfect number exists and has exactly 28 prime divisors, what is the largest value that the smallest prime divisor could take?

A history of perfect numbers

The history of perfect numbers is really the history of Mersenne primes. Indeed, the largest known prime number at any time is usually a Mersenne prime. Table 2 summarizes the search for perfect primes through the centuries. The first four perfect numbers were known to the ancients. The earliest discoverer of prime numbers that we know about is the German mathematician Regiomontanus, who found the fifth perfect number and possibly the sixth in 1456. The sixth and seventh perfect numbers were determined by the Italian Cataldi (1552–1626) by the direct procedure of dividing all primes less

Table 1

	p	digits	n	$= 2^{p-1}(2^p-1)$
1.	2	1	6	$=2^1(2^2-1)$
2.	3	2	28	$=2^2(2^3-1)$
3.	5	3	496	$=2^4(2^5-1)$
4.	7	4	8128	$=2^6(2^7-1)$
5.	13	8	33550336	$=2^{12}(2^{13}-1)$
6.	17	10	8589869056	$=2^{16}(2^{17}-1)$
7.	19	12	137438691328	$=2^{18}(2^{19}-1)$
8.	31	19	2305843008139952128	$=2^{30}(2^{31}-1)$
9.	61	37	2658455991569831744654692615953842176	$=2^{60}(2^{61}-1)$
10.	89	54	191561942608236107294793378084303638130997321548169216	$=2^{88}(2^{89}-1)$
11.	107	65	13164036458569648337239753460458722910223472318386943117783728128	$=2^{106}(2^{107}-1)$
12.	127	77	14474011154664524427946373126085988481573677491474835889066354349131199152128	$=2^{126}(2^{127}-1)$
			The search begins here (beyond 10^{150}) for an odd perfect number.	
13.	521	314		$=2^{520}(2^{521}-1)$
14.	607	366		$=2^{606}(2^{607}-1)$
15.	1279	770		$=2^{1278}(2^{1279}-1)$
16.	2203	1327		$=2^{2202}(2^{2203}-1)$
17.	2281	1373		
18.	3217	1937		$\vdots$
19.	4253	2561		
20.	4423	2663		
21.	9689	5834		
22.	9941	5985		
23.	11213	6751		
24.	19937	12003		
25.	21701	13066		
26.	23209	13973		
27.	44497	26790		
28.	86243	51924		
29.	110503	66530		
30.	132049	79562		
31.	216091	130001		
32.	756839	455663		$=2^{756838}(2^{756839}-1)$

than the square root of each candidate number. Since there are 128 primes less than the square root of M_{19}, Cataldi needed 128 divisions to determine if it was prime. The eighth perfect number was verified by Euler: $2^{30}(2^{31} - 1) = 2{,}305{,}843{,}008{,}139{,}952{,}128$. In 1811 Peter Barlow wrote in his book *Number Theory* that this is "the greatest thing that will ever be discovered, for as they are merely curious without being useful it is not likely that any person will attempt to find one beyond it." The twelfth perfect number, $2^{126}(2^{127} - 1)$, has 77 digits and is the largest perfect number discovered without a computer. It was found in 1876. The next three numbers in order of discovery are actually smaller. The discovery in 1883 of the ninth perfect number took 54 hours of paper-and-pencil computation. In 1911 and 1913 the tenth and eleventh perfect numbers were discovered. The world had to wait until 1952 for a computer to discover the thirteenth perfect number: $2^{520}(2^{521} - 1)$.

In 1952 the National Bureau of Standards' Western Automatic Computer (SWAC) was turned loose on the problem. R. M. Robinson discovered five perfect numbers that year (the thirteenth through seventeenth). As Albert H. Beiler describes it,

> Dr. D. H. Lehmer, who had spent very many hours on Mersenne's numbers, saw the machine do in 48 seconds what it had taken him over 700 hours of arduous labor with a desk calculator to demonstrate 20 years before: that $2^{257} - 1$ is composite. Mersenne had stated that all eternity would not suffice to tell if a 15- or 20-digit number is prime. Within a few hours, SWAC had tested 42 numbers, the *smallest* of which had 80 digits. It took 13½ minutes to determine that $2^{1279} - 1$ is prime.[2]

In 1978 two high school students, Curt Noll and Laura Nickel, discovered the twenty-fifth perfect number. A year later Noll discovered the twenty-sixth. In 1979 David Slowinski and Harry Nelson discovered the next perfect number. In 1983 and again in 1985 Slowinski found two more perfect numbers, but in 1988 two researchers found one that was smaller than Slowinski's discoveries. In early 1992 Slowinski found the thirty-second known Mersenne prime and is now tied with R. M. Robinson as the champion discoverer of perfect numbers.

Table 2

	M_p	Discoverer	Year
1.	M_2	?	?
2.	M_3	?	?
3.	M_5	?	?
4.	M_7	?	?
5.	M_{13}	Regiomontanus (Johann Müller)	1456
6.	M_{17}	P. A. Cataldi	1588
7.	M_{19}	P. A. Cataldi	1588
8.	M_{31}	L. Euler	1772
9.	M_{61}	I. M. Pervushin	1883
10.	M_{89}	R. E. Powers	1911
11.	M_{107}	E. Fauquemberque	1913
12.	M_{127}	E. Lucas	1876
13.	M_{521}	R. M. Robinson	1952
14.	M_{607}	R. M. Robinson	1952
15.	M_{1279}	R. M. Robinson	1952
16.	M_{2203}	R. M. Robinson	1952
17.	M_{2281}	R. M. Robinson	1952
18.	M_{3217}	H. Riesel	1957
19.	M_{4253}	A. Hurwitz	1961
20.	M_{4423}	A. Hurwitz	1961
21.	M_{9689}	D. B. Gilles	1963
22.	M_{9941}	D. B. Gilles	1963
23.	M_{11213}	D. B. Gilles	1963
24.	M_{19937}	B. Tuckerman	1971
25.	M_{21701}	Curt Noll and Laura Nickel	1978
26.	M_{23209}	Curt Noll	1979
27.	M_{44497}	David Slowinski and Harry Nelson	1979
28.	M_{86243}	David Slowinski	1982
29.	M_{110503}	W. N. Colquitt and L. Welsch Jr.	1988
30.	M_{132049}	David Slowinski	1983
31.	M_{216091}	David Slowinski	1985
32.	M_{756839}	David Slowinski	1992

Searching for perfect numbers

In 1876 the Frenchman E. V. Lucas discovered a test to determine if numbers of the form $2^p - 1$ are prime. In 1930 the American D. H. Lehmer published an improvement. The method is now known as the Lucas–Lehmer test for primality. To determine if $2^p - 1$ is prime, set $S_0 = 4$ and generate a sequence of $p - 2$ numbers:

$$S_{n+1} \equiv (S_n^2 - 2) \bmod (2^p - 1).$$

Then $2^p - 1$ is prime if and only if $S_{p-2} = 0$.

The only problem is that S_n^2 will quickly approach $(2^p - 1)^2$, which is huge. To verify that M_{756839} is a prime, some of the squares contain nearly a half-million digits!

Problem 6. In the Lucas–Lehmer test, how big will some of the numbers become in testing M_{756839}? Hint: since $(2^p - 2)^2$ is nearly equal to 2^{2p}, the question amounts to asking how many digits are in $2^{2(756839)}$.

Number theorists have invented many tricks to speed up computer searches. For example, it has been shown that all prime divisors of $2^p - 1$ must be of the form $8n \pm 1$. Another trick is to work in large bases. Because the CRAY-1 computer returns an exact 48-bit integer product when the operands are 24-bit integers, David Slowinski decided to work in base 2^{24} (that is, base 16,777,216) when searching for the twenty-seventh Mersenne prime. This reduced the size of his numbers from more than 26,000 digits to less than 2,000 digits.

Problem 7. (a) The twenty-seventh Mersenne prime $M_{44497} = 2^{44497} - 1$ contains 26,790 digits when expressed in base 10. How many digits does it require when expressed in base 2^{24}?

(b) Express $M_{44497} = 2^{44497} - 1$ in base 2^{24}. Hint: work by analogy.

Notice that $10000 \equiv 1 \bmod (10^4 - 1)$, so $12345678 =$

[2]Albert H. Beiler, *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains* (Dover, 1964), p. 18.

$12340000 + 5678 = 1234 \times 10^4 + 5678 \equiv 1234 + 5678 \equiv 6912 \pmod{10^4 - 1}$. This is no coincidence: again, $001101_2 \bmod (2^3 - 1) = 001_2 + 101_2 = 110_2 = 6_{10}$. And in general if n (not equal to $b^p - 1$) is represented with $2p$ digits in base b, then it's not hard to see that

$$n \bmod (b^p - 1) = (\text{first } p \text{ digits}) + (\text{second } p \text{ digits})$$

(the case $b = 10$, $p = 1$ is familiar from elementary arithmetic). In other words, no division is necessary to calculate the members of the Lucas–Lehmer sequence. It is the combination of software tricks and fast computers that make big-game hunting successful in the land of giant numbers. David Slowinski stated in 1979 that

> the CRAY-1 has a tremendous speed advantage over conventional computers. As an indication of the speed of the CRAY-1, the Lucas–Lehmer test for $p = 8191$ took 100 hours on the Illiac-I (D. Wheeler, 1959), 5.2 hours on an IBM 7090 [1962], 49 minutes on the Illiac-II [1963], 3.17 minutes on the IBM 360/91 [1971], and ten seconds on the CRAY-1 [1979]. The author's program independently discovered the 26th Mersenne prime, on February 23, 1979, two weeks (alas) after Noll (sans Nickel). The check for M_{23209}, which had taken Noll eight hours forty minutes on a CYBER-174, used less than seven minutes on the CRAY-1.[3]

Mersenne's conjecture

It isn't known if there are infinitely many Mersenne prime numbers. In 1644 Mersenne stated (without proof) that the only p's for which Mersenne numbers are prime are $p = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127$, and 257. In 1883 it was proven that $2^{61} - 1$ is a prime and the Mersenne list was shown to be incomplete; but some speculated (W. W. Rouse Ball among them) that a copyist had misread Mersenne's 61 as a 67. Then in 1903 an American mathematician by the name of Frank Nelson Cole settled the issue with a paper presented to a gathering of his peers. The mathematical and science fiction writer E. T. Bell was there and felt obliged to preserve the scene for posterity:

> When the chairman called on him for his paper, Cole—who was always a man of very few words—walked to the board and, saying nothing, proceeded to chalk up the arithmetic for raising 2 to the 67th power. Then he carefully subtracted 1. Without a word, he moved over to a clear space on the board and multiplied out, by longhand,
>
> $$193{,}707{,}721 \times 761{,}838{,}257{,}287.$$
>
> The two calculations agreed. For the first time on record, an audience of the American Mathematical Society vigorously applauded the author of a paper delivered before it. Cole took his seat without having uttered a word.
>
> When I asked Cole in 1911 how long it had taken him to crack M_{67}, he said "three years of Sundays."[4]

[3]David Slowinski, "Searching for the 27th Mersenne Prime," *Journal of Recreational Mathematics*, vol. 11, no. 4.

[4]E. T. Bell, *Mathematics: Queen and Servant of Science* (McGraw-Hill, 1951), p. 228.

Just who was this taciturn number-cruncher? Frank Nelson Cole graduated from Harvard, second in a class of 189. He studied in Germany and returned to teach at Harvard. Under Cole's effective leadership, the Harvard graduate school of mathematics became a world-famous center of study. Cole was honored by the American Mathematical Society with the establishment of an F. N. Cole prize in mathematics, which is still being awarded. A strange remark can be found under his name in the *Dictionary of Scientific Biography*: "He had married in 1888, but he had largely isolated himself from his family [a wife and three sons] since 1908. At the time of his death [in 1926], Cole lived in a rooming house under the name of Edward Mitchell and claimed to be a bookkeeper."

To return to the subject at hand: Mersenne's original list overlooked some primes (for $p = 61, 89, 107$) and included false primes (for $p = 67, 257$). His conjecture was proven wrong, but in fact his name is associated with a surprising curiosity. In the list of Mersenne primes, the twentieth is $2^{4423} - 1$. If the exponent is written in base 26 (because there are 26 letters in the alphabet), then starting with A = 0, B = 1, C = 2, and continuing to Z = 25, we can substitute letters for the coefficients:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

If the ghost of Father Mersenne is floating through the ether, he must occasionally smile at the connection of one of his numbers with the Word:

$$4{,}423 = (6 \times 26^2) + (14 \times 26^1) + (3 \times 26^0).$$

ANSWERS ON PAGE 59

Does your library have Quantum?

If not, talk to your librarian!

Quantum is a resource that belongs in every high school and college library.

"A first-class 'new' magazine. . . one can appreciate the meaning of quality and imaginative challenge. . .it is for anyone with an interest in science, particularly math and physics. Highly recommended."—***Library Journal***

"It should be in every high school library [and] in most public libraries. . .we owe it to our students to make *Quantum* widely available."—**Richard Askey, Professor of Mathematics at the University of Wisconsin, Madison**

Share the QUANTUM experience!

PTRA-PLUS

OPPORTUNITY FOR PHYSICS AND PHYSICAL SCIENCE TEACHERS
TO BE TRAINED BY
THE
AMERICAN ASSOCIATION OF PHYSICS TEACHERS
TO BECOME

PHYSICS TEACHING RESOURCE AGENTS

The American Association of Physics Teachers (AAPT) expects renewal of a grant from the National Science Foundation to continue the Physics Teaching Resource Agent (PTRA) program. Participants who are selected to attend will be given intensive training in basic physics content, equity awareness, issues in physics teaching, and leadership.

The summer 1993 program will be held from July 31 to August 12 at Boise State University in Idaho, and will include about six days of workshop training as well as participation in a national meeting of AAPT. Completed applications must be received at AAPT by Friday, March 5, 1993.

Teachers selected for the program will receive:

1] Training in physics content, physics teaching strategies, equity awareness and leadership by outstanding teachers,
2] Opportunities to interact with outstanding fellow teachers who are presently PTRAs,
3] Travel support for two summer PTRA-PLUS training sessions one in 1993, and the second in 1994,
4] Per diem expenses and $60.00 per day stipend for six training days each summer,
5] Financial support for conducting outreach workshops based on training, and
6] Certification by AAPT as a PTRA.

For additional information and an application, call Maria Elena Khoury at AAPT (301) 345 - 4200 or write to:

PTRA-PLUS Program
c/o American Association of Physics Teachers
5112 Berwyn Road
College Park, MD 20740 - 4100

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M71

Forget Cardano. Solve the cubic equation

$$x^3 + x^2 + x = -\frac{1}{3}.$$

(Y. Ionin)

M72

Equal sums of digits. A number b is obtained from a number a by a permutation of digits. Prove that the following pairs of numbers have equal sums of digits: (a) $2a$ and $2b$; (b) $a/2$ and $b/2$ (if a and b are even); (c) $5a$ and $5b$. (A. Lisitsky)

M73

Right at the incenter. In a right triangle one half of the hypotenuse (from a vertex to the midpoint of the hypotenuse) subtends a right angle at the triangle's incenter. Find the ratio of side lengths of the triangle. (B. Pitskel)

M74

A fifth wheel. An equilateral triangle is covered by five smaller equilateral triangles congruent to each other. Prove that it's possible to shift smaller triangles so as to cover the big one with four of them. (V. Proizvolov)

M75

Democracy and math (cont'd). The country of Anchuria is divided into 999 electoral districts with equal numbers of voters; one member of parliament is elected in each district. There are three parties A, B, and C that nominate candidates. Party A is supported by 15% of all voters, parties B and C by 30% and 55%, respectively. If none of three candidates gets more than 50% of the vote, a runoff election is held for the two candidates that received more votes than the third one. (The law for the case when candidates get the same number of votes is irrelevant for this problem.) In the runoff, parties A and B support each other, and C supports A. What is the greatest and smallest numbers of each party that can be elected to the parliament?

Physics

P71

Flight of the bumblebee. A bumblebee can fly vertically upward with a maximum speed v_1 and downward with a speed v_2. Assuming that the bumblebee's "thrust" F does not depend on the direction of flight and the air resistance is proportional to the bumblebee's speed, determine the bumblebee's maximum speed when it flies at an angle α with the horizontal. (B. Korsunsky)

P72

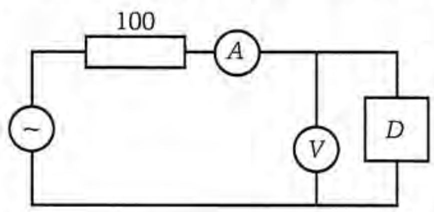
Center of mass of a semicircle. A closed figure in the form of a semicircle and the diameter connecting its ends was made by bending a piece of homogeneous thin wire. The radius of the semicircle is R. At what distance from the middle of the diameter does the figure's center of mass lie? (A. Chernoutsan)

P73

Vessel within a vessel. Inside a big vessel filled with helium under pressure P_0 = 1 atm and at temperature T_0 = 300 K there is a small vessel that has been evacuated. A small hole opens briefly in the small vessel and then closes. Some quantity of helium enters the small vessel. What will the temperature be inside the small vessel? (The walls of the small vessel do not allow heat to pass through.) (E. Butikov)

P74

Hungry for power. An electrical device D is connected in series to a 220-V AC network and a 100-Ω resistor. An ammeter shows a current of 0.5 A, and a voltmeter shows a voltage of 200 V. What is the device's power consumption? (A. Zilberman)



P75

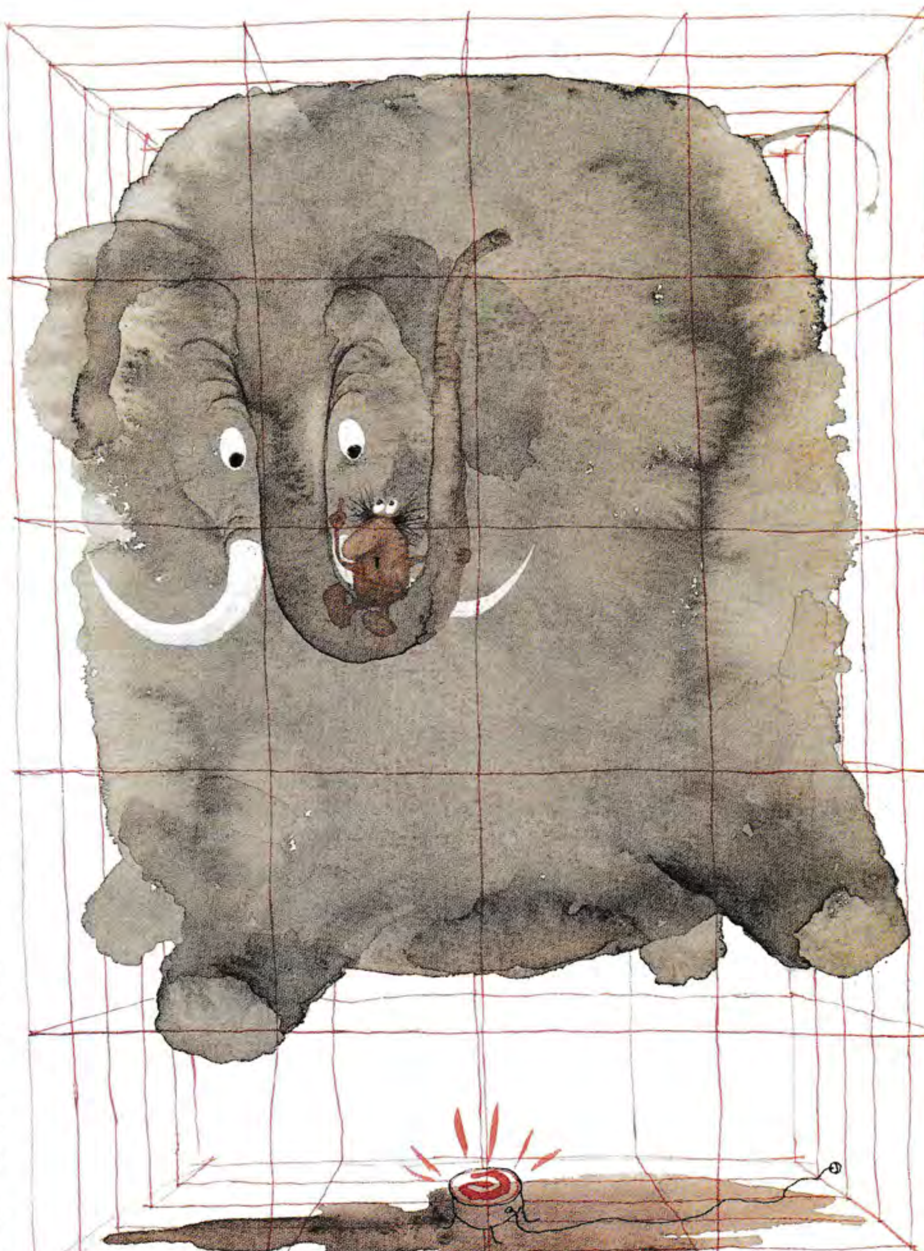
Nuclear winter. Suppose that a layer of soot has accumulated in the upper atmosphere as a result of widespread fires and that this layer absorbs practically all radiation from the Sun. What would be the average temperature on Earth? (It is now 300 K.) (A. Stasenko)

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

Glancing at the thermometer . . .

Why are we so sensitive to small changes in temperature?

by M. I. Kaganov



Art by Pavel Chernusky

ONE MORNING I FELT THAT my room was colder than usual. Glancing at the thermometer I saw that, in fact, the air temperature was 19°C instead of the usual 20°C. Complaining about the unreliability of my apartment building's services, I went to work. On the way my thoughts returned to the data from the thermometer, and I felt that something was wrong . . .

Temperature is a measure of the thermal motion of molecules. The average energy of thermal motion of molecules (for example, of gas in the air filling a room) is $\frac{3}{2}kT$, where $k \cong 1.4 \cdot 10^{-23}$ J/K (Boltzmann's constant), and the temperature has to be measured not in degrees Celsius but in kelvins, which are shifted relative to the Celsius scale by –273.16. So the temperature in my apartment was about 300 K. And I had *felt* a temperature change ΔT of the order of $\frac{1}{300}T$—that is, I felt that the energy of thermal motion of the air molecules changed by 0.3%! Not only that, without complicated instruments—by means of a simple wall thermometer—I verified my sensation: I had *measured* the 0.3% change in the energy of thermal motion of the molecules . . . I even felt a certain pride, thinking about the evolution that had created such sensitive mechanisms for perceiving temperature.

The importance of temperature for living organisms is well known: a change in body temperature of one degree is a sign of illness, and the range of allowable body temperatures is less than 10 degrees. So it's quite natural that living organisms feel temperature precisely. But how do they do it?

The wall thermometer gave me a hint. So let's start with that. How do we manage to measure a temperature change of one degree? Or in the case of a medical thermometer, 0.1 degree? Here, the measured property is the change in volume of a liquid (mercury, to be precise). With the increase or decrease in temperature, its volume V changes by ΔV, and the following formula holds:

$$\frac{\Delta V}{V} = \alpha \cdot \Delta T.$$

The factor α is called the coefficient of thermal expansion. Its order of magnitude is about 10^{-3} to 10^{-4} K^{-1}. There's only one way (and a very simple one) to "see" $\Delta V/V$ on the order of 10^{-4} and that is to "push" mercury in a thin capillary tube. Then $\Delta V = S \cdot \Delta l$, where S is the cross-sectional area and

$$\Delta l = \frac{V}{S}\alpha \cdot \Delta T$$

is the change in height of the mercury column. If S is small enough, you can get the required resolution. The capillary tube acts as an *amplifier*. If V is of the order of 1 cm^3, then to get Δl of the order of 1 mm = 0.1 cm with $\alpha \cdot \Delta T$ of the order of 10^{-4}, you need a capillary tube with a cross-sectional area $S = 10^{-3}$ cm^2. A very simple amplifier!

Now that we understand how the thermometer works, let's get back to living organisms. What acts as the amplifier in them? The determining role of temperature in processes of vital importance is connected with the fact that the rates W of most chemical reactions (without which life would be impossible) have a strong power-law temperature dependence:

$$W \propto e^{-U/kT}.$$

The value U, called the activation energy, is different for every reaction, but as a rule U greatly exceeds kT. Without going into details and without looking in any reference books, I worked it out like this. A chemical reaction is always a reconstruction of electronic states. For example, sodium (Na) is joined with chlorine (Cl) to form ordinary table salt (NaCl) in the following way: an atom of Na gives an electron to an atom of Cl; the ions Na^+ and Cl^- attract each other and create a molecule of NaCl, but the electrons in a molecule of NaCl are not distributed around the nuclei as they are in atoms of Na and Cl. To measure the energy of electrons in an atom, a special energy scale was created: the electron-volt (eV). One eV = $1.6 \cdot 10^{-19}$ J. The characteristic value of the activation energy U is of the order of 1 eV, and kT is about 1/40 eV at room temperature.

The relative change in the rate of a chemical reaction $\Delta W/W$ due to a relative change in temperature $\Delta T/T$ is given by[1]

$$\frac{\Delta W}{W} \cong \frac{U}{kT}\frac{\Delta T}{T}.$$

For $\Delta T/T \cong 1/300$, the relative change in the rate of the chemical reaction $\Delta W/W \cong 1/10$—that is, it is quite perceptible. Here the factor U/kT serves as an amplifier. Judging by our senses, it's quite reliable. It must be connected with the fact that there are many different chemical reactions in the body, and they are all (they must be!) carefully coordinated . . .

So it seemed that I understood everything and I calmed down . . . But another idea occurred to me. Until I had understood (though roughly and superficially) the mechanisms of amplification, I was surprised that I could feel and measure a relative change in thermal energy of a molecule approximately equal to 1/300. But in fact, we're discussing the change in the energy not of one but of all the molecules. The ratio $\Delta T/T$ is equal to the relative change in the energy of the gas when its temperature changes by ΔT. My new idea was this: what is the *absolute* value of the change in the energy of the gas if the temperature changed by 1 degree? Of course, I understood that it's easy to answer this question—Joule's mechanical equivalent of heat is well known. But I wanted to obtain an emotionally tinged answer, to feel whether it's large or not.

I decided to calculate the mass that can be lifted—say, h = 1 m—from an expenditure of the amount of energy needed to heat the air 1°C in a well-insulated room measuring 4 m × 5 m × 5 m = 100 m^3. The approximate calculation is quite simple. The gas energy is

$$E = \frac{3}{2}NkT,$$

where N is the number of gas particles; and the change in energy is

$$\Delta E = \frac{3}{2}Nk \cdot \Delta T.$$

It's easy to determine the number of gas particles in the room. A mole of gas under normal conditions has the volume V_0 = 22.4 liters ($22.4 \cdot 10^{-3}$ m^3), so there are $100/(22.4 \cdot 10^{-3}) \cong 5 \cdot 10^3$ moles of gas in the room. And the number of molecules in a mole (Avogadro's number) is $N_A \cong 6 \cdot 10^{23}$ mole^{-1}. So there are $N \cong 3 \cdot 10^{27}$ molecules in the room, and $\Delta E \cong 6 \cdot 10^4$ J. Now let's calculate the desired mass

---

[1] $W(T + \Delta T)$ is proportional to

$$e^{-\frac{U}{k(T+\Delta T)}} = e^{-\frac{U}{kT(1+\Delta T/T)}} \cong e^{-\frac{U}{kT}\left(1-\frac{\Delta T}{T}\right)},$$

where we have used the first two terms in the binomial expansion

$$(1+x)^n \cong 1 + nx + \frac{n(n-1)}{2}x^2 + \ldots,$$

which holds for $x \ll 1$ for all values of n. So

$$W(T+\Delta T) = W(T) \cdot e^{\frac{U}{kT}\frac{\Delta T}{T}} \cong W(T)\left(1 + \frac{U}{kT}\frac{\Delta T}{T}\right),$$

when $U \cdot \Delta T/kT^2 \ll 1$. We have used the approximation $e^x \cong 1 + x$, which is valid when $x \ll 1$.

Do you get the drift?

SECTIONS OF RAILWAYS OR PAVED ROADS THAT run in narrow gullies get snowbound even in the absence of a new snowfall. How does that happen? On the face of it the answer is clear: snow is transported by the wind. However, to gain a detailed understanding of this process, some serious research was needed.

In 1936 an English physicist by the name of Bangold examined the transport of sand by the wind in a wind tunnel.

It turned out that if the wind speed is less than a certain velocity v_1, the sand doesn't move. With a wind speed greater than v_1 but less than another velocity v_2, the sand may still stay put. But if a grain of sand from somewhere else lands on this motionless mass of sand, the impact knocks several other grains into the air. These grains get swept up by the wind, then fall, setting other grains in motion. Thus the sand is transported by the wind. If the wind speed exceeds v_2, the grains rise up and form a sand–air flux whose density is rather substantial but decreases with height. Figure 1 illustrates the trajectories of the grains.


Figure 1

Now it can be explained why gullies fill with snow in windy weather. In a gully the flow broadens (see figure 2, showing the lines of flow), and that is why its speed decreases. As a result, the equilibrium between the particles swept up and the particles falling down breaks down. The number of falling particles is greater than the number of rising particles, and so the gully gradually fills with snow.

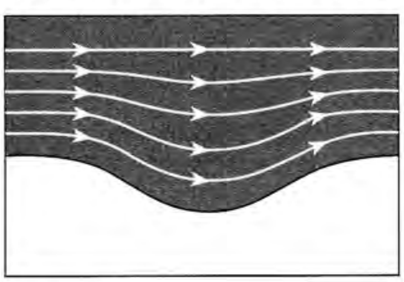


Figure 2

Analogous processes occur when snow transported by the wind meets an obstacle—for example, a tree. An upward movement of air occurs near the tree trunk on its windward side. This leads to the formation of a deep gully in the snow on the windward side of the trunk. Beyond the gully and slightly behind the trunk, where the wind speed is lower, a drift appears.

This phenomenon is used to protect sections of road in gullies from snowdrifts. At a certain distance a wooden fence is erected in front of the gully on the windward side. Beyond the fence a calm zone is created, with a gentle uniform wind, where all the snow swept by the wind is deposited.

The motion of sand dunes is explained similarly. Wind of sufficient force attacks a sand dune and sweeps the sand up on the windward side. On the leeward side, where the wind speed is lower, the sand falls. And this is how, over the course of time, dunes "wander."

—Lev Aslamazov

using the formula $\Delta E = Mgh$; $M \cong 6 \cdot 10^3$ kg, or 6 metric tons (!). This answer made me check the calculation three times. When I finally believed that the answer was correct, I recalled the appeals on television to conserve heat and the statement from the film "Life on Earth" that warm-blooded animals expend the greater part of the food they eat on keeping their body temperature constant. Heat is an expensive pleasure.

Nota bene

Everything you've read up to now can be considered a slightly organized stream of consciousness: I worked out practically everything in my head. After I had written and reread it, I thought: students are often frightened by the feeling that physics is a set of various seemingly disconnected facts, magnitudes, and relations. And here I have added to this heap. But I wanted to bring one very important idea to your attention.

Modern physics has penetrated so deeply to the very essence of things that it can evaluate, and often calculate exactly, innumerable parameters, constants—everything that science has accumulated at different stages of its development (often as experimentally obtained values). To make these calculations one need only use a few physical magnitudes bearing the mighty name of "universal constants." These are the electron charge $e \cong 1.6 \cdot 10^{-19}$ C; the electron and proton masses $m_e \cong 10^{-30}$ kg and $m_p \cong 1.7 \cdot 10^{-27}$ kg; Planck's constant $h \cong 6.6 \cdot 10^{-34}$ J $\cdot$ s (more often physicists use the constant $\hbar = h/2\pi \cong 10^{-34}$ J $\cdot$ s); and the speed of light $c \cong 3 \cdot 10^8$ m/s. Just think: all the magnitudes dealing with macroscopic physics[2] that can be measured can in principle be expressed by means of five universal constants![3] Such calculations are called calculations from first principles. Of course, theory isn't always so well developed that such a calculation can be carried through to the end with the required accuracy (which is why I said "in principle"). But it's quite clear that the calculation is possible, and there is no reason to expect that we'll come up against a problem that is unsolvable in principle.

I either remembered or looked up all the physical magnitudes in this article, but all of them can be obtained by calculating from first principles. I'll try to prove this by taking the activation energy U and the coefficient of thermal expansion α as examples. We won't go all the way to the bitter end, since the size of the atom a can undoubtedly be expressed by means of the universal constants listed above. For example, the size of the hydrogen atom $a_H = \hbar^2/m_e Ke^2$. You can find this in any book dealing with quantum mechanics.

Let's begin with the activation energy U. Since we're not going to develop a theory of the rates of chemical reactions but just show how the magnitudes are expressed with universal constants, we'll set a limit on the calculation of the ionization energy of the hydrogen atom U_{ion}—that is, we'll answer the question: what energy must be expended to wrench an electron from a proton?

The energy of an electron in a hydrogen atom is

$$E = \frac{m_e v^2}{2} - \frac{Ke^2}{a}.$$

In the Bohr model of the hydrogen atom, the centripetal force on the electron is provided by the electrostatic attraction

$$\frac{mv^2}{a} = \frac{Ke^2}{a^2}.$$

Therefore,

$$\frac{1}{2}mv^2 = \frac{Ke^2}{2a}.$$

This means that the kinetic energy of the electron is $-1/2$ its potential energy. So the energy is

$$E = -\frac{Ke^2}{2a}.$$

Substituting $a = a_H$, we get

$$U_{ion} = \frac{m_e K^2 e^4}{2\hbar^2} \cong 13.6 \text{ eV}.$$

The energy needed to reestablish electron states is usually less than U_{ion}. So when we evaluated the relative change in the rate of a chemical reaction, we took U to be of the order of 1 eV.

Calculating the coefficient of thermal expansion is more complicated. It requires that we know the structure of the expanding object. We'll have to limit ourselves to the simplest approach, keeping the main point in mind: that thermal expansion is the result of the temperature dependence of the average equilibrium distance between particles.

So, two particles are situated at a distance $d + x(t)$ from each other, where d is the distance between them when they are stable (at absolute zero) and $x(t)$ is the instantaneous (at moment t) deviation of the particle from the equilibrium position. The force F acting on the particle resembles the elastic force acting on a mass fastened to a spring—it's proportional to the displacement of the particle from its equilibrium position: $F = -\kappa \cdot x(t)$. But the average force $\overline{F}$ (for a sufficiently large period of time) must be equal to zero.[4] And this means that $\overline{x(t)} = 0$ as well—that is, the average distance between particles is equal to d and does not depend on the displacement of the atoms, and so it does not depend on temperature either. The scientific expression of the this result is: *a harmonic approximation cannot describe the thermal expansion of objects.*

[2]It's not by chance that we limit ourselves to macroscopic physics. We still aren't able to calculate, for example, the masses of the elementary particles—various mesons, hadrons, and so on.

[3]In the MKS system of units we also use Coulomb's constant $K = 9 \cdot 10^9$ N $\cdot$ m^2/C^2, but this can be absorbed into the value for the electronic charge by a change of units, as was done in the cgs system of units. We also choose to use a capital K to distinguish Coulomb's constant from Boltzmann's constant.—*Ed.*

[4]If the average force $\overline{F}$ were not zero, the particles would have to move somewhere.

I have given this strict scientific statement in order to introduce the word "approximation." The point is that the expression for the force that we used is approximate. Let's try to define it more precisely by taking into account the components that are nonlinear relative to the displacement $x(t)$:

$$F = -\kappa \cdot x(t) + \beta \cdot x^2(t) + \dots .$$

The theory based on this formula, or similar expressions, is called anharmonic, and the coefficient β is called an anharmonic coefficient. It follows from this formula that

$$\overline{x} = \frac{\beta}{\kappa}\overline{x}^2.$$

Therefore,

$$\frac{\Delta V}{V} = \frac{(d+\overline{x})^3 - d^3}{d^3} \cong \frac{3\overline{x}}{d} = \frac{3\beta}{d\kappa}\overline{x}^2.$$

The coefficient 3 appeared because a body can expand in three directions. In this expression V is the volume of the body at $T = 0$ K. So if we are to complete the calculation, we must be able to calculate the values of β, κ, and $x^2(t)$. Let's start with the last one. As the anharmonic component βx^2 is a small correction (we had to incorporate it in the force expression only because the answer without it was zero), we can say that the potential energy of movement is $\kappa \cdot x^2(t)/2$, and the total energy is

$$E = \frac{mv^2}{2} + \frac{\kappa x^2}{2}.$$

But the average kinetic energy and the average potential energy are equal. So

$$\overline{x}^2 = \frac{1}{\kappa}\overline{E}$$

The average energy of oscillatory motion is kT.[5] So

$$\frac{\Delta V}{V} = \frac{3\beta}{d\kappa^2}kT.$$

Now we have to figure out how to calculate the $3\beta/\kappa^2$ factor. This is certainly the most complicated part of the problem. But we'll make the maximum permissible simplifications (though even here you'll have to take one point on faith).

Suppose we consider ions with charges $+e$ and $-e$ situated a distance r apart. The electrostatic attractive force is equal to Ke^2/r^2. But this force can't be the only one—otherwise the ions would fall on each other. When they approach each other too closely, they repel one another, and the law of repulsion can be explained by equations of quantum mechanics.

And now we've come to the point where you just have to believe me. The total force for the ion interaction is

$$F = \frac{Ke^2}{r^2} - \frac{A}{r^{10}}.$$

If $r = d$, the force must be equal to zero. So $A = Ke^2d^8$ and, finally,

$$F = \frac{Ke^2}{r^2} - \frac{Ke^2d^8}{r^{10}}.$$

Substituting $r = d + x$, we expand F in powers of x (limiting ourselves to the first two powers of x—see footnote 1). The first power will give us the value of κ and the second power will give the value of β. Calculate it yourself, and you'll see that

$$\frac{\Delta V}{V} = \frac{3 \cdot 52}{64}\frac{kdT}{Ke^2}$$

—that is,

$$\alpha = \frac{3 \cdot 52}{64}\frac{kd}{Ke^2}.$$

The distance d between atoms is approximately equal to the size of the atom: $d \cong a \cong 3 \cdot 10^{-10}$ m = 30 nm, and when we substitute the values for k, K, e, and d, we get

$$\alpha \cong 5 \cdot 10^{-5}\ \text{K}^{-1}.$$

Look at the reference tables and you'll see that the estimate we obtained (even with our extreme simplifications) isn't bad at all.

Well, it seems we're through. But I'd like to mention that we could guess the order of magnitude of the coefficient of thermal expansion α. Look at the last expression for $\Delta V/V$. You can see that the *nondimensional* ratio $\Delta V/V$ is approximately equal to the ratio of the thermal energy kT (of one particle) to the binding energy of the particles (here approximately e^2/d^2). So we conclude: the stronger the molecular bonds in the body, the smaller the coefficient of thermal expansion. You might notice this law in the tables, where α is given together with the melting point for solid bodies or the boiling point for liquids. But, of course, such an assertion is not a law of nature. There may be exceptions to it . . .

[5]Compare this with the energy $\frac{3}{2}kT$ for a particle in a gas: the particle is free—that is, its potential energy is equal to zero; so there is $\frac{1}{2}kT$ for every degree of freedom. Here there are only two degrees of freedom.

Be a factor in the

QUANTUM

equation!

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

Send your inquiries to:

Managing Editor
Quantum
3140 N. Washington Boulevard
Arlington, VA 22201

When we say, "Let's see if it will fly," we mean it!

Embry-Riddle engineering students use CADD resources to aid in the design of the their two senior team projects. The curriculum includes a mix of theory and practical applications that fully prepare students for aerospace and aviation careers.

Over 67 international university teams entered the 7th Annual Aero Design Competition, co-hosted by the Embry-Riddle Society of Automotive Engineering Chapter. The objective of the competition is to lift as much weight as possible with a R/C aircraft designed and built by students.

Aerospace engineering students at Embry-Riddle don't have to wait until tomorrow to apply what they're learning today. The curriculum builds to two senior design projects that test your knowledge of aerodynamics, controls, materials, propulsion and structure as your team designs the internal and external details of an airplane or spacecraft. Both your imagination and abilities are put to the test, just as they will be in your professional life. We believe that's what education is all about.

Electrical Engineering students at Embry-Riddle receive a complete, broad-based education with full laboratory resources and equipment to complement academic study.

You may know about us, ***but did you know***:

- We offer over 20 different accredited aviation/aerospace majors at the Bachelor and Master's levels including aerospace engineering, electrical engineering, engineering physics, avionics engineering technology, aircraft engineering technology and aviation computer science.
- Our aerospace engineering enrollment at our Daytona Beach, Florida and Prescott, Arizona campuses is the largest in the United States. We are listed as one of the 18 best engineering schools in the nation by *U.S. News and World Report's 1992 College Guide* called "America's Best Colleges."
- Your university life is enriched through many social and professional clubs and organizations, like our SAE Chapter, that provide the opportunity for a variety of living and learning experiences, related to your major, outside the classroom.

Want to test your career design? Write to us today.

EMBRY-RIDDLE
AERONAUTICAL UNIVERSITY
Look at us now.

Mail to: Embry-Riddle Aeronautical University
Director of University Admissions
600 South Clyde Morris Boulevard
Daytona Beach, FL 32114-3900
Or Call 1-800-222-ERAU

Circle No. 1 on Reader Service Card

Please send me more information about Embry-Riddle.

Engineering Related Programs:	**Other Aviation/Aerospace Majors:**
☐ Undergraduate	☐ Undergraduate
☐ Graduate	☐ Graduate

Name ______________________

Address ______________________

City ______________ State ______ Zip ______________

Phone (   ) ______________________

MZ

KALEIDOSCOPE

What harmony means

Pythagoras struck a mathematical chord

by Vladimir Dubrovsky and Anatoly Savin

THE RENOWNED PYTHAGOras (6th century B.C.) was not only a great philosopher and mathematician. He was perhaps the first (at least, in European culture) to study the laws of musical harmony scientifically. According to one legend, he observed that three equally taut strings played together produce a particularly pleasant blend of sounds if their lengths are in the ratio 6 : 4 : 3. He found a neat relation between the three numbers: the reciprocal of the one in the middle is the arithmetic mean of the reciprocals of the two extremes: $1/4 = (1/6 + 1/3)/2$. Since then, the number $h = h(a, b)$ defined for arbitrary a and b in a similar way—as $h^{-1} = (a^{-1} + b^{-1})/2$, or

$$h = \frac{2ab}{a+b}$$

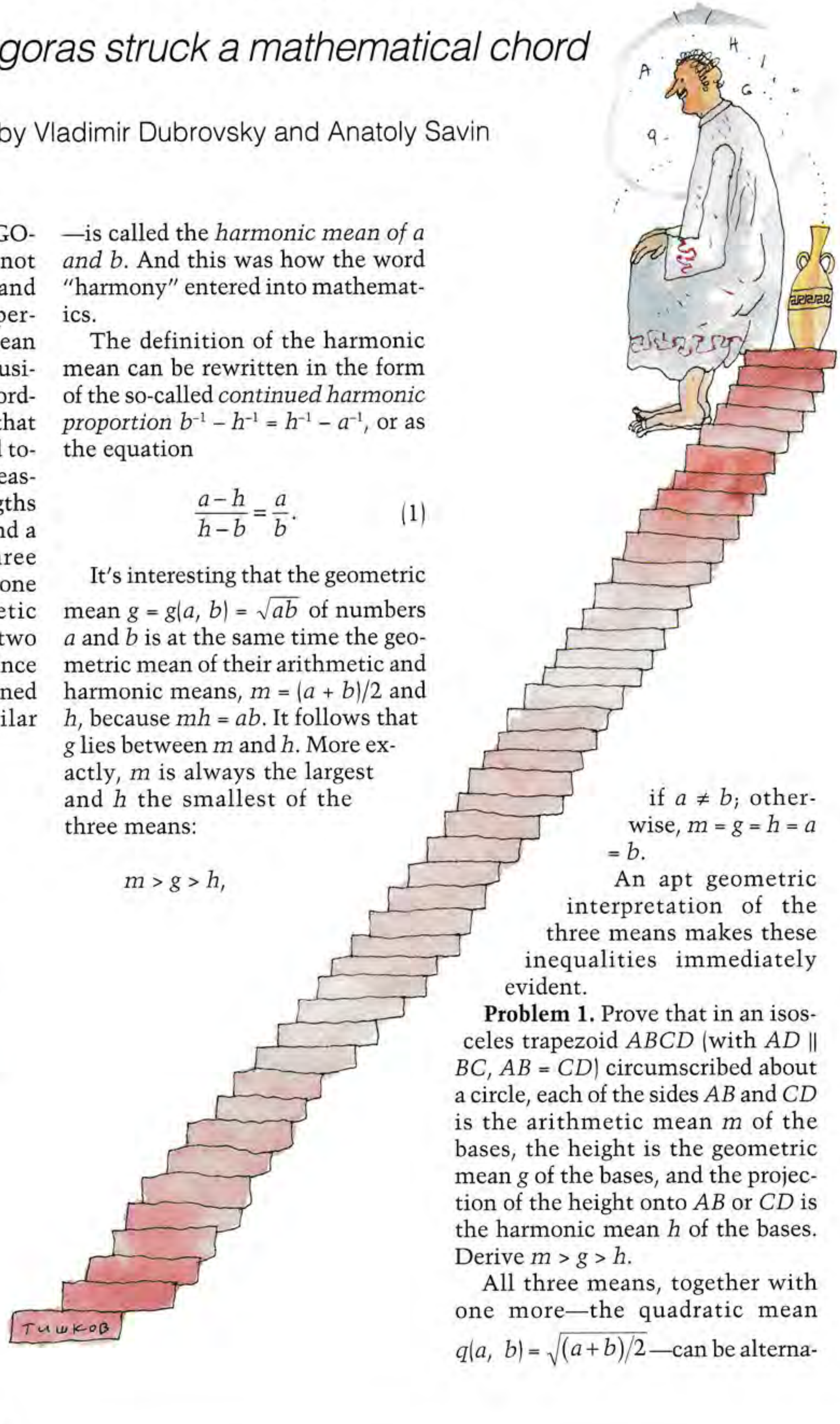
—is called the *harmonic mean of a and b*. And this was how the word "harmony" entered into mathematics.

The definition of the harmonic mean can be rewritten in the form of the so-called *continued harmonic proportion* $b^{-1} - h^{-1} = h^{-1} - a^{-1}$, or as the equation

$$\frac{a-h}{h-b} = \frac{a}{b}. \qquad (1)$$

It's interesting that the geometric mean $g = g(a, b) = \sqrt{ab}$ of numbers a and b is at the same time the geometric mean of their arithmetic and harmonic means, $m = (a + b)/2$ and h, because $mh = ab$. It follows that g lies between m and h. More exactly, m is always the largest and h the smallest of the three means:

$$m > g > h,$$



if $a \neq b$; otherwise, $m = g = h = a = b$.

An apt geometric interpretation of the three means makes these inequalities immediately evident.

Problem 1. Prove that in an isosceles trapezoid $ABCD$ (with $AD \parallel BC$, $AB = CD$) circumscribed about a circle, each of the sides AB and CD is the arithmetic mean m of the bases, the height is the geometric mean g of the bases, and the projection of the height onto AB or CD is the harmonic mean h of the bases. Derive $m > g > h$.

All three means, together with one more—the quadratic mean $q(a,\ b) = \sqrt{(a+b)/2}$—can be alterna-

Art by Leonid Tishkov

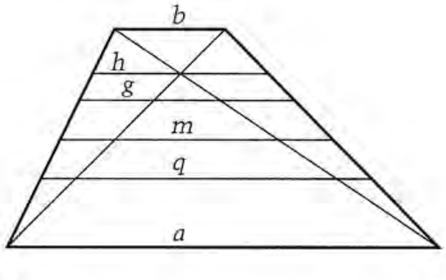


Figure 1

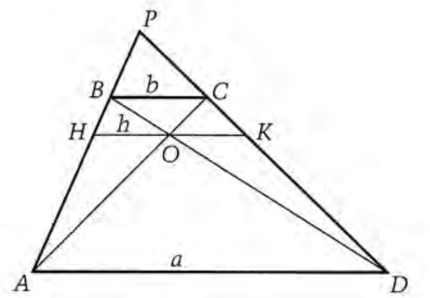


Figure 2

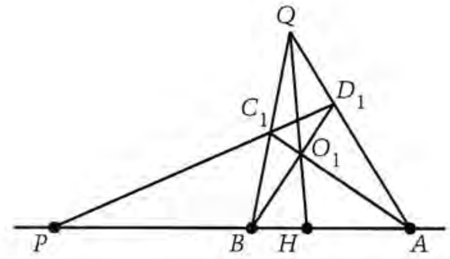


Figure 3

tively illustrated as segments in a trapezoid parallel to its bases a and b (fig. 1). It's well known that the median m is the arithmetic mean of a and b: $m = (a + b)/2$. The segment g splitting the trapezoid into two trapezoids similar to each other is the geometric mean of the bases, because the bases of the two smaller trapezoids are in the same ratio: $a/g = g/b$. Further, segment q is drawn to bisect the area of the trapezoid.

Problem 2. Show that q is the quadratic mean $q(a, b)$ of the bases, and that segment q is closer to, and segment g further from, the longer base than the median m (implying $q > m > g$).

The harmonic mean $h(a, b)$ is represented in figure 1 by the segment h through the intersection point of the trapezoid's diagonals. To prove this, extend the sides AB and CD of the trapezoid in figure 2 to meet at P. From similar triangles APD, HPK, and BPC (the notations are indicated in the figure),

$$a : h : b = AD : HK : BC = PA : PH : PB.$$

This observation, together with a little algebra, means that it will suffice to show that *PH is the harmonic mean of PA and PB*, or, applying equation (1), that

$$\frac{HA}{BH} = \frac{PA - PH}{PH - PB} = \frac{PA}{PB}. \qquad (2)$$

We've mentioned that the right side equals AD/BC. The left side is equal to OA/OC (because lines HO and BC are parallel), and so, by the similarity of triangles AOD and COB, to AD/BC too, thus completing the proof.

Problem 3. In figure 2, let $OH = t$ and $OK = u$. Use similar triangles such as ABC, AOH and BAD, BOH to show that $1/a + 1/b = 1/t = 1/u$, giving another proof that t is the harmonic mean of a and b.

Points A, B, H, and P are an example of a so-called *harmonic range* of points, defined as any four collinear points $(A, B; H, P)$ satisfying

$$\frac{PA \cdot HB}{PB \cdot HA} = -1, \qquad (3)$$

where the segments on the left side should be regarded as directed segments (that is, their lengths should be assigned so that the segments of the same direction—for instance, PA, PB, and HA in figure 2—receive the same sign, while those of the opposite direction—HB in figure 2—receive the opposite sign). With this sign convention, condition (3) implies equalities (2) irrespective of the order of the four points (for instance, $HA = PA - PH$ turns out to be valid for any three points along a line). Therefore, for any harmonic range $(A, B; H, P)$ the (directed) segment PH is the harmonic mean of PA and PB.

And not only that! Clearly, we can swap points in the pairs (A, B) and (H, P), and swap the pairs, without violating equality (3). So these rearrangements yield new harmonic ranges of the same points taken in a different order, and new harmonic means too. All in all there are four harmonic means hidden in a harmonic range: the directed segment from *any* of the four points to its "mate" is the harmonic mean of the directed segments from this point to the points of the other pair.

A simple way to complete a harmonic range $(A, B; H, P)$ given three of its points—say, A, B, P—and, by doing so, to construct $h(PA, PB)$ is shown in figure 3. It's done with a ruler alone. We draw two arbitrary lines through A and B, intersecting at point Q. Then draw any line through P, intersecting QA and QB in C_1 and D_1, respectively. Find the intersection O_1 of AC_1 and BD_1;

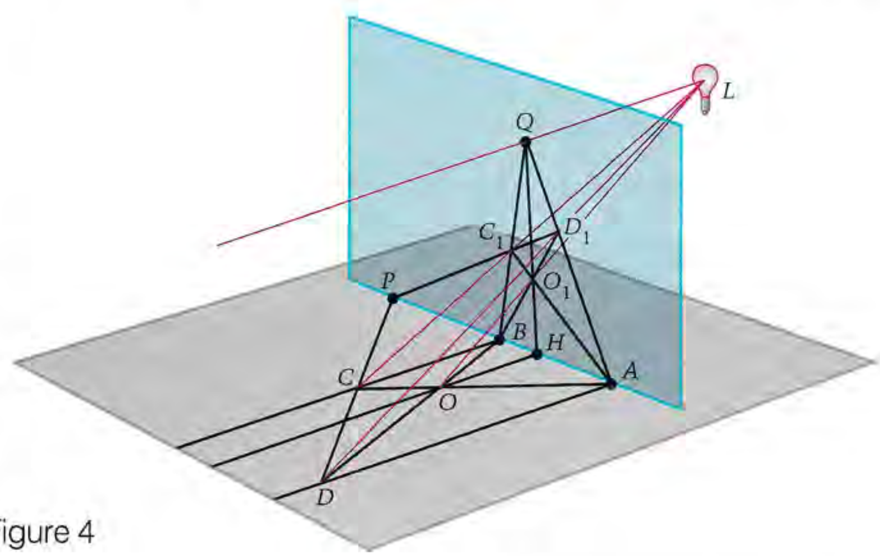


Figure 4

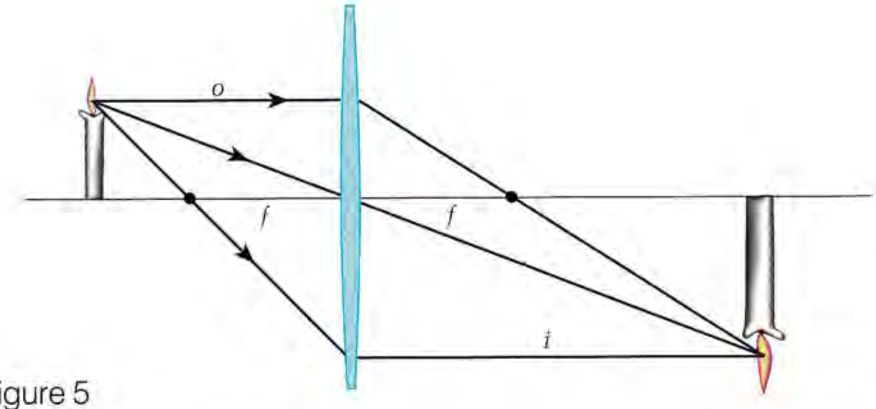


Figure 5

then line QO_1 cuts AB at H.

Indeed, let figure 3 be drawn on a sheet of glass with line PA running along its edge. Stand the glass upright on a plane surface (fig. 4) and shine a light on it from point L (the same height above the plane as point Q in our diagram). Now look at the shadow cast by the diagram on the plane. Since the shadow of a straight line is again a straight line, the entire shadow is basically the same as the original drawing except that lines meeting at Q become parallel in the shadow (because the ray LQ is parallel to the plane). So if we perform the construction on the glass, the corresponding shadows will reproduce figure 2, in which H indeed completes the harmonic range $(A, B; H, P)$.

Figures 3 and 4 show the case of P outside segment AB; but if you switch the labels of points P and H, C_1 and O_1, you'll see at once that the method works as well for the third point given between A and B.

This "optical" proof leads us directly to geometrical optics, where the harmonic mean presents itself in a very important and useful formula.

Figure 5 shows the construction of the image of a candle (it could be any other object) in a converging thin lens by the ray-tracing method. We follow the path of three rays of candlelight: the ray parallel to the lens's principal axis (horizontal line) is refracted to go through the focal point beyond the lens; the ray through the focal point in front of the lens becomes parallel to the axis; and the ray pointing at the center of the lens goes through without deviating. The three rays always meet at the same point, and this is where the image is formed. This is how the lens works—just take it for granted, if you haven't studied it yet. But look at the figure formed by the rays, the lens, and the axis! Do you recognize it? Of course you do! This is our trapezoid again. And free of charge we get the *thin-lens equation*: the distance between the focal points, or double focal length $2f$, is the harmonic mean of the object distance o and image distance i. This is usually written as

$$\frac{1}{o}+\frac{1}{i}=\frac{1}{f}.$$

Another, simpler, physical application of the harmonic mean emerges when we calculate an average speed. What's the average speed v of a jogger who runs the first half of his daily route at a speed v_1 and the second half at a speed v_2? The answer that springs to mind first, $v = (v_1 + v_2)/2$, is wrong! Indeed, dividing the total length of the route—say, $2l$—by the total elapsed time, we get

$$v=\frac{2l}{2l/v_1+2l/v_2}=\left(\frac{v_1^{-1}+v_2^{-1}}{2}\right)^{-1}$$
$$=h(v_1v_2).$$

Problem 4. An oval running track consists of two long parallel straight segments and two short curved segments connecting them. If the wind blows along the straight segments of the track, is it easier or harder for a runner to show the same time per lap as in calm weather? (Neglect the effect of the wind on the curved segments; assume the runner's speed with respect to the air is constant along the track.)

Before asking one more question that's partly physical, we must introduce the sequences that relate to the harmonic mean as arithmetic and geometric sequences relate to the corresponding kinds of mean—that is, sequences whose every term except the first is the harmonic mean of its neighbors. They're called *harmonic sequences* and can equivalently be defined as consisting of the reciprocals of the terms of an arithmetic sequence. The simplest of them—1, 1/2, 1/3, ...—can be obtained geometrically by way of the beautiful *Brianchon construction* shown in figure 6.

Problem 5. The lines AA_1 and BB_1 in figure 6 are parallel, and $AA_1 = 1$. Points $A_2, A_3, \ldots$ are constructed successively as shown in the figure (red lines illustrate the construction of A_4 after A_3). Prove that $AA_n = 1/n$.

Adding up successive terms of this sequence, we can obtain an ar-

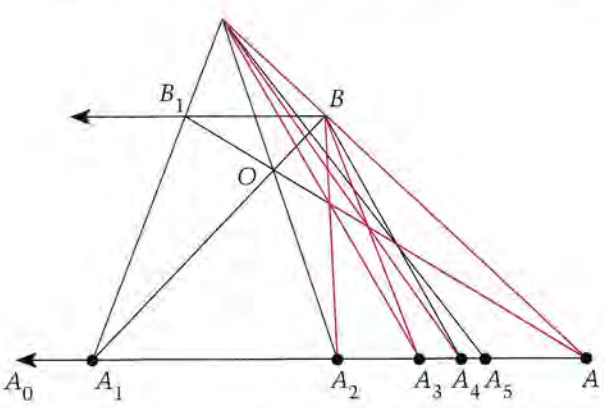


Figure 6

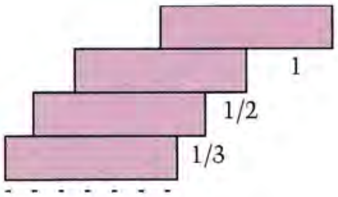


Figure 7

bitrarily large number. In other words, the *harmonic series* $1 + 1/2 + 1/3 + \dots$ is divergent, or the sequence $S_n = 1 + 1/2 + \dots + 1/n$ approaches infinity as $n \to \infty$.

Problem 6. Show that

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \geq \frac{1}{2},$$

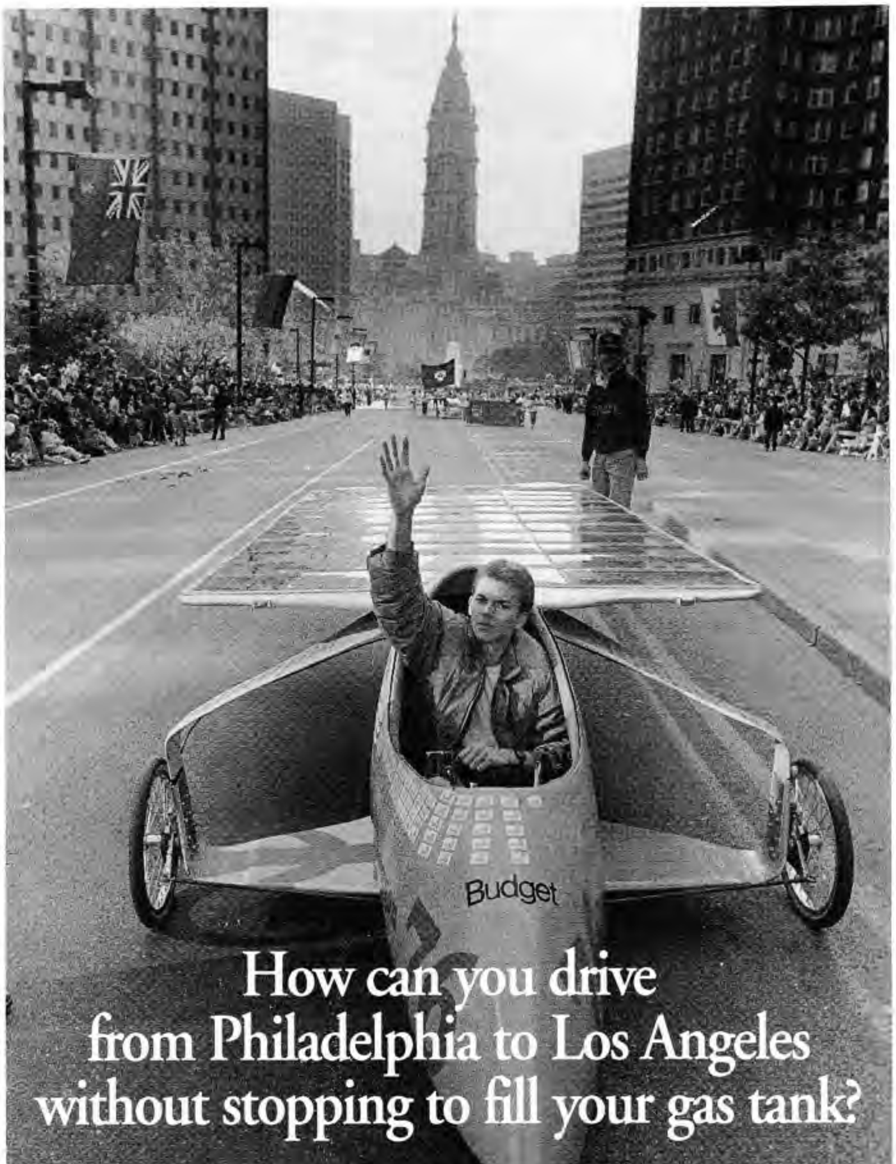
and use this estimate to prove the divergence of the harmonic series.

Here's a popular trick question: how far might a staircase of bricks laid one upon another, one brick per layer, protrude over its base without tumbling down? The answer is surprising: infinitely far! Figure 7 illustrates such a "leaning tower": the $(n + 1)$st brick counted from the top is shifted by $1/n$ with respect to the nth one, where the unit measure is half the length of a brick. Of course, such a tower of $n + 1$ bricks overhangs a length of $S_n = 1 + 1/2 + \dots 1/n$, which can be made arbitrarily long.

Problem 7. Show that the leaning tower of bricks doesn't topple down.

In fact, the underpinning of Pythagoras's discovery of the harmonic mean in the course of his musical experiments was the harmonic sequence 1, 1/2, 1/3, ... The point is that a sounding string clamped at both ends vibrates not only with its whole length but also with its halves, thirds, quarters, and so on. Similar things can be said of other acoustic devices, including the human ear. So the sounds produced by three strings whose lengths form the harmonic proportion are in a certain sense related to each other—a sense discerned by the ear as harmony. Of course, this is but a very rough explanation of the subtlest matter of musical harmony.

ANSWERS, HINTS & SOLUTIONS ON PAGE 59



Solar power.

Since 1989, teams of Drexel students and faculty members have designed, built and raced three working solar-powered cars—SunDragon I, II and III. We also have been selected to participate in two nationally sanctioned solar vehicle races.

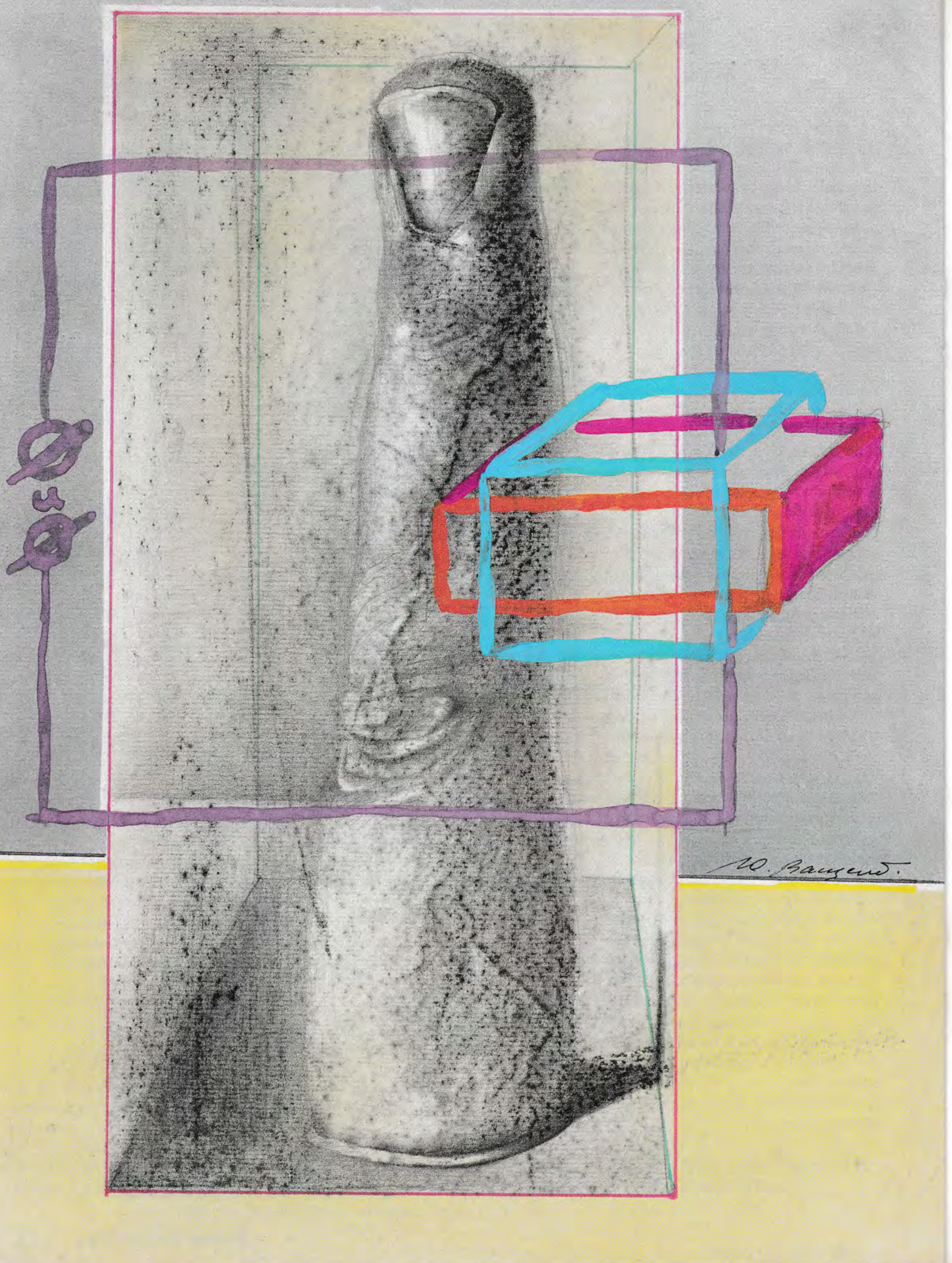
We're ready and waiting to tell you and your students all about our SunDragon projects and fill you in on: photovoltaics, aerodynamics, mechanics, thermodynamics, mathematics, project management, vehicle design, materials engineering, advanced composites, fund-raising and energy management.

Lesson plans and a show-and-tell set of transparencies outlining the development of our solar-powered vehicles are available for your science and mathematics classes.

For more information about these unique teaching resources or to obtain copies for your classes, contact Amy Warlin, admissions special projects coordinator, at **(215) 895-2400** or write to Michel Barsoum, associate professor of materials engineering, 4445 LeBow Engineering Center, Drexel University, Philadelphia, PA, 19104.

DREXEL
UNIVERSITY

Circle No. 8 on Reader Service Card


Ю. Ващенко.

Tactile microscopes

A sophisticated way to "feel your way around"

by A. Volodin

WE HAVE ALWAYS BEEN interested in the microcosm—the world of things that cannot be seen with the naked eye. Research into the nature of the microcosm cannot be overestimated. It has saved mankind from many terrible diseases, helped engineers to create the miracles of microelectronics, and allowed scientists to understand the nature of most of the phenomena that take place in the visible world.

But in order to understand the structure of the microcosm, we first of all must be able to see it. Wishing to see this invisible world, in 1674 the Dutch researcher Antonie van Leeuwenhoek invented the optical microscope, which you've certainly had occasion to use in your biology lab. It contains a system of refracting lenses that magnify the image of an object. Centuries passed, and the optical microscope worked so well in so many areas of science that it became a symbol of the scientific pursuit. Its capabilities were limited, however—it could reach only to the micron level. Scientists were just as interested in the submicron region, where dimensions are measured in nanometers. But today's optical microscopes differ little from Leeuwenhoek's first microscope. His could enlarge an image by a factor of 200, while its modern counterpart can't manage a magnification factor greater than 1,000. What's going on here?

Art by Yury Vashchenko

Abbe's prohibition

Over 100 years ago the German physicist and optician Ernst Abbe proved that there are fundamental limitations for any microscope working with light (or any other radiation) focused by lenses. The most important restriction is caused by diffraction—the ability of a wave to envelop the object. It "hides" details that are less than half the radiation's wavelength. As the wavelength of visible light is about half a micron, you can't see an object less than 0.1 micron in size with an optical microscope.

So to enter the submicron world, it's logical to use radiation of shorter wavelength—for example, X-rays or an electron flux (like all elementary particles, the electron is also a wave). In the 1930s, electron microscopy made its appearance.

"Particles" of any radiation, regardless of its origin, have a universal characteristic: their energy. The higher the energy, the shorter the wavelength. An instrument built on the principle of the optical microscope but working with an electron beam reflected by special magnetic lenses is called an electron microscope. Electron waves are shorter than light waves by a factor of about 1,000, so the magnification with the best electron microscopes is up to 1,000,000x.

But it wasn't so easy to get to that point. An electron microscope is thousands of times larger, more expensive, and more complicated than an optical microscope. It also suffers from a serious flaw. It destroys the object being examined. Electrons with energies of tens of electron-volts kill all living things. They also cause defects in crystals, disrupting the regular placement of the atoms. Nevertheless, electron microscopy has made it possible to take a great investigative step into the submicron world.

But—was there some other way to plumb the depths of this miniature universe?

How Abbe's prohibition was overcome

In the middle of the 1980s, the development of microscopic instruments changed direction sharply. Earlier progress had been achieved within the framework of Abbe's prohibition by decreasing the wavelength of the radiation that forms the image in the microscope. Now this prohibition was simply evaded. A new generation of microscopes—scanning probe microscopes—appeared. These devices made it possible to investigate a surface from a very close distance. Microscopes had previously been based on sight; these new instruments were based on touch. They not only show the shape and location of the smallest details of an object, they also provide information on other characteristics of the object. For instance, the magnetic probe microscope "feels" the heterogeneity of an object's magnetization; the electrical probe microscope senses the microdistribution of electric fields; the thermal probe microscope can tell the difference in temperature; and so on. As a result, every basic physical parameter has a corresponding probe microscope. Probe microscopes based on electron tunneling and forces of interatomic interaction even allow us to see individual atoms.

But we're getting ahead of ourselves. Let's take a closer look at the optical probe microscope.

Feeling with light

Imagine a nontransparent conical sheet with a tiny hole in the tip, its diameter less than half the wavelength of light (fig. 1). If we transmit light through such a cone, the light won't get very far—the light wave won't "crawl" through a hole of this size, so it will reflect back. Light can still be found on the other side of the hole, but only close by—about half a wavelength. The new generation of optical microscopes makes use of this "sagging" light. (This sagging is quantum-mechanical in nature. It is characteristic of any wave or particle and is called "tunneling.")

Let's place the object we want to examine near the hole, at a distance less than its diameter. A spot of light, approximately equal in size to the diameter of the hole, will appear on its surface. Light reflected by the object can be caught by a photoconverter—an instrument that converts weak light fluxes into an electrical signal. This signal can be amplified and again be depicted as a light spot on a monitor screen. The brightness of the point on the screen will correspond to the intensity of the captured light.

Now let's move the point-probe along the surface of the object, line by line. The spot of light from the probe will run along the surface. This procedure is called "scanning" the surface. If the path traveled is marked by points of different brightness on the monitor screen, we'll obtain an image of the surface. The resolution of the image corresponds to the diameter of the spot illuminating the surface—less than $\lambda/2$. (The size of the smallest visible detail of an object is called the resolution of the microscope.)

So Abbe's prohibition was overcome. The new instrument that made it possible was called the near-field optical scanning microscope. As paradoxical as it may seem, this microscope allows us to see details that are much smaller than the wavelength of light!

Piezoelectric fingers

How can the surface of an object be scanned so precisely? That's where piezoelectric manipulators come in. The simplest of these is shown in figure 2. It's made of a special ceramic whose size changes slightly with changes in the applied electric field. The manipulator is placed between two capacitor plates, which are often simply electrodes in the form of thin metal layers. By changing the voltage between the electrodes by 0.1 V, you can increase the length of the bar by 0.1 nm—that is, by the diameter of an atom. (The metal layer on the surface is thin enough so as not to interfere with this displacement.)

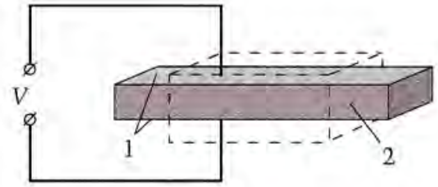


Figure 2
Piezoelectric manipulator. 1—electrodes; 2—piezoceramics.

A simple device made of three manipulator bars connected perpendicularly (fig. 3) can move the probe, situated at the point of intersection, in any spatial direction. Three driving voltages V_x, V_y, and V_z determine the coordinates x, y, and z of the probe's position. By changing the voltages V_x and V_y, we can move the probe along the surface of the object under investigation, scanning it along parallel lines at a given distance from one another (like the beam inside a television set). The voltage V_z, on the other hand, moves the probe up and down. If V_z is constant, the scanned surface moves toward or away from the probe because of the object's unevenness. This presents a problem for the recording system, though—the signal changes drastically, and the

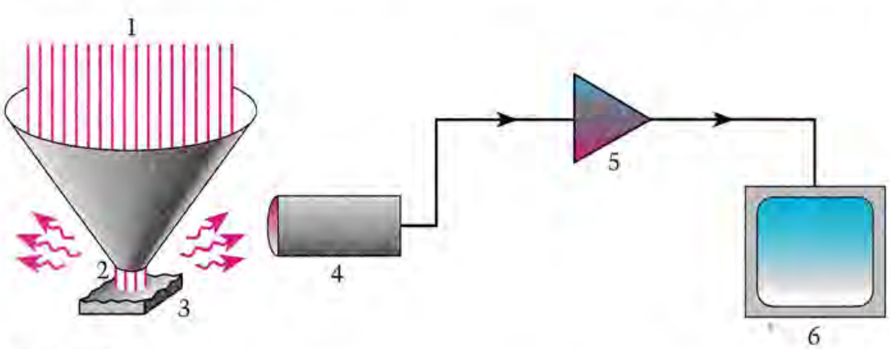


Figure 1
The near-field optical scanning microscope. 1—incident light; 2—"sagging" light; 3—object under investigation; 4—photoconverter; 5—amplifier; 6—monitor.

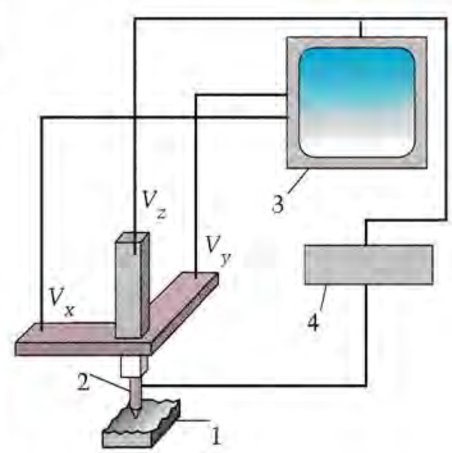


Figure 3
Scanning probe microscope. 1—object under investigation; 2—probe; 3—monitor; 4—feedback system.

probe may even collide with a large bump on the object. To get around this, an element of "self-control," or negative feedback, is introduced. It forces the probe to move up and down according to the surface relief.

Touching atoms

Let's take a closer look at this feedback mechanism, taking the scanning tunneling microscope as an example. This was in fact the first probe microscope. G. Binning and G. Rorer, working at IBM in Zurich, Switzerland, were awarded a Nobel Prize for their work on this invention. The probe in this microscope is an extremely sharp metal needle. The role of the hole in the probe of the optical scanning microscope is played here by the tip of the needle, from which (playing the role light played) quantum-mechanical waves of electrons contained in the metal of the point "sag" (fig. 4). The length of these electron waves is smaller than the wavelength of light by a factor of about 1,000, and so they can "illuminate" an area that is correspondingly smaller than the area accessible to the optical probe. When such an electron wave touches the surface being investigated (at distances of about 1 nm between the probe and the surface), an electron from the point can jump onto the surface—that is, it "tunnels." Tunneling means that an electric current arises in the probe–surface circuit—a very weak current (several billionths of an ampere). But amplification of such a current presents no problem to modern electronics. It's important that it has a pronounced dependence on the distance between the surface and the point of the probe. Decreasing the distance by 0.1–0.2 nm—that is, the size of an atom—increases the tunnel current by a factor of a thousand. This dependence is described by an exponential function with base $e = 2.718\ldots$.

Now let's get back to the feedback mechanism that ensures the proper functioning of any probe microscope. It's a rather complicated and sensitive electronic circuit that registers the change in the tunnel current and the varying voltage V_z applied to the vertical manipulator. The piezoelectric manipulator shifts the probe so that the tunnel current remains constant. This is possible only if the distance between the probe and the surface remains constant. So the feedback mechanism does not allow the probe either to move too far away from the surface or to collide with it. Because of the extreme sensitivity of the tunnel current to the distance from the surface, the precision of the feedback mechanism system is very high—0.01–0.001 nm. As a result, the point moves along a trajectory that closely mimics the relief of the scanned surface. As the voltage V_z is proportional to the height of the point above the surface at a given moment, it provides a good measure of the relief. Information on surface relief is fed to a computer and, after processing (filtering out noise and parasitic signals), is drawn on a monitor as the "topographic map" of the surface (fig. 5). It is sometimes rendered as a half-tone image, in which the height of the relief is marked by the intensity of coloration.

Scientists have used scanning tunneling microscopes to obtain detailed images of the surfaces of many crystal and polymer materials with resolution down to the atomic level. The scanning tunneling microscope provides the unprecedented magnification of 100,000,000x!

Researchers have already gotten used to the fact that the piezoelectric manipulators can be shifted with a precision corresponding to the size of an atom. They have even learned to use the point of the tunneling microscope as a working tool in the nanometer microcosm. The point—one atom wide—can be positioned at exactly the chosen place in a molecule and slice it in two. An atom can be caught and carried to where one wants it. In an IBM laboratory scientists managed to make inscriptions out of chains of atoms. One such inscription—the IBM logo—was made of separate atoms of xenon on a surface of nickel crystal. Chaotically distributed atoms of xenon adhering to the nickel were

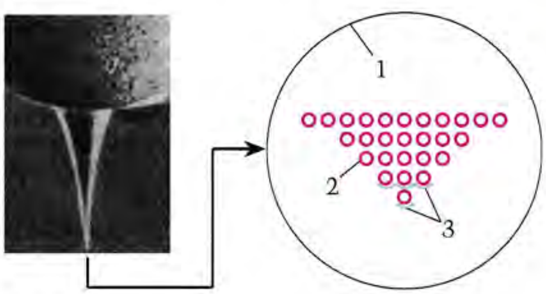


Figure 4
Probe of tunneling scanning microscope. 1—"enlarged" point; 2—atoms; 3—electron clouds. At left: the point (electron-microscopic photo).

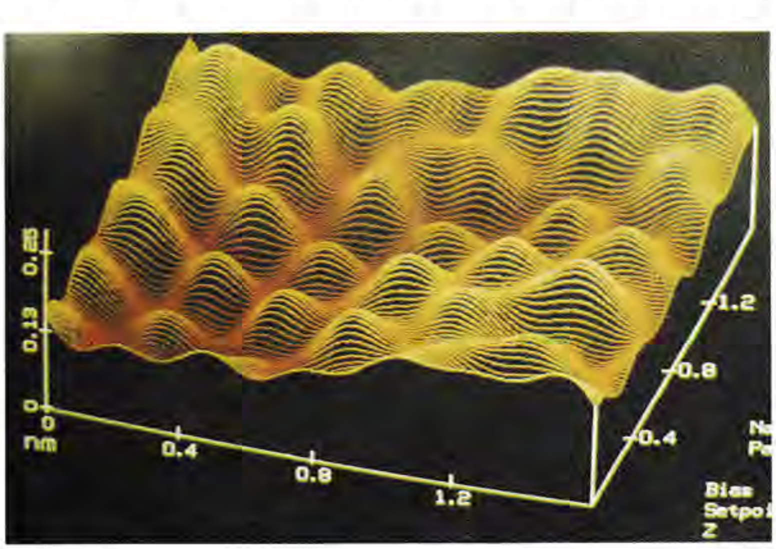


Figure 5
"Topographic map" of the surface of a molybdenum disulfide crystal. This image was taken in a student lab at Montana State University by H. David Sheets and Darryl L. Steinert.

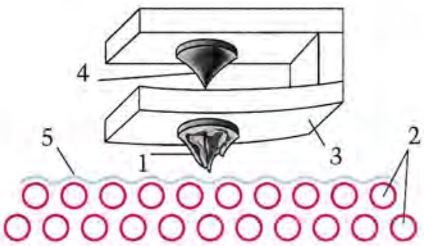


Figure 6
Probe of the atomic force microscope. 1—diamond grain; 2—atoms; 3—spring strip; 4—point; 5—electron clouds.

collected by the point of the tunneling microscope. To prevent displacement of atoms on the surface as a result of thermal motion, the experiment was conducted at a very low temperature (–269°C). This was just a publicity stunt, of course, but it demonstrates the development of nanotechnology—the ability to construct artificial structures in the microcosm that will become the basis of fantastically small electronic devices.

Feeling atomic repulsion

Although the scanning tunneling microscope has many attractive features, it still suffers from one very serious defect: it can be used only with electrically conductive materials. But most materials are covered with an insulating layer of oxides. Biological objects that are of interest to researchers likewise conduct electricity rather poorly. Can the scanning tunneling microscope feel "nonconducting" atoms? Yes, it can. One need only place a diamond grain, attached to a thin metal strip, between the point of the microscope and the surface of the object under investigation (fig. 6). The sharp point of the grain will be repulsed by the electron clouds of the surface atoms. Of course, the grain must be brought so close that the electron clouds at its tip and those of the surface overlap. The metal strip acts as a spring, pressing the grain against the surface.

Now let's scan the surface with this piggyback probe. The grain will move up and down to follow the unevenness of the surface. Its movement will be recorded by the change in the tunneling current flowing from the point on the metal strip. This microscope is called an atomic force microscope (AFM), and it has a feedback mechanism much like the one in the scanning tunneling microscope. By moving this probe vertically, the system maintains a constant separation between the strip and the grain (and therefore keeps the repulsive force constant as well).

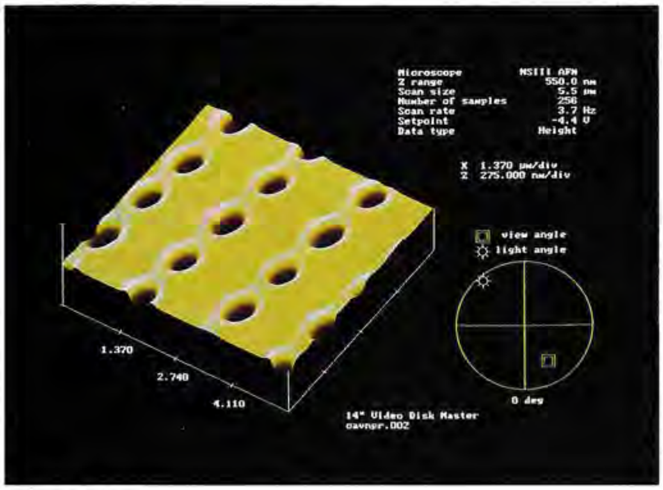
The atomic force microscope also has extremely high, atomic resolution. Figure 7 gives you the chance to take a close look at a common medium for storing information. Digital data are recorded on a video disk (as on an audio CD) by means of tiny holes on its surface.

Bits under a microscope

You undoubtedly know what a "bit" of information is. When information is recorded by a computer on a disk, the surface layer of the disk contains magnetized and unmagnetized regions of micron and submicron size. But you can't see anything, no matter which microscope you use.

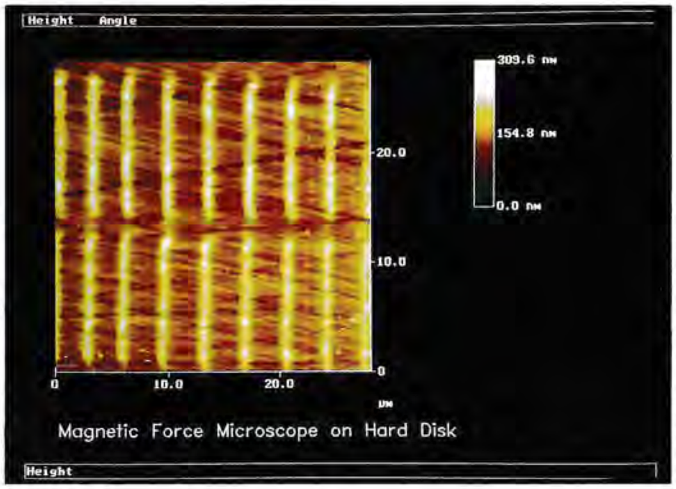
Well, a new instrument has appeared—the magnetic force microscope—that can "touch" these magnetic regions on the disk. It's easy to turn the atomic force microscope into a magnetic force microscope—just replace the diamond grain with a grain of magnetic material (iron or nickel). The magnetic grain will feel the influence of the fields of the magnetized regions. By scanning the surface of the disk, we can obtain a map of the distribution of magnetic forces—bytes and bits of information made visible (fig. 8).

So now you've been introduced to the large and powerful family of scanning probe microscopes. In spite of their relative youth, they are already capable of some pretty amazing feats. As they "grow up," no doubt they'll open up the microcosm of atoms and molecules for us. But they'll also give us the opportunity to work in that miniature world—to change it and use it in a wide variety of applications.



Courtesy of Digital Instruments, Inc.

Figure 7
AFM image of a video disk master.



Courtesy of Digital Instruments, Inc.

Figure 8
Image of a hard disk taken by a magnetic force microscope.

MATH INVESTIGATIONS

The Worm Problem of Leo Moser

Part I: This just about covers it!

by George Berzsenyi

LEO MOSER WAS ONE OF THE most imaginative and creative of mathematicians, and his wonderful problems still intrigue many of us long after his untimely death. One of these is known as his Worm Problem. It appeared in 1966 as problem 9 in a set of mimeographed noted entitled "Poorly Formulated Unsolved Problems in Combinatorial Geometry," a collection of fifty well-thought-out challenges to fellow mathematicians. In this problem he asked: What is the region of smallest area that will accommodate every arc of length 1? More precisely, the region is usually restricted to a convex region (in which every segment connecting a pair of points in the region is in the region), and the arc is understood to be planar. My preferred formulation of Moser's problem is: **Find the area of the smallest convex blanket that will cover every worm of unit length.**

This famous and elusive problem remains unsolved, in spite of many excellent efforts during the last 25 years. First it was shown that a circular disk of diameter 1 will cover the worm; to see this, place the center of the circle at the midpoint of the worm. Then it was shown that the square of diagonal 1 will also cover the worm; the proof of this fact is a bit more difficult. Next it was shown that a semidisk of diameter 1 will also do the job; this accomplishment also called for some very clever reasoning. My next challenge to my readers is: **Verify the above claims independently.** You will thus experience the first few steps of the initial progress made on the problem.

In part II of this account, I'll describe five more steps of the progress, including the latest one, which reduced the area of the minimum region to 0.27524. This is the record to date, and I'm particularly happy that two of my longtime friends and former colleagues were instrumental in achieving this result. I'll credit them more properly in part II. In part III, I'll offer you some of their conjectures. For the present, I'm being purposefully vague, since I don't want to spoil your fun in making your own discoveries. It's often best not to be aware of the methods of attack attempted by others, but to explore uncharted territories on your own. You should be encouraged by the fact that most of the results obtained thus far required no "fancy tools of mathematics," merely a bit of geometry, some trigonometry, and a healthy dose of careful reasoning.

In closing, I wish to warn my readers that the area of the minimal cover cannot be reduced below 0.21946. (This lower bound was established several years ago.) Thus, any blanket of a smaller area will necessarily leave part of some worm exposed to the elements. We wouldn't want our worm to catch a cold! ◘

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

Art by Sergey Ivanov

PHYSICS CONTEST

Row, row, row your boat

"It is not possible to step twice into the same river."—Heraclitus

by Arthur Eisenkraft and Larry D. Kirkpatrick

ONE OF THE EARLY ARGUMENTS against a spinning Earth held that objects would not fall straight down. Given that we now know (with the help of Eratosthenes) that our home planet has a diameter of 6,500 km (4,000 miles), then parts of our planet must be moving at 1,700 km/hr (1,000 mph). If the critics had been correct and you drop an object that takes 0.5 s to reach the ground, the object would land 240 m behind you. Along comes Galileo to refute what appears to be common sense.

Galileo proposes that a person climb the mast of a ship. If the ship is not moving and a ball is dropped, it will certainly fall straight down. The defenders of the stationary Earth would then predict that if the ship were moving, the ball dropped from the mast would land toward the rear of the boat. This is because, they would say, the boat glides forward while the ball is descending. Galileo suggested the correct behavior. The ball maintains the original horizontal motion of the ship and lands in the identical location as when the ship stood still.

If this works for a ship, it should also work for the Earth. The vertical motion is independent of the horizontal motion. A ball on a stationary ship or a gliding ship will land in the same place whether the ship is moving or stationary. This may seem obvious to some of our readers, but it is quite subtle and still confounds many people.

Imagine hopping aboard a rowboat and paddling from one shore to the opposite shore with no current. The trip takes you 15 minutes. If you return to the river and venture across again, paddling to the opposite shore with the same strokes, but with a stiff current dragging you downstream, will you arrive at the opposite shore in less time, in more time, or in the same time? You probably recognize that you would land further downstream on this second journey. Since you traveled further, maybe it should take more time. But your velocity is actually the sum (the vector sum!) of your paddling velocity and the velocity of the current. With this faster speed, maybe the journey should take less time. Or perhaps, the longer distance is exactly compensated by the greater speed and you arrive at the opposite shore in the same time. Our readers can use the fact that the motion across the river and the motion downstream are perpendicular to each other and are therefore unaffected by each other. The time is determined by the motion across the river independent of the speed of the current. The current determines where the boat lands downstream, but does not change the time.

Once again, this is quite subtle, and our readers should attempt to explain the solution of this puzzle to people not accustomed to thinking the way physicists do. If you can convince someone of this, then you, as a teacher, must really understand it.

Let's complicate the situation. What happens if you don't paddle straight across the river, but rather choose to paddle at some angle? Now you'll find that there is a component of your velocity that helps you across the river and a component that takes you upstream or downstream. In this way, you can head upstream and end your journey directly across from where you embarked.

So here is our contest problem for this issue. Assume that you wish to end up directly across the river and that you were permitted to walk on the far shore if you land upstream or downstream. What path takes the least time? Let's add some specific numbers (suggested by Resnick and Halliday in their *Fundamentals of Physics*): the river is 500 m wide; your rowing speed is 3,000 m/h; the river flows at 2,000 m/h; and your walking speed on the opposite shore is 5,000 m/h.

A. Solve for the possible range of angles qualitatively.

B. Describe the path quantitatively.

C. Calculate this minimum length of time.

". . . Merrily, merrily, merrily, merrily, life is but a dream."

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

Art by Tomas Bunk


A
B
E=mc²
Tom Bunk

Sources, sinks, and gaussian spheres

A completely correct solution to the contest problem in the July/August issue was submitted by Eric Joanis of Gatineau, Quebec.

A. Because we have a spherically symmetric distribution of charge, we choose a spherical gaussian surface of radius r centered on the spherical charge distribution. We can then use the result given in the statement of the problem:

$$E4\pi r^2 = \frac{q_{enc}}{\varepsilon_0}, \qquad (1)$$

where

$$q_{enc} = \rho V = \rho\frac{4}{3}\pi a^3. \qquad (2)$$

Combining these two equations and solving for the electric field E, we get

$$E_o = \frac{a^3\rho}{3\varepsilon_0 r^2}, \qquad (3)$$

where we have added the subscript "o" to indicate that this is the value outside the sphere—that is, for $r \geq a$.

We can show that this has the same form as Coulomb's law by substituting the value of ρ from equation (2) into equation (3).

B. We can use the same technique to find the electric field E_i inside the sphere ($r \leq a$) if we remember that only the charge *inside* the gaussian sphere contributes to the field. Therefore, the enclosed charge is given by

$$q_{enc} = \rho\frac{4}{3}\pi r^3,$$

where we use the radius r of the gaussian sphere instead of the radius a of the complete sphere as before. Therefore,

$$E_i = \frac{r\rho}{3\varepsilon_0}. \qquad (4)$$

At the surface of the sphere $r = a$ and

$$E_o = E_i = \frac{a\rho}{3\varepsilon_0},$$

as expected.

C. The only complication caused by the spherical hole at the center of the sphere is in the calculation of the enclosed charge. We can calculate this charge by taking the charge of the complete sphere and subtracting off the contribution due to the hole. There are three regions. Outside the sphere ($r \geq a$) we have

$$q_{enc} = \frac{4}{3}\pi a^3\rho - \frac{4}{3}\pi b^3\rho = \frac{4}{3}\pi\rho(a^3 - b^3),$$

with the resulting electric field

$$E = \frac{\rho(a^3 - b^3)}{3\varepsilon_0 r^2}.$$

In the spherical shell ($b \leq r \leq a$) we have

$$q_{enc} = \frac{4}{3}\pi r^3\rho - \frac{4}{3}\pi b^3\rho = \frac{4}{3}\pi\rho(r^3 - b^3).$$

Inside the spherical hole ($r \leq a$) $E = 0$, since there is no enclosed charge.

D. When the hole is moved off-center, we must be careful to remember the vector nature of electric fields. Inside the bigger sphere (assumed to be completely filled with charge ρ), the electric field E_B is given by equation (4), but now we write it in vector form:

$$E_B = \frac{\rho}{3\varepsilon_0}\mathbf{r},$$

where $\mathbf{r}$ is a radial vector from the origin to the point of interest. We now express this in terms of rectangular coordinates:

$$E_B = \frac{\rho}{3\varepsilon_0}(x, y, z).$$

We can do the same thing for the smaller sphere of negative charge, but we must remember that the center of this sphere has been shifted to $x = c$:

$$E_S = \frac{-\rho}{3\varepsilon_0}(x - c, y, z).$$

When we add the two contributions to the field, we find that the y- and z-components cancel, and we are left with a constant x-component:

$$E = E_B + E_S = \frac{\rho}{3\varepsilon_0}(c, 0, 0).$$

Notice the surprising result that the electric field inside the hole has a constant value independent of the size of the hole, the size of the larger sphere, and the location within the hole. It depends only on the amount of offset and the charge density. ◘

THINK	Think about a stimulating change of pace. Independent schools offer competitive salaries, small classes, college prep curriculum. Each year over 400 private schools nationwide list positions seeking qualified science and math teachers.
INDEPENDENTLY	IES, a non profit placement service that works exclusively with private schools, will refer you to positions that match your experience and career needs. No placement fee to candidates. Positions available K-12 in chemistry, biology, physics, IPS, environmental and earth sciences and math. Call Carol Compton at 1 800 257-5102 for more info.

Circle No. 14 on Reader Service Card

ALISTON THOMAS
B.S. in Chemistry; Ph.D. Graduate Fellow and Teaching Assistantship in Chemistry at the University of Illinois, Urbana-Champaign; summer internship Pralex Corporation, Christiansted, St. Croix, USVI.

Join the *SUPERACHIEVERS* in *SCIENCE* and *MATH* at CLARKE COLLEGE

DUBUQUE, IOWA

DEANNA ANGST
B.S. in Chemistry and Mathematics; summer internship for ICI Americas, Brussels, Belgium; Teaching Assistantship & Emily Berger Fellowship for Ph.D. in Chemistry, University of Kansas, Lawrence.

PROFESSIONAL & GRADUATE OPPORTUNITIES

✓ After freshman year, students can enhance academic programs each summer with professional paid work experiences for credit in companies and research institutions throughout the United States and in Europe. Recent opportunities have included Proctor and Gamble, Cincinnati, OH; Monsanto Agriculture New Products Division, St. Louis, MO; Owens Corning Fiberglass, Granville, OH; Michael Reese Medical Research Center, Chicago, IL; Brookhaven National Laboratory, Upton, NY; Argonne National Laboratory, Argonne, IL; Meharry Medical College, Nashville, TN.

✓ 95% of Clarke graduates are employed or in graduate school within six months of graduation.

Celebrating 150 years as a four-year liberal arts college.

STATE OF THE ART EQUIPMENT AND INSTRUCTIONAL MATERIALS

✓ Selected for National Science Foundation Grant for **Multimedia Electronic Classroom Laboratory** for Math, Biology and Computer Science.

✓ Awarded National Science Foundation Grant for the acquisition of **Nuclear Magnetic Resonance** (NMR) instrument to be used in chemical structural analysis.

✓ Computers incorporated into hands-on investigative approach to all science and math courses.

✓ ***Mathematica***, a graphics/symbolic computation program, used in calculus and other math classes

✓ Designated as an **IBM test site** for evaluating prototype hardware and software. Faculty and students involved in research which integrates new computer technology into chemistry analysis.

QUALITY ACADEMIC ENVIRONMENT

✓ Clarke's **12:1 teacher ratio** guarantees attention from faculty who will challenge you to achieve.

✓ Academic excellence rewarded with renewable non-need **scholarships**.

✓ Supports student chapters of the American Chemical Society (**ACS**); Mathematics Association of America (**MAA**); and Association for Computing Machinery (**ACM**) where students compete in national computing programming contest.

✓ Offers a chapter of National Computer Honor Society, **Upsilon Pi Epsilon.**

✓ **Strong undergraduate research programs** in biology developed for on-site study.

Located in scenic northeast Iowa on the bluffs of the Mississippi River bordering Illinois and Wisconsin.

UNDERGRADUATE SCIENCE MAJORS & PROGRAMS:

Biology/Molecular Biology
Chemistry
Computer Information Science/Management Information Science
Computer Sciences
Education
Health:
- Medical Technology
- Pre-Chiropractic
- Pre-Dentistry
- Pre-Medicine
- Pre-Pharmacy
- Pre-Veterinary

Mathematics
Nursing
Physical Therapy
3-2 Engineering

✓ Recent graduates offered opportunities for **totally funded** teaching fellowships and stipends for graduate studies at universities throughout the country including University of Arizona, University of Iowa, University of Minnesota, University of Notre Dame, University of Oklahoma, University of Osteopathic Medicine and Health Science, Virginia Polytechnic Institute and State University, Washington State University

For more information, call 319-588-6316 OR toll-free 1-800-383-2345 (1-800-255-2255 ext. 6316 U.S. Virgin Islands & Puerto Rico)

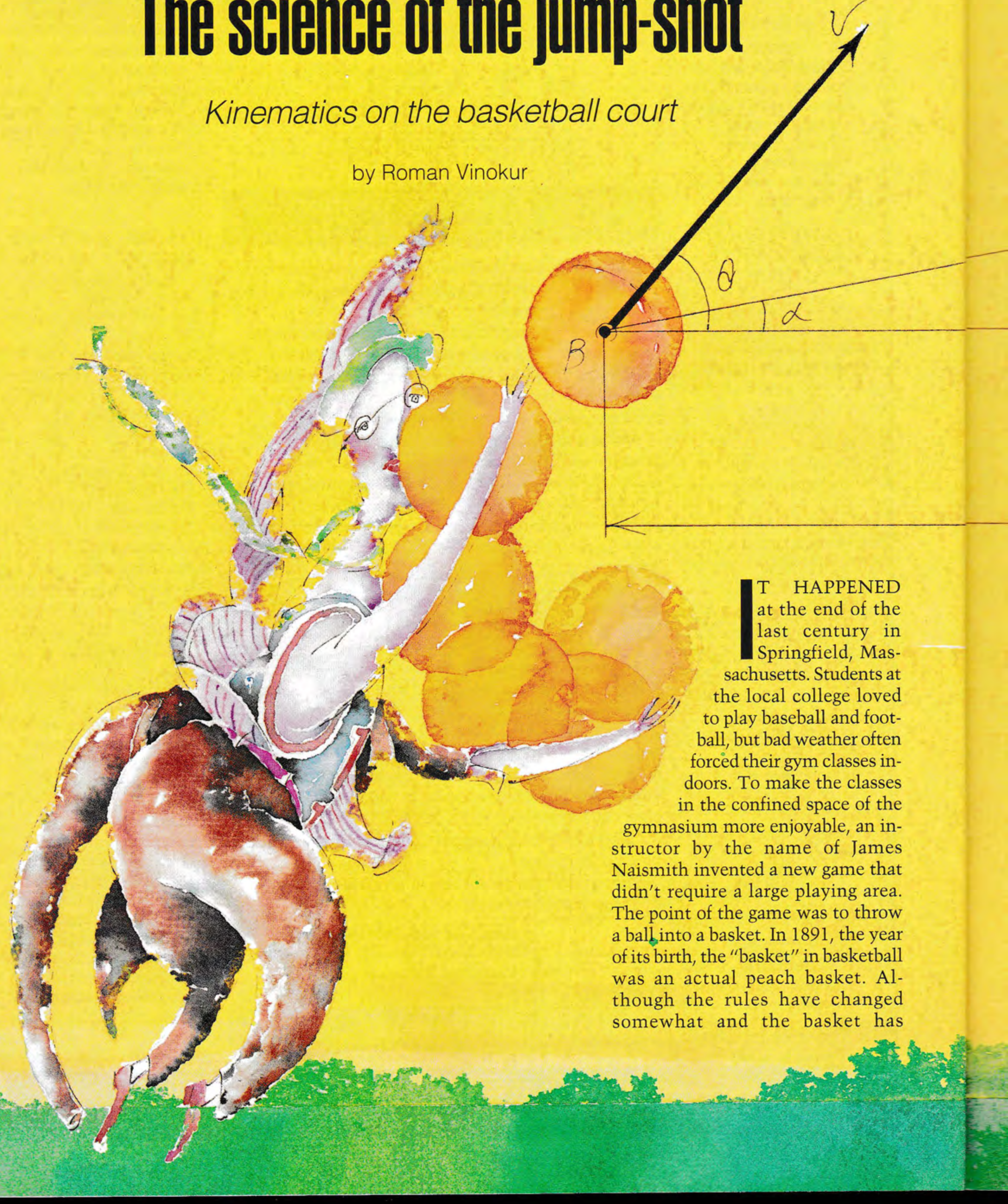
Circle No. 7 on Reader Service Card

AT THE
BLACKBOARD II

The science of the jump-shot

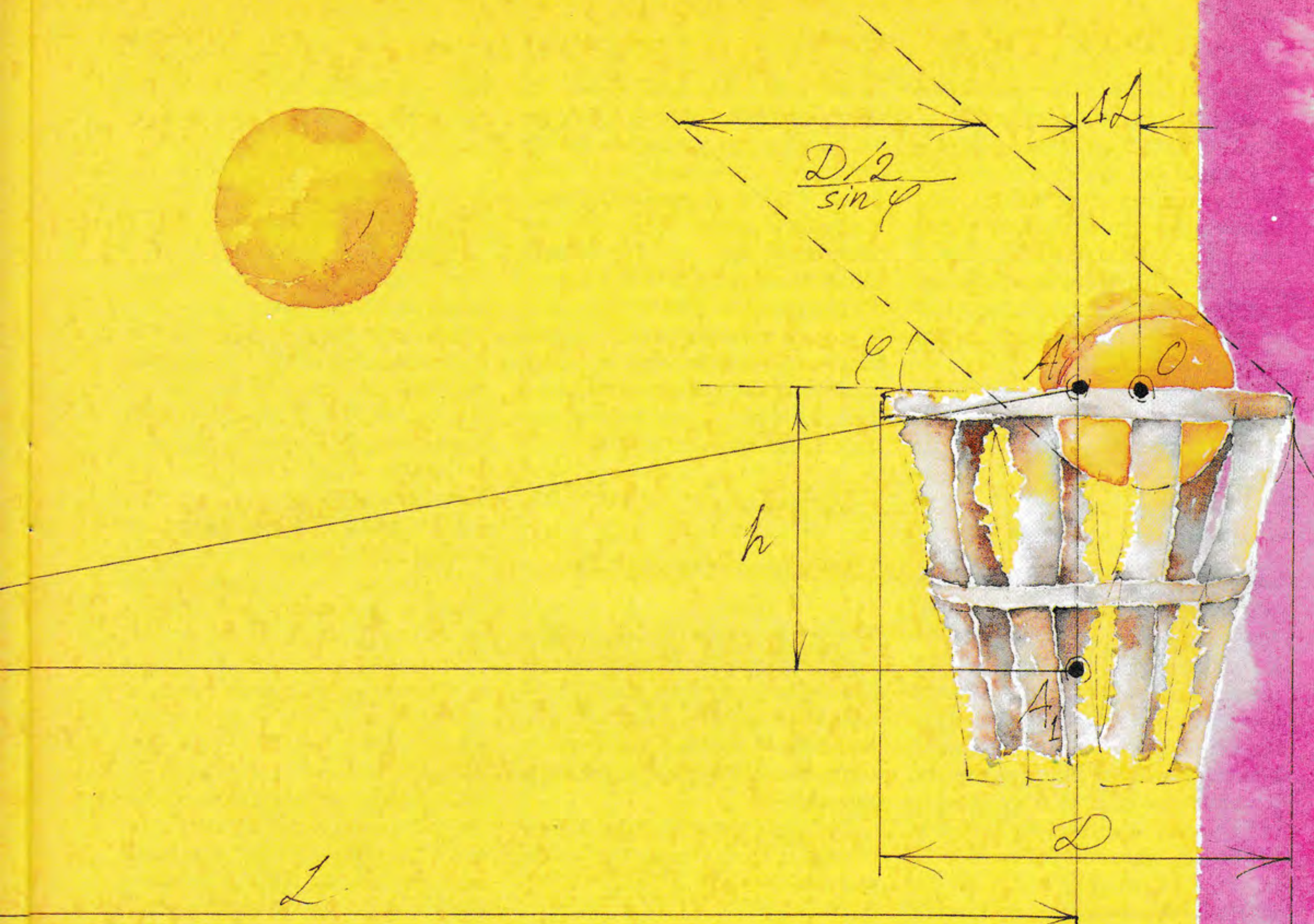
Kinematics on the basketball court

by Roman Vinokur



Art by Sergey Ivanov

IT HAPPENED at the end of the last century in Springfield, Massachusetts. Students at the local college loved to play baseball and football, but bad weather often forced their gym classes indoors. To make the classes in the confined space of the gymnasium more enjoyable, an instructor by the name of James Naismith invented a new game that didn't require a large playing area. The point of the game was to throw a ball into a basket. In 1891, the year of its birth, the "basket" in basketball was an actual peach basket. Although the rules have changed somewhat and the basket has



evolved into an abstraction of its former self, the basic thrust of basketball remains unchanged: tossing the ball into the basket.

In addition to basketball games themselves, another form of competition became popular in the US: shooting contests. In 1977 Ted Martin made 2,036 free throws in a row. A year later, Fred Newman made only 88 straight—with his eyes closed! Other countries can boast of hot shooters too: Drazan Petrovic of Yugoslavia and the Muscovite Sergey Belov display a nice touch with the basketball.

If you want to be a good shooter, inborn talent is obviously important. But you also have to work at the technique of shooting. For example, Drazan Petrovic would go to the gym every morning and take 500 shots at the hoop from various spots on the court. Practice makes perfect, as they say, but a grasp of theory doesn't hurt either. Let's see if we can work out a simple theory of accurate basketball shooting.

The main phases of the ball's movement when it is tossed with one hand are shown in the picture. We need to find the best angle θ with the horizontal that produces the most accurate shot.

We'll restrict ourselves to an approximation based on elementary physics and mathematics. The height of the hoop above the floor is H = 3.05 m (or 10 feet—the same height at which Naismith nailed the peach basket in Springfield). The inner diameter of the hoop D = 0.45 m, and the diameter of the basketball is about half that. Suppose the ball enters the hoop at an angle ϕ with the horizontal and the trajectory of the ball's center lies in the vertical plane passing through the center of the hoop. (We'll ignore possible deviations of the ball from this plane, assuming that the reason for a bad shot is that it's either overthrown or underthrown.)

The condition under which the ball passes through the hoop without touching it can be written as

$$|\Delta L| \le l = \frac{D}{2}\left(1 - \frac{1}{2\sin\phi}\right), \qquad (1)$$

where ΔL is the deviation of the ball's center O from the hoop's center A. This condition makes sense if $\phi > 30°$. If $\phi < 30°$, the ball will definitely hit the rim and—what usually happens (especially when the ball's velocity is large enough when it hits the rim)—bounce off without touching the net. If we increase the angle ϕ, we increase our chances of getting the ball into the basket, since l increases as well. So, if $\phi = 40°$, then $l \cong 0.05$ m; at $\phi = 60°$, $l \cong 0.095$ m—almost double the first. The maximum value of $l \cong 0.112$ m (for $\phi = 90°$).

It's obvious that the angle ϕ increases as the angle θ (the angle at which the player tosses the ball at the basket) grows steeper. But if you try to shoot at an extremely steep angle ($\theta \ge 70°$), it's rather difficult to put the ball through the hoop—that is, if you're any distance from the basket. Sometimes it's hard not only to make a basket but just to get the ball to the rim—it takes a great deal of effort. When basketball players shoot at steep angles, it's not because they like to—it's because of the long arms of the opposing players.

To simplify things, though, we can neglect such obstacles to a basketball shot. This usually happens when a player is far from the basket and, thus, far from the defending players. And let's also neglect the air resistance at this stage of our analysis, even though its influence (unlike that of opposing players) increases with distance from the basket.

Let the basketball leave the player's hand at time t_0, when the ball's center is at point B. It reaches the center of the hoop (point A) in time t. The initial speed of the ball is V and the distance of the shot (the projection of segment BA on the horizontal plane) is equal to L. Here we are examining a "clean" shot—that is, one that is not banked off the backboard. We can write the kinematic equations describing the movement of a material point thrown at an angle θ with the horizontal:

$$L = (V\cos\theta)t,$$

$$h = L\tan\alpha = (V\sin\theta)t - \frac{gt^2}{2}.$$

Here $g \cong 9.81$ m/s^2 is the acceleration due to gravity, θ is the value in question (the angle at which the ball leaves the player's hand), and h is the height of the hoop's center relative to the initial point of the ball's trajectory (B).

The angle ϕ at which the ball enters the hoop is determined by the equation

$$\tan\phi = \frac{|V_{yA}|}{V_{xA}} = \frac{|V\sin\theta - gt|}{V\cos\theta},$$

where V_{xA} and V_{yA} are the horizontal and vertical components, respectively, of the ball's velocity at point A. The kinematic equations can be transformed (by eliminating the unknown t) into

$$L = \frac{V^2}{g}\frac{\sin(2\theta - \alpha) - \sin\alpha}{\cos\alpha}, \qquad (2)$$

$$\phi = \arctan(\tan\theta - 2\tan\alpha). \qquad (3)$$

Let's analyze these equations to arrive at some useful conclusions. It follows[1] from equation (2) that when $2\theta - \alpha = 90°$, the distance for a good shot is achieved with the minimum velocity and, consequently, with the least expenditure of energy. So the optimal initial angle for the shot is

$$\theta = \theta_{opt} = 45° + \frac{\alpha}{2}. \qquad (4)$$

The angle $\alpha = \arctan(h/L)$ is shown in the picture and depends on the player's height and, if the player jumps while shooting, the height of the jump. The angle α also depends on L. In the case of a long shot ($h << L$), this angle is pretty small, and, in accordance with equation (4), the optimal initial angle of the shot equals about 45° (or perhaps a few degrees more).

Let's check this conclusion, assuming the ball is at a height of 2 m in the initial phase of the shot (which would be the case if you're not a giant and you shoot without jumping). Then $h = 3.05 - 2.0 = 1.05$ m. Taking the distance $L = 6$ m, we get $\alpha = \arctan(1.05/6) \cong 10°$—that is (see equation (4)), $\theta_{opt} = 45° + 5° = 50°$. Substituting this result in equation (2), we calculate that the initial speed of the ball should (in our case) be equal to 8.37 m/s. If you're able to run this fast, you'd run the 100-meter dash in 11.9 seconds—not bad for the average athlete, but far short of the world record. Unlike track-and-field events, though, high speeds aren't necessary here: the greater the ball's initial speed, the greater its speed when it enters the basket and the more likely it is to bounce off if it touches the rim.

It's interesting that when the initial angle is optimal, the initial velocity of the ball is minimal. It's a nice discovery—you could find it by analyzing equation (2). But the main advantage of using the optimal angle is a surprising phenomenon: the length of the shot depends hardly at all on slight deviations of the initial angle from the optimal value! This result is especially useful because unfortunately (or fortunately), we aren't robots—we make mistakes, and not just during basketball games.

Let's look at a shot at angle θ (not necessarily optimal) and initial velocity V that goes right through the middle of the hoop. If the shot is made at an angle $\theta + \Delta\theta$, where $\Delta\theta$ is a slight deviation, the ball's center will cross

[1]To show this, solve the equation for V^2 and set the first derivative with respect to θ equal to zero, or use graphical techniques.—*Ed.*

the horizontal plane that includes the hoop not at the hoop's center but somewhere else. So the length of segment BA_1 will be equal to $L + \Delta L$, where ΔL is small. Using equation (2) (see the appendix, part 1), we get

$$\Delta L = L \cdot 2\frac{\Delta\theta \cdot \cos(2\theta-\alpha)-(\Delta\theta)^2\sin(2\theta-\alpha)}{\sin(2\theta-\alpha)-\sin\alpha}. \quad (5)$$

(Here $\Delta\theta$ is expressed in radians.)

It follows from equation (5) that the error caused by the angular deviation is proportional to the length of the shot. This isn't surprising. But here's an interesting result. If we ignore the term in equation (5) that includes the very small value $(\Delta\theta)^2$, then if $\theta = \theta_{opt}$, we come upon a paradox (at first glance): $\Delta L = 0$. It's amazing, really: shooting at an angle θ_{opt}, you can't miss—as long as you judge the initial velocity correctly. But if we do not neglect any part of equation (5), we get

$$\Delta L = -\frac{2L(\Delta\theta)^2}{1-\sin\alpha}. \quad (6)$$

Thus, ΔL depends on the deviation $\Delta\theta$, albeit insignificantly. So the main problem is to limit the deviation $\Delta\theta$. How much deviation is acceptable? Let's estimate it for the special case $\alpha = 0°$, $L = 6$ m. From equation (4), it follows that the optimal angle is 45°. Using equation (3), we get $\phi = 45°$ as well. Substituting the necessary data in equations (1) and (6), we get, finally, $|\Delta\theta| \le 4.2°$.[2]

Is this possible? Experience says that it is, but let's check it. We won't even need a basketball.

Mark two points B and A on a sheet of paper. Put the tip of a ballpoint pen on point B and, keeping your eye on point A, quickly draw a line connecting these points. The resulting line 1 (fig. 1) isn't an ideal straight line, of course, and maybe doesn't even pass through point A. Now take a straightedge and draw the straight lines BA and BC, which approximately coincide with line 1 where it begins. Measure the angle ABC with a protractor. It probably won't be more than 3° to 4°. With a little practice, you can get even better results.

Figure 1

Although this little experiment has no direct relation to basketball, it nevertheless characterizes the precision of your hand's movement in a visually assigned direction. It's not the same as shooting a basketball, but it's encouraging.

Let's get back to our formulas and find the angle ϕ at which the ball enters the hoop if it is tossed at the optimal angle $\theta = \theta_{opt} = 45° + \alpha/2$. After some straightforward trigonometric transformations (appendix, part 2), using equations (2) and (4), we get

$$\phi(\theta_{opt}) = 45° - \frac{\alpha}{2}.$$

This means that for $\alpha > 30°$, a shot at the optimal angle will not be accurate, because in this case $\phi < 30°$ and the ball won't be able to go through the hoop cleanly. Even if $\alpha \cong 20°$ (that is, $\phi = 35°$), our chances aren't great because l will be relatively small (see equation (1)).

Similar situations (characterized by a steep angle α) arise when one is shooting from a short distance (not more than 2 m). What should we do? Since it's easier to make a basket from close in, it's not so important to achieve the optimal angle. We can shoot at steeper angles. And we can improve the shooting conditions by jumping: not only do we move the ball's center closer to the hoop's level (that is, you diminish the angle α), we also get up over the defenders.

Another way to decrease the angle for short shots is to artificially increase the length of the shot. You do this by banking the ball off the backboard. In figure 2, you can see the projection of the ball's trajectory on the horizontal plane (for a perfectly elastic collision). The ball is tossed at the hoop's mirror image. It's clear that $|BA| < |BA'|$. Naturally the difference between $|BA'|$ and $|BA|$ is less the bigger $|BA|$ is. So bank shots are useful only from rather short distances.

In all our calculations we have neglected air resistance. But we all know how significant this factor is in many physical phenomena. For instance, a rifle bullet would travel ten times farther if there were no atmosphere. The air resistance depends on the shape of the moving body (think of the streamlined shapes of racing cars and airplanes), on the cross section perpendicular to the direction of travel (which is why a bicyclist bends over while racing), and especially on the body's velocity. You've probably noticed that the air changes the trajectory of tennis balls, soccer balls, and so on. A basketball is influenced by air to a lesser degree, primarily because its motion is much slower. A ball acquires a higher initial velocity when hit or kicked than when pushed by the fingers. So tennis balls travel as fast as 30 m/s, while a basketball usually moves no faster than 10 m/s.

To account for air resistance in a basketball shot, we'd need to resort to differential equations. Let's just look at some results of such an analysis. Although air resistance shortens the distance a ball travels, its influence on the optimal angle of a shot is insignificant in actuality. For example, the optimal angle calculated according to equation (4) for a shot from 6–7 m away is only about 2–3° less.

The movement of a ball in principle also depends on its rotation

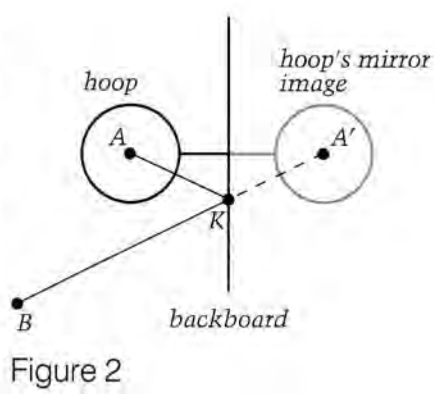


Figure 2

[2]The result in radians was translated into degrees by multiplying by $180°/\pi$.

about its center. The adjacent layers of air become involved in this rotation; this combines with the head-on movement of the air against the ball to produce a force acting on the ball perpendicular to its trajectory. Because a well-executed shot imparts a slight reverse rotation to the ball, this force produces a slight "lift" that may partially compensate the effect of air resistance. At any rate, the role of a ball's rotation is much more pronounced in tennis, baseball, golf, and other sports in which balls of lower mass are propelled at much higher speeds.

I hope this article has convinced you that it's not so hard to mix business with pleasure, science with sports, study with games.

Appendix

1. The change in the length of the shot in the situation mentioned in the article is

$$\begin{aligned}\Delta L &= L' - L\\ &= \frac{V^2}{g}\frac{\sin(2\theta' - \alpha) - \sin\alpha}{\cos\alpha} - \frac{V^2}{g}\frac{\sin(2\theta - \alpha) - \sin\alpha}{\cos\alpha}\\ &= L\frac{\sin(2\theta' - \alpha) - \sin(2\theta - \alpha)}{\sin(2\theta - \alpha) - \sin\alpha}.\end{aligned}$$

Transforming the difference of sines in the numerator of the resulting equation, we get

$$\begin{aligned}\sin(2\theta' - \alpha) - \sin(2\theta - \alpha) &= 2\cos(2\theta + \Delta\theta - \alpha)\sin(\Delta\theta)\\ &= 2[\cos(2\theta - \alpha)\cos(\Delta\theta)\\ &\quad - \sin(2\theta - \alpha)\sin(\Delta\theta)]\sin(\Delta\theta).\end{aligned}$$

If we express $\Delta\theta$ in radians, then the approximate equations $\cos(\Delta\theta) \cong 1$ and $\sin(\Delta\theta) \cong \Delta\theta$ are valid (since $\Delta\theta << 1$). Using them in the formulas derived above, we can obtain equation (5).

2. Substituting $\theta = \theta_{opt} = 45° + \alpha/2$ in equation (3), we get

$$\begin{aligned}\phi(\theta_{opt}) &= \arctan(\tan\theta_{opt} - 2\tan\alpha)\\ &= \arctan\left(\frac{1 + \tan(\alpha/2)}{1 - \tan(\alpha/2)} - \frac{4\tan(\alpha/2)}{1 - \tan^2(\alpha/2)}\right)\\ &= \arctan\left(\frac{1 - \tan(\alpha/2)}{1 + \tan(\alpha/2)}\right)\\ &= 45° - \frac{\alpha}{2}.\end{aligned}$$

Calling all modem maniacs!

What did you like in this issue of *Quantum*? If you find pen-and-paper communication too old-fashioned, you can send your comments, questions, and suggestions to the managing editor by electronic mail at the following address:

72030.3162@compuserve.com

We look forward to hearing from you.

SUMMER STUDY IN RUSSIA AND THE U.S.

A unique opportunity awaits advanced mathematics and science students and teachers interested in participating in a summer institute in the U. S. or Russia. The program will feature three weeks of **advanced classes** in mathematics, physics, and molecular biology taught by **prominent Russian and American professors**, trips to major **scientific laboratories**, as well as many **cultural and recreational activities**. Following the three-week study program, there will be a one-week cultural program in Washington, DC, or St. Petersburg.

The U.S. session will take place from June 27 to July 24, 1993, in Long Island, NY. The Russian portion of the program will take place from July 6 to August 3, 1993, at Moscow State University.

Scholarships Available!

For more information and application, please fill out the coupon below and mail to: Dr. Edward Lozansky, International Educational Network, 1800 Connecticut Avenue, NW, Washington, DC 20009, Phone: 202-362-7855, Fax: 202-364-0200.

Please send me _____ brochures to distribute among interested high school teachers and students.
Last Name______________________ First______________________
Address______________________ City______________________
State________Zip code______________ Phone number______________________
I am a teacher_______ or a student_______.

HAPPENINGS

US team places second at IMO

Moscow hosts an international celebration of math talent

by Cecil Rousseau and Daniel Ullman

THE 33RD INTERNATIONAL Mathematical Olympiad (IMO) was held last summer in Moscow, Russia, where six high school students representing the US performed with distinction and earned a combined team score second only to that of China. This year's IMO team from the US was composed of Wei-Hwa Huang of North Potomac, Maryland; Kiran Kedlaya of Silver Spring, Maryland; Robert (Bobby) Kleinberg of Wales Center, New York; Sergey Levin of Providence, Rhode Island; Lenhard (Lenny) Ng of Chapel Hill, North Carolina; and Andrew Schultz of Evanston, Illinois. Kiran, Bobby, and Lenny had all represented the US last year at the 32nd IMO in Sweden, all earning silver medals. This year they brought home gold. (Kiran, in fact, had earned a gold medal in 1990 at the 31st IMO in Beijing, China, before he was a sophomore in high school.) Wei-Hwa, Sergey, and Andrew all earned silver medals in Moscow. This ties the 1986 team for best American medal performance at an IMO.

Kiran and Sergey are currently freshmen at Harvard, while Andrew attends the University of Illinois. Wei-Hwa, Bobby, and Lenny are all seniors in high school, which means of course that they are eligible to represent the US next year at the 34th IMO in Istanbul, Turkey.

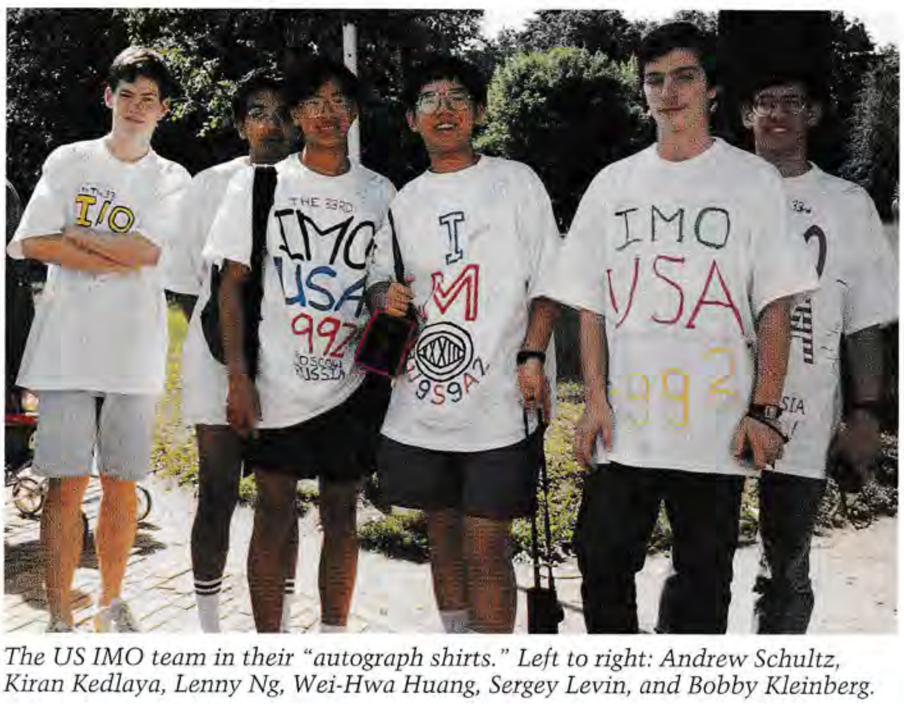
Historic and modern Moscow

The team arrived at its hotel, located about five miles from the Kremlin, on the evening of July 10. The days of July 11, 12, and 13 were devoted to tours of Moscow sites of historic interest, primarily the gardens, museums, and churches of the tsars. The opening ceremony of the IMO was on July 14. This is the forum for the official welcome by the Russian hosts, speeches by Russian mathematical dignitaries, and entertainment in the best Russian tradition. We enjoyed spectacular performances of folk singers and dancers, some ballet, and an act from the famed Moscow circus.

The organizers of this 33rd IMO overcame many obstacles. It was four years ago that the 1992 IMO was officially slated for Moscow, but it was unclear for most of these four years exactly which country Moscow would be in. The collapse of the Russian economy forced other constraints on the organizers. But they beat the odds and the event transpired, owing to the volunteer efforts of dozens of Russian mathematicians, most of whom themselves grew up supported by a network of local, regional, and national olympiads that is the envy of the rest of the world. In Russia there have been olympiads for all age groups at all levels for many years. This has undoubtedly been a major factor in the development of the world-renowned Russian community of research mathematicians.

The competitive portion of the olympiad took place on July 15 and 16. On each day the participants had 4½ hours to do three problems.

On the evening following the first



The US IMO team in their "autograph shirts." Left to right: Andrew Schultz, Kiran Kedlaya, Lenny Ng, Wei-Hwa Huang, Sergey Levin, and Bobby Kleinberg.

day of the exam, the US team couldn't hide its disappointment. The exam looked relatively easy, and discussion with members of other teams suggested that, although the American team had done well, their total team score would be lower than that of the teams from the Commonwealth of Independent States (CIS), China, Germany, Romania, Russia, the United Kingdom, and perhaps others. The US team has competed for 18 years at the IMO, only once finishing below fifth place. This fact was not lost on the team.

US team turns it around

The second day of the exam boosted the US team's morale. Two of the problems on that day were especially tough, yet the Americans did almost as well as they had on the first day. Other teams had faltered. And after all the problems were scored, the US team slipped into second place, ahead of Romania, the CIS, and the UK but still far behind China.

This was a real cause for celebration. The team earned all gold and silver medals, a consistent performance that had not been accomplished by a US team since 1986. (Gold medals are awarded to the top 1/12 of the participants, silver medals to the next 1/6). There was plenty of rejoicing. Amidst the frequent reports that American students are far behind their international peers in mathematics, it certainly is nice to learn that our best students are as good as those from anywhere in the world.

It's a common American bias to think that talents of whatever type are innate rather than acquired. This undervalues the effort of those who excel and unnecessarily discourages those who falter. Mathematics is learned. Those who learn a lot may be able to win competitions and may be able to make major contributions to the discipline. It's not a matter of luck but rather a matter of hard work and dedication. Students should recognize that what it takes to excel in mathematics is not special mental faculties but motivation and support.

On the days of July 17 and 18, with the pressure of the exam off their shoulders, the team enjoyed a boat ride and some tours while the team leaders and official coordinators began the arduous task of scoring the papers.

The closing ceremony took place in the afternoon of July 20. The organizing committee awarded the medals. Numerous dignitaries spoke about international friendship and cooperation and using mathematics to solve the problems of the world. There was more traditional entertainment. And early on the morning of July 21, the team headed home.

The road to IMO

The six students who represented the US in Moscow earned a place on the IMO team by high achievement on a series of mathematical competitions, beginning with the American High School Mathematics Exam (AHSME), a 30-question, multiple-choice contest conducted in February 1992. Over 350,000 students participated in this event. Wei-Hwa, Bobby, and Lenny all wrote perfect papers on the AHSME, as did Ren Shi of Los Angeles, California.

Participants whose score on the AHSME exam exceeded 100 (out of a possible 150) were then invited to take the American Invitational Mathematics Exam (AIME), a 15-question, three-hour exam administered in March 1992, each of whose answers is an integer from 0 to 999. This year, 4,669 students qualified to take the AIME.

From these competitors, 138 of the highest scorers were invited to participate in the USA Mathematical Olympiad (USAMO), a five-question, 3½-hour contest administered in April, whose problems require written answers containing complete justifications. From this pool of high-achieving students, 25 were selected to take part in a monthlong training session, the Mathematical Olympiad Program (MOP), held at the US Naval Academy in Annapolis, Maryland.

Enrichment rather than acceleration

The 25 students chosen for the MOP were kept busy. They were bombarded with four hours of class per weekday and three or four Olympiad-style tests per week. In addition, "team contests"—a MOP tradition involving the presentation of solutions at the blackboard—were scheduled once or twice a week. One afternoon at the MOP was devoted to solving problems from the latest issue of the *American Mathematical Monthly*. Twice students were asked to compose their own original problems. In between these mathematical activities, soccer, frisbee, chess, and bridge were popular pastimes.

All told, the four weeks are intended to enrich rather than accelerate the students. The focus is on enjoying mathematics rather than training in test-taking skills. The primary goal at the MOP is not victory at the IMO but promoting interest in mathematics.

The first three afternoons at the MOP were devoted to three qualifying tests which were used, together with the USAMO score, to select the six members of the IMO team. This year, as last, every one of the team members was a MOP veteran. In fact, Kiran, Bobby, and Andrew had been to two previous MOPs and Lenny to three.

The MOP ended on July 7, and the members of the IMO team had a few days to catch their breath before departing for Moscow on July 9.

In addition to the six team members, the following students attended the MOP this year: Jeremy Bem of Ithaca, New York; Ruth Britto-Pacumio of Binghamton, New York; Hal Burch of Ponca City, Oklahoma; Christopher Chang of Palo Alto, California; Hank Chien of Forest Hills, New York; Timothy Chklovski of St. Louis Park, Minnesota; Matthew Crawford of Birmingham, Alabama; Andrew Dittmer of Vienna, Virginia; Craig Helfgott of Teaneck, New Jersey; Alex Heneveld of Savannah, Georgia; Jacob Lurie of Bethesda, Maryland; Elizabeth Mann of Silver Spring, Maryland; Adam Meyerson of Severna Park, Maryland; Akira Negi of Charlotte, North Carolina; Daniel Schepler of Beavercreek, Ohio; Jade Vinson of Savannah, Georgia; Stephen Wang of St. Charles, Illinois; Jonathan Weinstein of Lexington, Massachusetts; and Huan Yao of Honolulu, Hawaii.

The MOP was directed by professors

Anne Hudson of Armstrong State College, Cecil Rousseau of Memphis State University, and Daniel Ullman of George Washington University. Professors Rousseau and Ullman accompanied the team to the IMO.

What the IMO means

The IMO is more a celebration of mathematical talent than a competition for mathematical supremacy. The team scores are unofficial and informal, and the international rivalries are friendly and good-natured. The focus is on cultural exchange, friendships, and mathematics. The competition is there to invigorate the exchange, intensify the friendships, and deepen the mathematics.

The real impact of the IMO has little to do with winning and losing. The importance of the IMO is as a forum for recognizing mathematical achievement in the same sort of way that our culture recognizes athletic achievement. Mathematics is required to support the scientific, business, and health enterprises, increasingly with deeper mathematics.

We must foster mathematical achievement and encourage the development of the next generation of mathematicians. Our welfare depends on it.

Bulletin board

PROMYS at Boston University

Boston University's 1993 Program in Mathematics for Young Scientists (PROMYS) will be held June 27 to August 7. A residential program designed for 60 high school students entering grades 10 through 12, PROMYS offers a lively mathematical environment in which ambitious high school students explore the creative world of mathematics.

Through their intensive efforts to solve a large assortment of unusually challenging problems in number theory, the participants practice numerical exploration, formulation and critique of conjectures, and techniques of proof and generalization. More experienced participants may also study combinatorics and modern geometry. Problem sets are accompanied by daily lectures given by research mathematicians, and a highly competent staff of counselors lives in the dormitories and is always available to discuss mathematics with students.

Admission decisions will be based on applicants' solutions to a set of challenging problems included with the application packet, teacher recommendations, high school transcripts, and student essays explaining their interest in the program. Application materials can be obtained by writing to PROMYS, Department of Mathematics, Boston University, 111 Cummington Street, Boston, MA 02215, or by calling 617 353-2563. Applications will be accepted from March 1 until June 1, 1993.

Science bulletin board

A new nationwide electronic bulletin board system for science teachers and students has been created by the US Department of Energy's Argonne National Laboratory. Called NEWTON, the bulletin board is free and open to anyone who teaches or studies science, computer science, mathematics, or technology at any level.

NEWTON services for students include "Ask a Scientist," where callers can leave questions to be answered by scientists as part of a growing collection of questions and answers everyone can read; discussion of hobbies and other special interests; local and worldwide electronic mail over Internet; electronic exchange of computer files and software; and news about Argonne's educational and scientific programs. Services for teachers include ideas for classroom demonstrations, activities, and field trips; on-line conferences with teachers and scientists; calendars of conferences, lectures, and workshops; and several Argonne publications.

To use NEWTON, you need a telephone, a computer with a modem, and communications software. The phone number for access to NEWTON is 708 252-8241; modem settings are N, 8, 1, F. It can also be reached over Internet at the address "newton.dep.anl.gov" (or "130.202.92.50").

FUTURES: math in careers

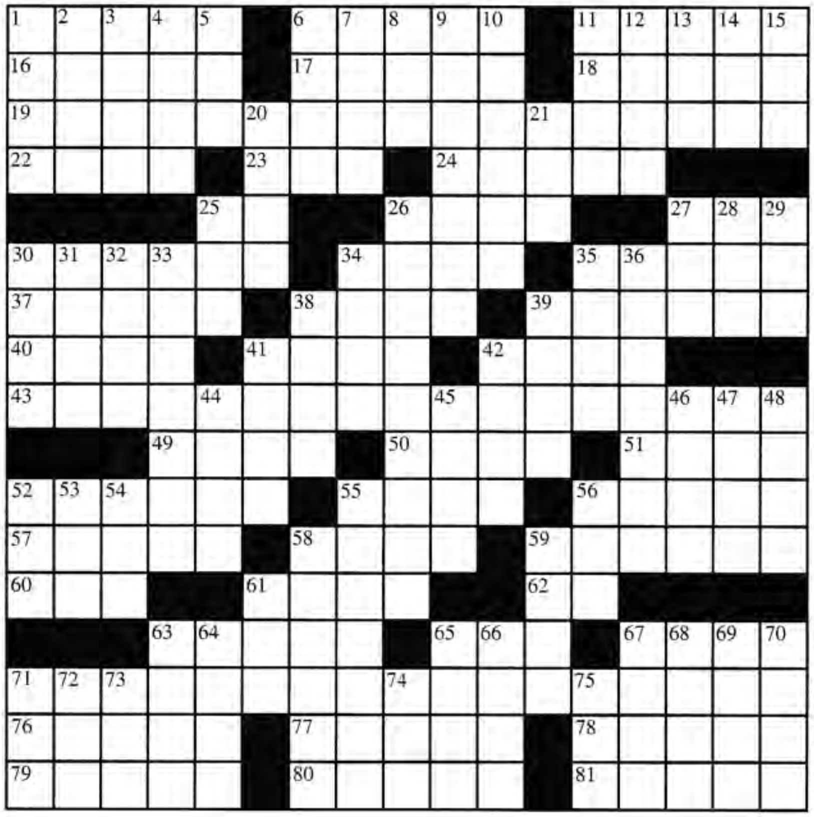
Enthusiastic math students are often faced with a question: where will they be able to use higher math once they graduate? There are the well-known technical fields such as the many branches of engineering, but what about animal care, ocean exploration, and physical fitness? As a series of educational videos called *FUTURES* shows, math is intrinsic to these and other careers.

Through interviews with professionals in dozens of fields, from graphics to space exploration, *FUTURES* shows that careers using math don't have to be dull office jobs; in fact, people using math have some of the most exciting jobs around. In the "Advanced Transportation" episode, Drs. James Powell and Gordon Danby discuss the principles they used in inventing magnetic levitation vehicles; and in "Future Habitats," the late Dr. Gerard O'Neill tells us that we would technically be able to establish colonies in space in the next ten years. Their hard work in studying math, these professionals say, has enabled them to have creative jobs they love.

The 24 shows in the *FUTURES* series are produced by the Foundation for Advancements in Science and Education (FASE). To find out how you can see *FUTURES*, call 213 965-8794, or write to FASE Productions, 4801 Wilshire Boulevard, Suite 215, Los Angeles, CA 90010.

Criss × cross science

by David R. Martin



© 1993 David R. Martin

Across

1 One who fears (suff.)
6 Protective ditches
11 Experimental ____
16 Brother of Moses
17 A gazelle
18 Make straight
19 Traditional repairman
22 Building wings
23 ___ Franklin (abbr.)
24 All of us are able (2 wds.)
25 Hello!
26 Head coverings
27 Dental org. (abbr.)
30 Ancient calculator
34 Military branch
35 The Standard ___
37 Future site of large accelerator?
38 Finished
39 Instrument part
40 ___ James (Am. Revolution statesman)
41 Units of length
42 A nomadic Moslem person
43 Supreme ETs (3 wds.)
49 Used to be Persia
50 Country in Indochina
51 End of an axis
52 Miserly
55 Fertilizer (abbr.)
56 Oils (slang)
57 Computer codes (abbr.)
58 Blows on
59 Goes by
60 Yang's partner
61 Charles Lamb
62 Element 103 (abbr.)
63 Movie award
65 Female sheep
67 Long story
71 Not classical (2 wds.)
76 City in N.Y.
77 High tidal wave in an estuary
78 Greek letter
79 Unit of length
80 Nose
81 More sage

Down

1 Step
2 The ___ effect
3 Type of exam
4 A bump (or Bruce Springsteen)
5 Naval off. (abbr.)
6 Weapon of the Middle Ages
7 Algerian sea port
8 Be in poor health
9 Wave guide propagation
10 Drowsy
11 ___ equation (for ion density)
12 Verve
13 Element 50
14 One (comb. form)
15 Chest muscle (slang)
20 A wading bird
21 Units of volume (abbr.)
25 Jan ___ (Bohemian religious reformer)
26 Lyrical passage of music
27 Sum
28 A differential operator
29 Tavern drink
30 Fund. Particle
31 Greek letter
32 The optic ___
33 Object from a mold
34 Holiday song
35 Just
36 Son of Jocasta
38 College bigwig
39 Knobs
41 Become raveled
42 After square or granny
44 Units of energy
45 Units of corn
46 People of wealth (Brit. slang)
47 Joy
48 A school term (abbr.)
52 Watch secretly
53 Three (comb. form)
54 Charged particle
55 A just person (2 wds.)
56 Ancient Roman guardian spirit
58 Burning gases
59 Earnest request
61 French coin
63 One time
64 Large ball of plasma
65 Beige
66 Hone
67 ___ Valley (city in Cal.)
68 Experts
69 Thickness
70 Like a wing
71 City in Iran
72 Am. Indian
73 River island
74 Self
75 Here's partner

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE NOVEMBER/DECEMBER PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M71

Clearly the equation can be rewritten in the form $3x^3 + 3x^2 + 3x + 1 = 0$, or $(x + 1)^3 = -2x^3$. It follows that $x + 1 = -\sqrt[3]{2}x$, so the equation has the unique real root

$$x = -\frac{1}{\sqrt[3]{2}+1}$$

M72

We start with the following lemma: *Let $S(x)$ be the sum of digits of a positive integer x, and $N(x)$ be the number of its digits that are greater than 4. Then*

$$S(2x) = 2S(x) - 9N(x).$$

To prove this, imagine we add a number x to itself, digit by digit. The sum of the digits in the result $S(2x)$ may differ from $2S(x)$ only if in some decimal places a 1 is carried to the next place. A carry from the kth place occurs if and only if the digit in this place is 5, 6, 7, 8, or 9, and in this case the 10 appearing at this point in the kth place of the sum $x + x$ is replaced by a 1 in the $(k + 1)$st place. So each carry diminishes $2S(x)$ by 9, the number of carries being equal to $N(x)$, which yields the formula stated in the lemma. The following example illustrates this reasoning:

$$\begin{array}{r} \overleftarrow{1}\ \ \overleftarrow{1}\overleftarrow{1}\ \\ 71983 \\ +\ \ 71983 \\ \hline 143966 \end{array}$$

$x = 71983$
$2S(x) = 14 + 2 + 18 + 16 + 6 = 56$
$S(2x) = 1 + 4 + 3 + 9 + 6 +6 = 29$
$N(x) = 3$, so $S(2x) = 29 = 56 - 27$
$= 2S(x) - 9N(x)$

Now it's easy to complete the solution as follows.

(a) Since $S(a) = S(b)$, $N(a) = N(b)$, we have $S(2a) = S(2b)$ by the formula in the lemma.

(b) Rewriting the formula as $S(x) = [S(2x) + 9N(x)]/2$, and substituting $a/2$ and $b/2$ for x, we see that all we have to do here is prove $N(a/2) = N(b/2)$. Take another look at the proof of the lemma above, and you'll see that the kth digit of x is greater than 4 if and only if the $(k + 1)$st digit of $2x$ is odd (because of the carry). So $N(a/2)$ and $N(b/2)$ are equal to the numbers of odd digits in a and b, respectively, which are obviously the same.

(c) Clearly $S(10x) = S(x)$, so, by the statement in section (b),

$$S(5a) = S(10a/2) = S(10b/2) = S(5b).$$

M73

In figure 1, ABC is the given triangle, I and M are its incenter and the midpoint of its hypotenuse AB, and angle BIM is the right angle from the statement. We'll prove that

$$BC : CA : AB = 3 : 4 : 5.$$

To this end, we'll show that $\tan \alpha = 1/3$, where α denotes half the measure of angle A of the triangle. Then

$$\frac{BC}{CA} = \tan\angle A = \frac{2\tan\alpha}{1-\tan^2\alpha} = \frac{2/3}{8/9} = \frac{3}{4},$$

and the ratio CA/AB is found from the Pythagorean theorem.

Noting that angle BIA equals $180° - (\alpha + \beta) = 135°$ (since $2\alpha + 2\beta = 90°$), we get the measure of angle $MIA = 135° - 90° = 45°$. The law of sines for triangle AMI yields

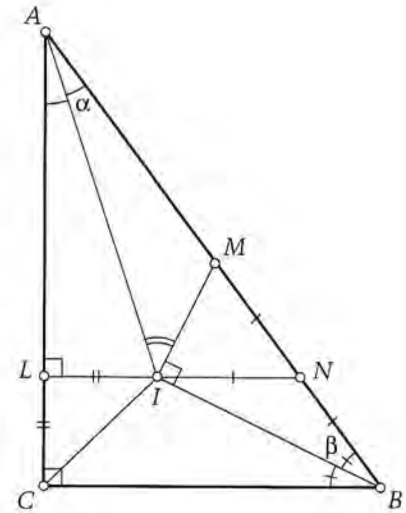


Figure 1

$$\frac{\sin\alpha}{\sin 45°} = \frac{MI}{AM} = \frac{MI}{MB}$$
$$= \sin\beta = \sin(45° - \alpha)$$
$$= \sin 45° \cdot \cos\alpha - \cos 45° \cdot \sin\alpha$$
$$= \frac{\sqrt{2}}{2}(\cos\alpha - \sin\alpha),$$

or

$$\sin\alpha = \frac{\cos\alpha - \sin\alpha}{2}.$$

Now, dividing by $\cos\alpha$ and solving for $\tan\alpha = \sin\alpha/\cos\alpha$, we get the required equality $\tan\alpha = 1/3$.

This solution can be readily generalized for any given value of the angle BIM. But a nice, simple answer like the one we've obtained often implies the existence of a nice geometrical solution without calculations. Such a solution follows.

Draw a line through the incenter I and the midpoint N of the hypotenuse BM of the right triangle BIM (see figure 1). It is known that N is the circumcenter of this triangle, so $IN = NB$ and, therefore, $\angle NIB = \angle NBI = \angle CBI$ (since BI is the bisector of the angle B). It follows that line NI is parallel to BC, so it divides the sides AB and AC in the same ratio:

$$\frac{CL}{LA} = \frac{BN}{NA} = \frac{BN}{NM+MA} = \frac{1}{3}$$

(because $MA = MB = 2BN$). But $CL = LI$, since NL is perpendicular to AC, $\angle LCI = \frac{1}{2}\angle ACB = 45°$, and so CLI is an isosceles right triangle. Finally, $\tan \alpha = \tan \angle LAI = IL/LA = 1/3$. (V. Dubrovsky)

M74

Consider six points: the vertices of the triangle and the midpoints of its sides (fig. 2). One of the five smaller triangles must cover two (or more) of these six points. Therefore, its side length is not less than half the side length of the big triangle. So each of the four triangles into which the big triangle is divided by its midlines can be covered by one of the smaller triangles.

Figure 2

M75

Party C can win any number of seats from 100 to 999, party B any number less than 900, and party A can get any number of seats less than 600.

Candidates of party C can win in all 999 districts if, for instance, the fraction of voters supporting each party is the same in all districts, as shown in figure 3a. Let's estimate from above the number of the districts where party C can lose. If x is the number of districts in which party C is supported by not more than a half of the voters, then the total number of its supporters—$0.55 \cdot 999N$, where N is the number of voters in a district—is not greater than $xN/2 + (999 - x)N = 999N - xN/2$. It follows that $x \le 2 \cdot 0.45 \cdot 999 = 899.1$. So candidates of party C necessarily win in at least 100 districts. Figure 3b shows the distribution of votes such that C wins exactly 100 seats: in each of 899 districts C is supported by $N/2 - 3$ voters, A by $N/6 + 1$ voters, B by $N/3 + 2$ voters; in the remaining districts—from the 900th to the 999th—everybody votes for C. In order that the supporters of C get the votes of 55% of all voters, the number N must satisfy $899(N/2 - 3) + 100N = 0.55N \cdot 999$, from which we get $N = 53{,}940$.

Clearly, parties A and B may not win a single seat (fig. 3a) and can't win more than 899 seats (at least 100 seats always go to C). Party B can win all of these 899 seats, if all candidates of A drop out after the first round, and so supporters of A will vote for B in the second round (fig. 3b).

It remains to figure out the maximum number y of the districts in which party A's candidate can win. To win in a district, a candidate must have at least $N/4$ supporters (otherwise, either one of the remaining two candidates gets more than $N/2$ votes and wins in the first round, or each of them gets more than $N/4$ votes, and so the first candidate drops out). So $y \cdot N/4$ doesn't exceed the total number of party A's supporters, $0.15 \cdot 999N$. It follows that $y \le 4 \cdot 0.15 \cdot 999 = 599.4$—that is, $y \le 599$, because y is an integer. The example of $y = 599$ is seen in figure 3c: in districts 1 through 599, parties A, B, and C have $N/4 + 8$, $N/4$, and $N/2 - 8$ supporters,

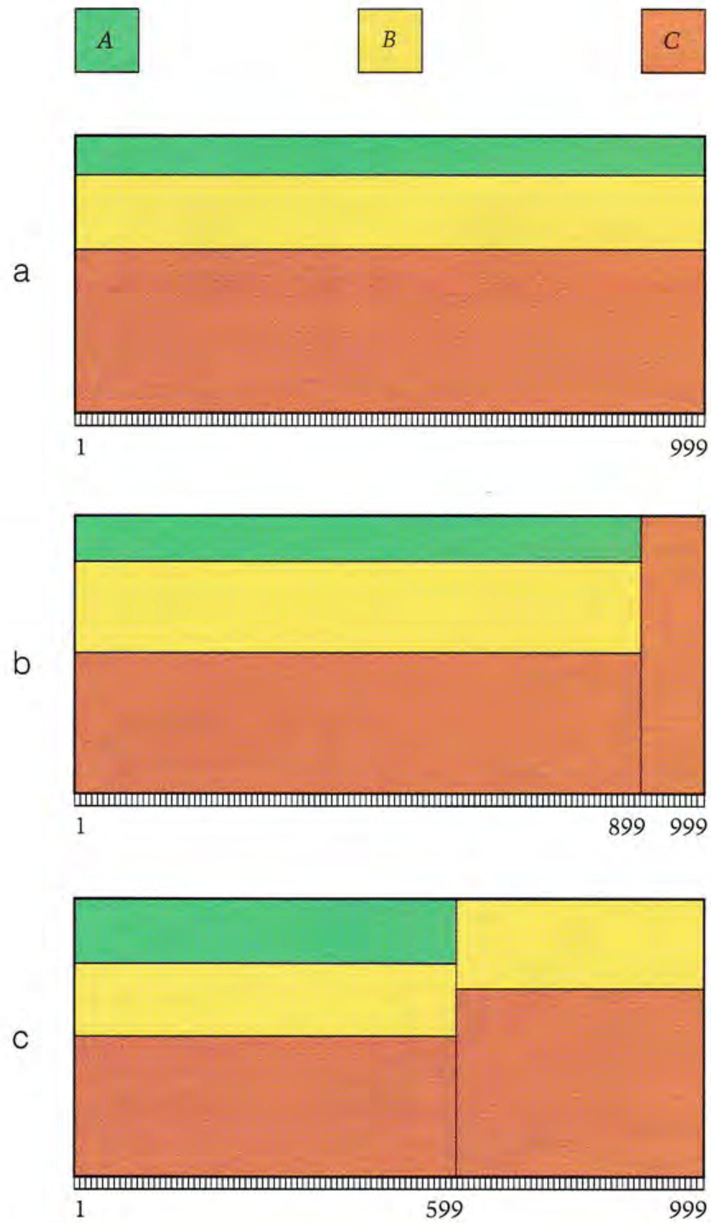


Figure 3

respectively; and if we take $N = 11{,}980$, so that $599 \cdot (N/4 + 8) = 0.15 \cdot 999N$, the total number of supporters of party A will total exactly 15% of all voters, as required. We can easily distribute votes between parties B and C in the remaining 400 districts so as to make the number of B's supporters less than that of C's in each district but total 30% of the entire electorate (which is an integer).

Physics

P71

When the bumblebee flies vertically upward with a constant speed v_1, the air resistance is directed downward:

$$F - mg - k \cdot v_1 = 0.$$

When it flies downward,

$$F + mg - k \cdot v_2 = 0.$$

Solving these two equations for the thrust F and the force of gravity mg, we obtain

$$F = k\frac{(v_1 + v_2)}{2};$$

$$mg = k\frac{(v_2 - v_1)}{2}.$$

Now, we consider the general case. If the bumblebee flies at a constant speed V at an angle α with the horizontal (see figure 4), then the net force acting on the bumblebee must be zero:

$$\mathbf{F} + m\mathbf{g} - k\mathbf{V} = 0.$$

Using horizontal and vertical components, we have

$$\boxed{V = \frac{\sqrt{(v_2 - v_1)^2 \sin^2\alpha + 4v_1 \cdot v_2} - (v_2 - v_1)\cdot\sin\alpha}{2}}$$

$$F \sin\theta = kV \sin\alpha + mg,$$
$$F \cos\theta = kV \cos\alpha.$$

Squaring each equation and adding them together, we get

$$F^2 = (mg)^2 + (kV)^2 + 2mgkV \cdot \sin\alpha.$$

From this we obtain the final expression for V (see box above).

P72

You don't need to do any extensive calculations to determine the position of the figure's center of mass—simple notions concerning energy are sufficient. Let's suspend the figure along the horizontal axis passing through the semicircle's center O, perpendicular to its plane. The figure's center of gravity C is located under the point of suspension at a distance X from it (we must find the value of X). Let's rotate the figure about the suspension point by a very small angle α, as shown in figure 5. The center of gravity rises by the distance $h = X \cdot (1 - \cos\alpha)$. On the other hand, the diameter's center of gravity remains at the same place, and the change in the semicircle's center of gravity is determined by the fact that a tiny piece (whose length is $R \cdot \alpha$) moves from the left to the right. Its center of gravity rises by the distance $R \cdot \alpha/2 - (-R \cdot \alpha/2) = R \cdot \alpha$. If mass of the entire figure is M, then the mass of the piece is

$$m = \frac{M \cdot R \cdot \alpha}{2R + \pi \cdot R} = \frac{M \cdot \alpha}{2 + \pi}.$$

Equating the changes in the potential energy, calculated by the two different methods—

$$Mgh = mgR \cdot \alpha$$

—and taking into account that $1 - \cos\alpha = 2 \cdot \sin^2(\alpha/2) \cong \alpha^2/2$ (for small angles), we get

$$X = \frac{R}{1 + \pi/2}.$$

P73

Under the conditions stated in the problem, the gas inside the vessel is not very rarefied. This means that the gas that enters the vessel is pushed into it by external pressure (see figure 6). The work W performed by the pressure can be expressed in terms of the number of moles n of gas inside the vessel:

$$W = P_0 \cdot V = n \cdot R \cdot T_0,$$

where V is the volume of n moles of gas at temperature T_0 and pressure P_0. We neglect the heat exchange between this portion of the gas and the environment because of the short time involved. Then the internal energy of this gas will increase by the amount of work:

$$\begin{aligned} n \cdot C_V \cdot T_x &= n \cdot C_V \cdot T_0 + W \\ &= n \cdot C_V \cdot T_0 + n \cdot R \cdot T_0 \\ &= n(C_V + R) \cdot T_0, \end{aligned}$$

$$T_x = T_0 \cdot \frac{C_V + R}{C_V} = 5 \cdot \frac{T_0}{3}.$$

P74

Voltage across the resistor is $100\ \Omega \cdot 0.5\ \text{A} = 50$ V. The sum of the voltages across the resistor and the device equals the circuit voltage of 220 V. Let's draw

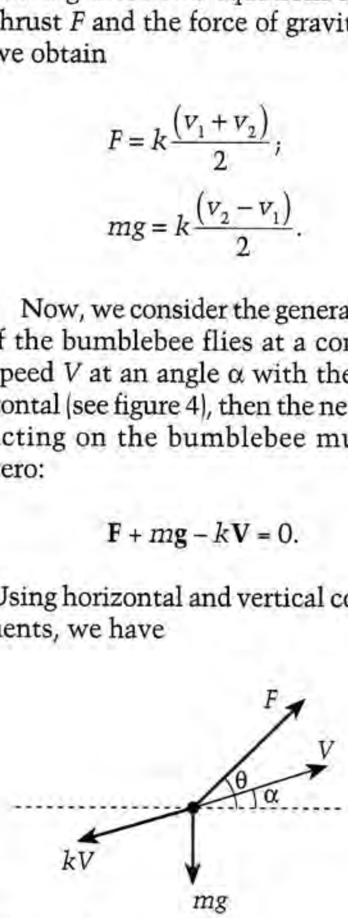


Figure 4

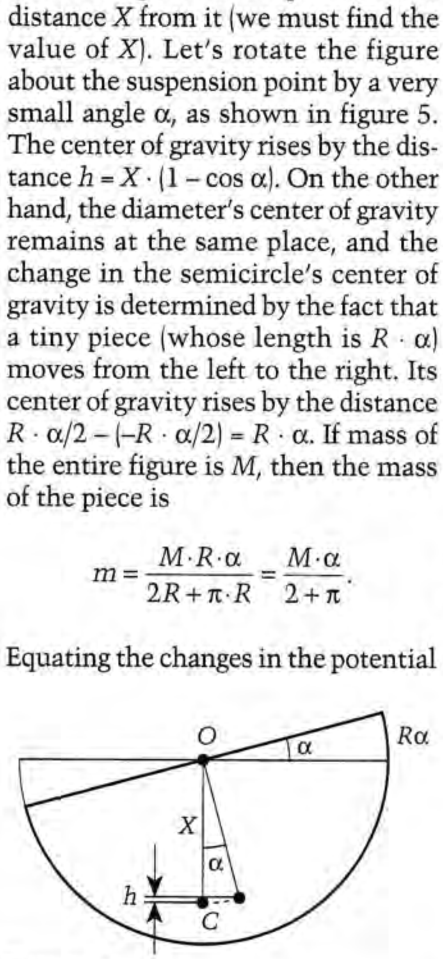


Figure 5

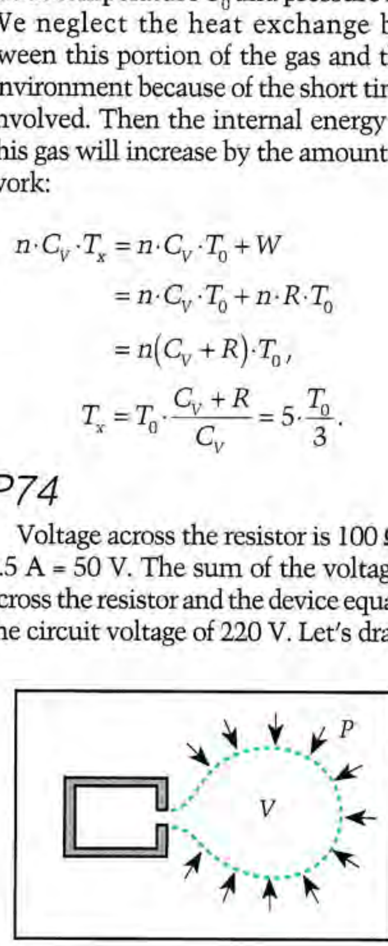


Figure 6

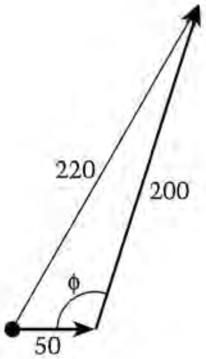


Figure 7

the vector diagram and find the angle between voltage across the device and current in the network (fig. 7). The phase of this current coincides with the phase of the voltage across the resisistor:

$$50^2 + 200^2 - 2 \cdot 50 \cdot 200 \cdot \cos\phi = 220^2,$$
$$\cos\phi \cong -0.3.$$

Now we can find the amount of power the device consumes from the network:

$$P = V \cdot I \cdot |\cos\phi| = 30 \text{ W}.$$

P75

To get a rough estimate, we can use the fact that the temperature on Earth is such that it leads to a balance between two processes: (1) absorption of solar energy falling on the Earth's surface and (2) radiation of energy from the Earth into the surrounding space. If the Earth is surrounded by a layer of soot, the solar energy will be absorbed by this layer. The layer in turn radiates the absorbed energy partly from its inner surface, partly from its outer. So only half of the energy absorbed by the layer of soot falls on the Earth: the energy emitted from its inner surface. Because half as much energy reaches the Earth, the temperature on Earth will decrease by a factor of $\sqrt[4]{2}$ (radiated energy is proportional to the fourth power of the temperature) and will be

$$T_x = \frac{T_0}{\sqrt[4]{2}} \cong 250 \text{ K} = -23°\text{C}.$$

It's very important that the Earth's radiation at such a temperature is concentrated in the infrared range of wavelengths. In this range, the layer of soot is practically transparent.

Brainteasers

B71

Yes, it can. My friend's birthday is on December 31, and our conversation took place on January 1.

B72

If you cover the left half of each pictograph (which is the reflection of the right half—see figure 8a), you'll immediately recognize the sequence 1, 2, ..., 7 in a special kind of simplified writing. (Actually, this is the standard way the digits of a zip code must be written on an envelope in Russia so they can be read by a machine.) So the next figure in the line must be 8 plus its mirror image (fig. 8b).

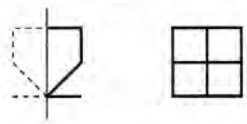


Figure 8 a b

B73

The unusual behavior of the water described in the problem can be explained by the siphon principle. Figure 9 is a schematic drawing of a karstic cavity that periodically empties the pool. Water fills the pool up to the level AA'; after that, the siphon is "switched on" and almost all the water leaves the pool. Pools where karstic cavities form siphons as described above are called intermittent pools.

B74

The difference between the time it would take to go back home and then to school and the time to go straight to school is 10 + 8 = 18 minutes. The difference between the corresponding distances is simply twice the distance from the spot where Alice remembered about *Quantum* to her home. So this spot is 9 minutes away from home, which is 9/20 of the entire distance to school.

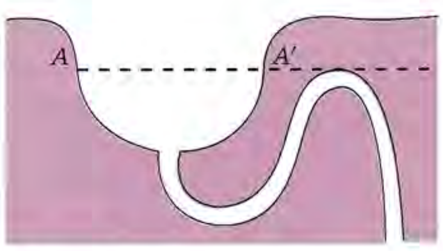


Figure 9

B75

The required slices are shown in figure 10.

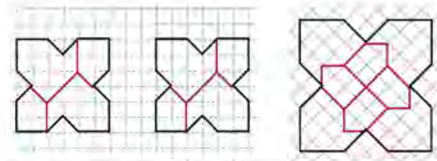


Figure 10

Democracy

1. For profile P_1, candidate *a* gets the highest score (9 points) and *c* is second with 8 points. If *b* withdraws his candidacy, we get profile P_2, where *c* wins with 7 points. Of course, there are many other examples.

	Number of votes			
Points	2	1	2	1
4	*a*	*c*	–	–
3	*b*	*d*	*a*	*c*
2	*c*	*e*	*c*	*d*
1	*d*	*a*	*d*	*e*
0	*e*	*b*	*e*	*a*
	Profile P_1		Profile P_2	

2. If *R* is the Relative Majority Rule, $R(S) = \{a\}$, $S = \{b, c, d\}$, the profile after *a* is excluded is A_1. Now $R(S) = \{d\}$, and the profile after the exclusion of *d* is A_2, which shows that *c* wins the third step and *b* is last.

Number of votes			
5	3	5	4
d	*d*	*b*	*c*
c	*b*	*c*	*d*
b	*c*	*d*	*b*

Profile A_1

Number of votes			
5	3	5	4
c	b	b	c
b	c	c	b

Profile A_2

Other rules are treated similarly.

3. Consider the profile in which k prefers d to a, a to c, and all the other voters prefer c to d, d to a. Since $\{k\}$ is decisive for a against c, $a \geq c$. By Unanimity, $d > a$. So, by Transitivity, $d \geq c$. Eliminating a by Independence, we conclude that $d \geq c$ whenever k prefers d to c and all the other voters don't.

Obsession

1. (a) If $p = 2k$, the expression $2^{2k} - 1 = (2^k - 1)(2^k + 1)$, which is just the difference of two squares.

(b) First convince yourself that $2^{35} - 1 = 2^{7\times5} - 1 = (2^7)^5 - 1 = (2^7 - 1) \times [(2^7)^4 + (2^7)^3 + (2^7)^2 + (2^7) + 1]$. Then in general we see that $2^{pq} - 1 = (2^p - 1) \times [(2^p)^{q-1} + (2^p)^{q-2} + (2^p)^{q-3} + \ldots + (2^p) + 1]$.

2. Since $2^{p-1}(2^p - 1) = 2^p(2^p - 1)/2 = n(n + 1)/2$ for $n = 2^p - 1$, all perfect numbers are triangular.

3. An even perfect number is expressible as $2^{p-1}(2^p - 1) = 2^{p-1}[(2^p - 1)/(2 - 1)]$. This is just the sum of p terms in a geometric series starting with 2^{p-1} and proceeding with the common ratio 2: $2^{p-1} + 2^p + 2^{p+1} + \ldots + 2^{2(p-1)}$. Therefore, all even perfect numbers are equal to the sums of successive powers of 2.

4. This is merely a restatement of the previously proved fact that even perfect numbers are sums of successive powers of 2.

5. The smallest prime divisor must be 23 or less, because the largest number less than 29 (28 + 1) that is prime is 23.

6. The question of how many digits are in 2^{2p} (p = 756,839) is equivalent to asking what is $x + 1$ in the expression $10^x = 2^{2p}$. Taking the logarithm to base 10 of both sides yields $x = 2p\ \log_{10}2$, or 461,082 digits in this case.

7. (a) M_{44497} requires 1,855 digits when represented in base 2^{24} because 44,497 ÷ 24 = 1,854.04… . In other words, $(2^{24})^{1854}$ is less than $2^{44497} - 1$, but $(2^{24})^{1855}$ is greater than $2^{44497} - 1$. And since the positional notation starts at zero, we need 1,854 + 1 digits to represent M_{44497}.

(b) Notice that $2^{10} - 1$ in base 2^3 is 1,777. By analogy, if z is the largest digit in the base ($z = 2^{24} - 1$), then the base 2^{24} representation of $(2^{44497} - 1)_{10}$ is a 1 followed by 1,854 z's.

Kaleidoscope

1. Since the given trapezoid is circumscribed about a circle (fig. 11), $2AB = AB + CD = AD + BC = a + b$, so $AB = m = (a + b)/2$. The height can be found from the Pythagorean theorem in the right triangle ABE, in which $AE = (a - b)/2$ (to see why, drop the height from C):

$$BE^2 = AB^2 - AE^2$$
$$= \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2$$
$$= ab,$$

so $BE = g = \sqrt{ab}$. By the similarity of right triangles BFE and BEA, $BF/BE = BE/BA$, or $BF = BE^2/BA = g^2/m = 2ab/(a + b) = h$. So the inequalities $m > g > h$ are reduced to $AB > BE > BF$, which is obvious. We note that this argument holds for any two positive numbers a and b, because one can always use them as the bases of an isosceles trapezoid that possesses an inscribed circle.

2. Extend the sides of the trapezoid to meet at P (fig. 12). The areas of the similar triangles APD, QPR, and BPC are proportional to the squares of their corresponding parts—in particular, to a^2, q^2, and b^2, respectively. On the other hand, area(APD) − area(QPR) = area($AQRD$) = area($QBCR$) = area(QPR) − area(BPC), which means that $a^2 - q^2 = q^2 - b^2$, or $q = \sqrt{(a^2 + b^2)/2}$. Further, consider the ratios of the distances from each of the segments g, m, and q to the bases b and a of the trapezoid ($b < a$—see figure 1 in the article). For segment g, this ratio is the ratio of similarity of the two smaller trapezoids: $b/g < 1$. For segment m, this is exactly 1, obviously; for segment q, it is greater than 1, because the upper trapezoid has the smaller sum of bases ($b + q < q + a$), but the same area as the lower one, so its height must be greater.

3. From similar triangles ABC, AOH, $t/b = BH/AB$. From similar triangles BAD, BOH, $t/a = HA/AB$. Adding, we have $t/a + t/b = (BH + HA)/AB = 1$, or $1/t = 1/a + 1/b$. A similar argument will show that $1/u = 1/a + 1/b$. Hence $t = u = ½h$, and $2/h = 1/a + 1/b$. Solving for h shows that h is the harmonic mean of a and b.

4. If the athlete's speed with respect to the air is v and the wind speed is u, we can assume that half of the track is run at a speed of $v + u$ and the other half at a speed of $v - u$. So the average speed is the harmonic mean of $v + u$ and $v - u$, which is less than the arithmetic mean v. This means that the runner must work harder to show the same average speed as in calm weather.

5. For $n > 2$, the red part of figure 6 in the article shows that point A_{n+1} completes the harmonic range (A, A_n; A_{n-1}, A_{n+1}). Therefore, AA_n is the harmonic mean of AA_{n-1} and AA_{n+1}, so that $1/AA_n$ is the arithmetic mean of $1/AA_{n-1}$ and $1/AA_{n+1}$.

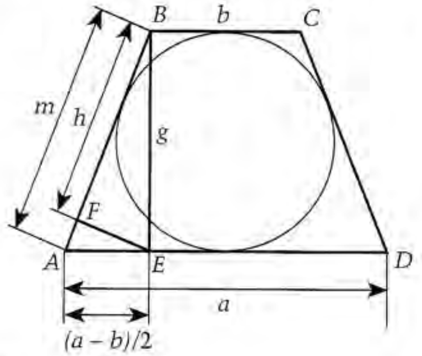


Figure 11

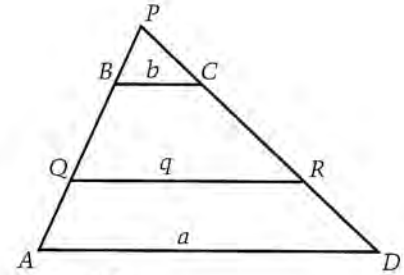


Figure 12

The case $n = 2$ requires special treatment. If we draw line TU parallel to AA_1 through R (T lying on A_1B, U on AB_1, and R on OA_2), then we proved, as part of problem 3, that $TR = RU$. Hence OR is a median in triangle OTU. Since triangle OA_1A is obtained from OTU by a dilation about point O, this means that OA_2 is a median in triangle OAA_1 too, so $AA_2 = 1/2$.

Thus, $1/AA_1$, $1/AA_2$, ... is an arithmetic sequence such that $1/AA_1 = 1$, $1/AA_2 = 2$, ...—that is, the sequence 1, 2, 3, ... So $1/AA_n = n$, and $AA_n = 1/n$.

You may have noticed that the construction of A_2 does not really differ from that of its successors, yet we had to speak about it differently. We can fix this if we take for AA_0 a "point at infinity" on line AA_1, so that $AA_0 = \infty$, and $1/AA_0 = 0$ (A, A_1, A_0, and A_2 provide the limiting case of an ordinary harmonic range).

6. Obviously

$$\begin{aligned}&\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{2n}\\&\geq\frac{1}{2n}+\frac{1}{2n}+\cdots+\frac{1}{2n}\\&=\frac{1}{2}.\end{aligned}$$

From this estimate it follows that the sum of the first 2^k terms of the sequence 1, 1/2, 1/3, ... is greater than $(k + 1)/2$:

$$\begin{aligned}&1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{2^k}\\&=1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)\\&\quad+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)+\cdots\\&\quad+\left(\frac{1}{2^{k-1}+1}+\cdots+\frac{1}{2^k}\right)\\&\geq\frac{k+1}{2},\end{aligned}$$

so it can be made arbitrarily large.

7. It suffices to show that the center of gravity of any "subtower" of n upper bricks falls exactly on the edge of the $(n + 1)$st brick (counting from the top). We'll prove it by induction. For $n = 1$ this is certainly true. Assume that it's true for the tower of n bricks. What happens when we add the next brick from below? The center of gravity moves $1/(n + 1)$ the distance from its old position C to the center O of the $(n + 1)$st brick (because the n-brick tower is n times heavier than one added brick—see figure 13 for $n = 2$). Since it was exactly above the edge of the $(n + 1)$st brick, by the inductive assumption, the center of gravity now falls on a point shifted by $1/n$ of a unit from this edge. This is just where the edge of the next—that is, $(n + 2)$nd—brick is set.

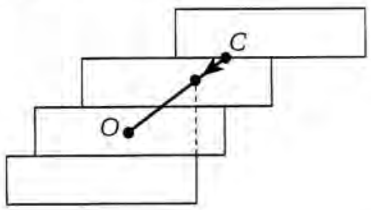


Figure 13

Hit or miss

(See the Kaleidoscope in the November/December issue)

1. The usual answer is that a log twice as thick but half as short weighs the same. But that's wrong. By doubling the transverse size (diameter) of a log, the weight becomes *four* times greater; by halving the length of a log, the weight is reduced by a factor of only *two*. So the thickened, shortened log doubles in weight in comparison with the longer, thinner log—that is, it weighs 60 kilograms.

2. When immersed in water, any (solid) iron object loses an eighth of its weight. So the weights under water will weigh only $\frac{7}{8}$ of their former weight, and the nails also $\frac{7}{8}$ of their former weight. Because the weights were a tenth as heavy as the nails, underwater they remain a tenth as heavy as the nails. As a result, the decimal scale remains balanced underwater.

3. Let's follow the movement of the hands at twelve o'clock. At that moment the hands coincide. Because the hour hand moves a twelfth as fast as the minute hand (the hour hand completes a revolution in 12 hours, the minute hand in 1 hour), then in the next hour the hands cannot meet.

Figure 14

But after one hour, the hour hand is positioned as in figure 14, after having gone through $\frac{1}{12}$ of a complete revolution; the minute hand has made a complete revolution and stands again as in figure 14—a twelfth of a revolution behind the hour hand. Now the conditions of the race are different than before: the hour hand moves more slowly than the minute hand, but it is situated in front of it, and the minute hand must overtake it. If the race continued an entire hour, then the minute hand would make a complete revolution and the hour hand would only make $\frac{1}{12}$ of a revolution—that is, the minute hand would make $\frac{11}{12}$ of a revolution more. In order to overtake the hour hand, the minute hand has to travel farther than the hour hand, but only the $\frac{1}{12}$ of a revolution that separates them. It doesn't need a full hour to accomplish that; it needs less time in the same proportion as $\frac{1}{12}$ is less than $\frac{11}{12}$—that is, $\frac{1}{11}$. So the hands meet in $\frac{1}{11}$ of an hour—that is, in $60/11 = 5\frac{5}{11}$ minutes.

Therefore, the hands meet $5\frac{5}{11}$ minutes after one hour passes—that is, at $5\frac{5}{11}$ minutes past one o'clock.

When will the hands meet next? This meeting will take place 1 hour, $5\frac{5}{11}$ minutes later—that is, $2{:}10\frac{10}{11}$; the next will once again take place 1 hour, $5\frac{5}{11}$ minutes later—that is, at $3{:}16\frac{4}{11}$; and so on. There will be eleven such meetings, the eleventh one occurring $1\frac{1}{11} \times 11 = 12$ hours after the first one—that is, at twelve o'clock; in other words, the eleventh meeting coincides with the very first one, and then the entire process repeats itself.

Here are the times when the hands coincide:

1st meeting—1:05$\frac{5}{11}$,
2nd meeting—2:10$\frac{10}{11}$,
3rd meeting—3:16$\frac{4}{11}$,
4th meeting—4:21$\frac{9}{11}$,
5th meeting—5:27$\frac{3}{11}$,
6th meeting—6:32$\frac{8}{11}$,
7th meeting—7:38$\frac{2}{11}$,
8th meeting—8:43$\frac{7}{11}$,
9th meeting—9:49$\frac{1}{11}$,
10th meeting—10:54$\frac{6}{11}$,
11th meeting—12:00.

A short way to see this solution is to count the number of times the hands meet (eleven) and divide the 12 hours by this number.

4. The solution to this problem is similar to that of the previous one. Let's begin again at twelve o'clock, when the hands coincide. It's necessary to calculate the time the minute hand needs to overtake the hour hand by exactly half a revolution—in this case, the hands will point in opposite directions. We've seen (in the previous problem) that in one hour the minute hand goes $\frac{11}{12}$ of the way toward catching the hour hand; in order to catch up by $\frac{1}{2}$ of a revolution, less than a full hour is needed—less in the same proportion as $\frac{1}{2}$ is less than $\frac{11}{12}$; that is, $\frac{6}{11}$ of an hour is needed. So the hands point in opposite directions for the first time in $\frac{6}{11}$ of an hour after twelve o'clock—at 32$\frac{8}{11}$ minutes past 12:00. Look at the clock at exactly 32$\frac{8}{11}$ past 12:00 and you'll convince yourself that the hands point in opposite directions.

Is this the only time the hands point in opposite directions? By no means! They take the same relative position 32$\frac{8}{11}$ minutes *after each coincidence (or meeting)*. As we've seen, there are 11 such meetings every 12 hours; so the hands point in opposite directions 11 times every 12 hours. It's easy to find these times:

12:00 + 32$\frac{8}{11}$ minutes = 12:32$\frac{8}{11}$,
1:05$\frac{5}{11}$ + 32$\frac{8}{11}$ minutes = 1:38$\frac{2}{11}$,
2:10$\frac{10}{11}$ + 32$\frac{8}{11}$ minutes = 2:43$\frac{7}{11}$,
3:16$\frac{4}{11}$ + 32$\frac{8}{11}$ minutes = 3:49$\frac{1}{11}$
. . .

—you can compute the rest of them yourself.

5. The usual answer is "7 seconds," but that's wrong. When the clock strikes three we observe two intervals: (1) between the first and the second strikes; (2) between the second and the third strikes. Both intervals together last 3 seconds. Therefore, each of them lasts half as long—that is, 1$\frac{1}{2}$ seconds. When the clock strikes 7:00, we have 6 such intervals, and $6 \times 1\frac{1}{2}$ seconds make 9 seconds. Consequently, the clock strikes 7:00—that is, it makes 7 strikes—in 9 seconds.

6. The row of peas would be much longer than a dinner table. The diameter of a pea is approximately $\frac{1}{3}$ cm to $\frac{1}{2}$ cm. If we take the larger number, then in a cube with an edge length of 1 cm one can pack at least $2 \times 2 \times 2 = 8$ peas (packed loosely; with tight packing, the number will increase). So in a glass with a volume of 200 cm^3, the number of peas is likely to be at least 1,600 (the exact number will depend on the shape of the glass and how the peas are packed). Lining the peas up in a row, we'll get a length of $\frac{1}{2} \times 1{,}600$ cm = 800 cm, or 8 m—much longer than your ordinary dinner table.

If we use $\frac{1}{3}$ cm as a pea's diameter, then one can probably fit at least $3 \times 3 \times 3 = 27$ peas in a cubic centimeter; and in a glass, at least 5,400 peas. The length of a row of such peas is $\frac{1}{3} \times 5{,}400$ cm = 1,800 cm, or 18 m—even more than when the size of an individual pea was taken to be larger.

7. Not only a house—you could encircle an entire town with these leaves lined up in a row, because such a row would reach about 10 km! There are at least 200,000–300,000 leaves on an old tree. Even if we choose the lower bound of 200,000 and consider each leaf to be 5 cm wide, the row will extend 125,000 cm—that is, 12,500 m, or 12$\frac{1}{2}$ km.

8. A million steps will take you much farther than 10 km—farther even than 100 km. If your step is about $\frac{3}{4}$ m, then 1,000,000 steps is 750 km. That's just about the distance from New York City to Washington, D.C. (as the crow flies).

9. Both men counted the same number of passers-by. The one standing at the door counted people coming from both directions, and the one pacing back and forth counted these same people coming toward him.

10. This problem yields readily to algebra. We give here a more intuitive arithmetic solution.

If the child is now a third the age of the parent, then the parent is older by the doubled age of the child. Five years ago the parent was, of course, older than the child by the doubled *present* age of the child, since the difference in their ages remains the same. On the other hand, at that time the parent was four times older than the child; therefore, the parent was older by the tripled age of the child *at that time*. Consequently, the doubled *present* age of the child is equal to the child's tripled *former* age; in other words, the child is now 1$\frac{1}{2}$ times older than five years ago. It's not hard to see that five years is half the child's former age; this means that five years ago the child was 10 years old and is now 15 years old.

Thus, the child is now 15 years old and the parent is 45 years old; five years ago the parent was 40 years old and the child was 10 years old—that is, a fourth as old.

11. Mercury is much heavier than water; so we would expect that mercury will pour out more quickly. But Toricelli knew that this line of thinking is wrong: the rate at which it pours out doesn't depend on the density of the fluid. It's determined by Toricelli's formula

$$v = \sqrt{2gh},$$

where v is the velocity of the escaping stream, g is the acceleration due to gravity, and h is the height of the fluid in the vessel. As we can see, density isn't involved in the formula.

This paradoxical law becomes quite understandable, however, if we take into account the fact that the driving force is the weight of the upper layers of fluid. For the heavier fluid, this force is greater than for the lighter fluid, but the mass put into motion in the first case is also greater than in the second case and by the same factor. It's not surprising that the accelerations and, consequently, the velocities in both cases turn out to be equal.

12. A fire engine can't pump boiling water, because there will be vapor with a partial pressure of 1 atm under the piston instead of rarified air.

13. It has often been written and said that the same temperature—4°C—reigns year-round at the bottom of deep rivers, because at this temperature water has the greatest density. This is true for standing reservoirs of fresh water. But in rivers—despite what many textbooks say—the distribution of temperature is something else entirely. In river water there is not only the visible *longitudinal* current, there are also the *transverse* currents. All the water in the river is constantly being stirred; that's why its temperature near the bottom is the same as at the surface. As M. A. Velikanov (author of *Land Hydrology*) says, "for all oscillations of the air temperature, these oscillations very quickly penetrate down to the bottom of a stream, and the most exact thermometers could not record the difference in temperature of different layers of water even at considerable depths."

So the correct answer is this: at the bottom of a very deep river, the water in summer is warmer than in winter by the same number of degrees as the summer air is warmer than the winter air.

14. At high temperatures, steel beams lose a considerable portion of their strength. At 500°C the fracture strength of steel is half that at 0°C; at 600°C—a third; at 700°C—about a seventh. (Here are more exact data: if we take the strength at 0°C to be 1, then at 500°C the strength is 0.45; at 600°C—only 0.3; and at 700°C—0.15.) So in a fire, steel structures collapse under their own weight.

15. Most people are sure that steam is white, and so they're very surprised to hear that steam is absolutely transparent, invisible, and, consequently, colorless. The white fog that is usually called "steam" is really not vapor in the physical sense but water dispersed into small droplets. Clouds also consist not of steam but of minute droplets of water.

16. This isn't a trick question but a perfectly serious problem in physics. As a match burns, heat develops and energy is liberated. How many joules does a burning matchstick develop per second? In other words, what is the power of a burning matchstick in watts? As you see, there's no funny business in this problem.

There's no reason to think the energy is ridiculously small. It's not hard to see that it's actually rather significant. Here are the calculations. A matchstick weighs about 100 mg, or 0.1 g (its weight can be determined by just weighing it, or—if you don't have a scale—by measuring the volume of a match and assuming a specific gravity of 0.5). The thermal output of wood can be taken as 3,000 cal/g. It can easily be determined that a match burns for about 20 s. So from 300 cal ($3,000 \times 0.1$) liberated while the match burns, in 1 second we get $300 \text{ cal} \div 20 = 15$ cal. Each calorie corresponds to 4.2 joules; thus, the power of a burning matchstick is $4.2 \text{ J/cal} \times 15 \text{ cal} = 63$ W. Consequently, the burning matchstick surpasses a 50-watt light bulb in power.

17. The solubility of the overwhelming majority of solids in water increases with temperature; for example, at 0°C sugar can be dissolved in water to a concentration of 64%, and at 100°C to a concentration of 83%. Salt, however, is an exception; its solubility in water is almost independent of temperature: at 0°C, it is soluble to a concentration of 26%; at 100°C—to 28%. At 40°C and 70°C equal amounts of salt can be dissolved—namely, to a concentration of 27%.

18. The sound we hear when we put a cup or shell to our ear is due to the fact that these objects are *resonators* amplifying numerous sounds in our environment, which we usually don't notice because they're so weak. This mixed sound resembles the roar of the sea, which gave rise to numerous legends about the "sea" trapped inside seashells.

19. It would seem that nothing surpasses black velvet in blackness or snow in whiteness. But these classic examples of black and white, dark and light, present themselves quite differently when approached with an impartial physical instrument—a photometer. Then it turns out that the blackest velvet in sunlight is *lighter* than the purest snow in moonlight.

This is because a black surface, however dark it may seem, doesn't completely absorb all the incident rays of visible light. Even soot and platinum black—the blackest colors known—diffuse about 1–2% of incident light. Let's take the figure 1% and assume that snow diffuses all 100% of incident light (which is, of course, an exaggeration—new-fallen snow diffuses about 80% of incident light). It's known that the Sun's illumination is 400,000 times stronger than the Moon's. Therefore, 1% of sunlight diffused by black velvet is 1,000 times stronger than 100% of moonlight diffused by snow. In other words, black velvet in sunlight is many times lighter than snow illuminated by the Moon.

The same is true, naturally, not only of snow but of the very whitest paints (the lightest of which diffuse 91% of incident light). Because no surface (if it hasn't been heated) can reflect more light than falls on it, and the Moon sends 1/400,000 of the Sun's light, then the existence of white paint that is lighter—as objectively measured—in moonlight than the blackest paint in sunlight is inconceivable.

20. In order for gold to lose its characteristic yellow color, one has to observe it in light from which yellow rays have been excluded. Newton achieved this by blocking the yellow color of the spectral band, letting the other colors pass, and bringing them together again with a lens. "If one stops yellow rays before they enter the lens," Newton wrote, "then gold (illuminated by the other rays) appears as white as silver."

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Title of publication: Quantum. A. Publication No.: 008544. (2) Date of filing: 10/1/92. (3) Frequency of issue: Bimonthly. A. No. of issues published annually, 6. B. Annual subscription price $28.00 (4) Location of known office of publication: Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. (5) Location of the headquarters of general business offices of the publishers: 175 Fifth Avenue, New York, NY 10010. (6) Names and addresses of publisher, editor, and managing editor: Publisher: Bill G. Aldridge, National Science Teachers Association, 3140 North Washington Boulevard, Arlington, VA 22201, in cooperation with Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010; Editor(s): Larry D. Kirkpatrick / Mark Saul / Constantine Bogdanov / Vladimir Dubrovsky, National Science Teachers Association, 3140 North Washington Boulevard, Arlington, VA 22201; Managing Editor: Timothy Weber, National Science Teachers Association, 3140 North Washington Boulevard, Arlington, Va 22201. (7) Owner: National Science Teachers Association, 3140 North Washington Boulevard, Arlington VA 22201. (8) Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total of bonds, mortgages or other securities; none. (9) The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: has not changed during the preceding 12 months. (10) Extent and nature of circulation. A. Total no. copies printed (net press run): Average no. copies each issue during the preceding 12 months, 10,667; no. copies single issue nearest to filing date, 10,500. B. Paid circulation: 1. Sales through dealers and carriers, street vendors, and counter sales: Average no. copies each issue during preceding 12 months, 12; no. copies single issue nearest to filing date, 12. 2. Mail subscriptions: Average no. copies each issue during preceding 12 months, 6,394; no. copies single issue nearest to filing date, 6,206. C. Total paid circulation: Average no. copies each issue during preceding 12 months, 6,406; no. copies single issue nearest to filing date, 6,218. D. Free distribution by mail, carrier, or other means. Samples, complimentary, and other free copies: Average no. copies each issue during preceding 12 months, 429; no. copies single issue nearest to filing date, 429. E. Total distribution: Average no. copies each issue during preceding 12 months, 6,835; no. copies single issue nearest to filing date, 6,647. F. Copies not distributed. 1. Office use, left-over, spoiled after printing: Average no. copies each issue during preceding 12 months, 3,832; no. copies single issue nearest to filing date, 3,853. 2. Return from news agents: Average no. copies each issue during preceding 12 months, 0; no. copies single issue nearest to filing date, 0. G. Total. Average no. copies each issue during preceding 12 months, 10,667; no. copies single issue nearest to filing date, 10,500. I certify that the statements made by me above are correct and complete.

Craig Van Dyck

Craig Van Dyck
Vice-President, Production

Corrections

Vol. 3, no. 1:

p. 58, col. 1, l. 11: *for* figure 4 *read* figure 5.

p. 60, col. 1, third line from bottom: *for* triangle *CBD read* triangle CB_1D_1.

Vol. 2, no. 6:

An editorial mishap resulted in an error in the solution to problem 4 in the article "Summertime, and the Choosin' Ain't Easy" (p. 61). There is no need to multiply the number of arrangements of 0's and 1's by six, so there are actually only 28 ways for the rich father to distribute six pennies to his three children. Many thanks to reader Marnold Ostby for pointing this out.

Vol. 2, no. 5:

Reader Douglas McIlory gives a beautifully intuitive solution to problem M55 (see pages 19 and 57). McIlory says:

First, consider the two-dimensional problem, where the planets are unit discs in the plane. Wrap a string tightly around the outside of the set of planets. The string will take on a shape that consists of straight segments between planets alternating with arcs in contact with planets. The arcs touch precisely the regions of the planets from which no other planets can be seen. These are of course also the regions that cannot be seen from any other planet. Now shrink the lengths of the straight segments to zero so that only the arcs remain. Together the arcs form a circle. (They have the same radius, join smoothly, and enclose an area of the plane.) Thus the measure of the regions in question is equal to the measure of the boundary of one disc.

The three-dimensional problem is solved the same way, by stretching a rubber sheet around the configuration of spheres and shrinking all the flattenable (plane and cylindrical) parts until only spherical parts are left. These spherical parts arose from the original regions of invisibility. Together they form a sphere. (Like the arcs, they have the same radius, join smoothly, and enclose a volume in space.)

Armed with this intuition, which puts the invisible regions in one-to-one correspondance with the surface of a sphere, one can easily fill in the details.

Vol. 2, no. 3:

Some typographical errors crept into the solution to problem M43 (p. 79). The hidden edges issuing from point Q in the cube are not QX, QY, and QZ: they are not named by points in the diagram. The points A_1, A_2, A_3, A_4, A_5, A_6 mentioned at the end of the third paragraph are, in fact, points A_1', A_2', A_3', A_4', A_5', A_6'. Plane PXZ referred to in the next paragraph should be plane QWZ, and line $K'M'$ is the intersection of the plane determined by K', M', and L' with this plane.

Index of Advertisers

American Association of Physics Teachers	24
Beloit College	9
Clarke College	45
Drexel University	35
Embry-Riddle Aeronautical University	31
Independent Educational Services	44
International Educational Network	50
NSTA Special Publications	Cover 4

The stomachion

If you stay calm, this "infuriating" game cannot cause heartburn

by Yuli Danilov

THE "STOMACHION" CAN rightfully be considered the grandaddy of slice-and-rearrange games. One familiar member of that family is the tangram.[1] The stomachion is more than 2,000 years old and has certainly passed the test of time. I'd like to introduce you to this ancient but not decrepit game. It is a true classic.

Let's get started

Figure 1 shows the most common "raw material" for the stomachion: a rectangle *ABCD* whose side lengths have a ratio of 1 to 2. Sometimes a square or even an arbitrary parallelogram is used instead of such a rectangle. Whichever initial figure is chosen, though, the way it's sliced up remains the same.

Let *ABCD* be an arbitrary parallelogram, *F* and *E* the midpoints of its sides *AD* and *BC*. (Figure 1 shows a rectangle *ABCD*, but you can easily imagine it to be any parallelogram.) Drawing the line *EF*, divide the initial parallelogram into parallelograms *ABEF* and *FECD*. Draw the diagonals *AC*, *BF*, and *FC*. Let *L* and *R* be the points of intersection of diagonal *AC* with diagonal *BF* and line *EF*. Through the midpoint *G* of the line *BE* draw a line parallel to the side *AB* of the initial parallelogram that meets diagonal *BF* at the point *H*, and a segment *GK* of line *GA* (*K* is the point of intersection of line *GA* and diagonal *BF*). Finally, join the corner *B* to the

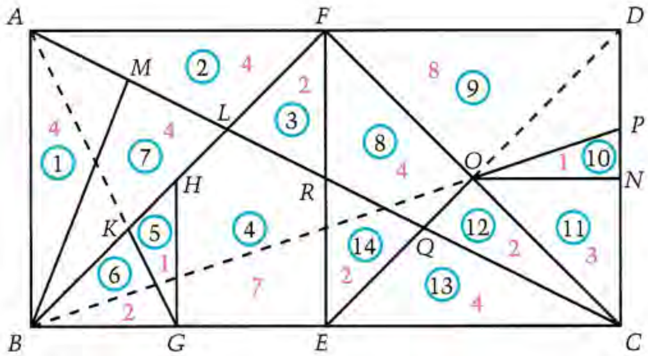


Figure 1
How to create stomachion pieces.

midpoint *M* of *AL*, thus dividing parallelogram *ABEF* into seven pieces.

Let *N* and *O* be the midpoints of line segments *CD* and *FC*. Draw lines *ON* and *OE*. Let *Q* be the point of intersection of line *OE* with diagonal *AC*, and let *P* be the point of intersection of line *BO* with side *CD* of the initial parallelogram. Draw line *OP*. Then parallelogram *FECD* is also divided into seven pieces.

Thus the initial parallelogram *ABCD* is divided into 14 pieces.

In creating new figures, you can flip over pieces of the initial figure. But you must use all 14 pieces. All the silhouettes you see in figures 2 through 4 are composed strictly in accordance with this indispensable rule. When you've mastered the secrets of stomachion mosaics by uncovering their structure, you'll be in a position to construct your own figures.

A bit of history

Among geometric puzzles, the stomachion is notable not only for its venerable age but for its distinguished origin.

The game was known in ancient times. It's also mentioned by Roman writers in the 4th through 6th centuries A.D. It was thought that the stomachion was invented by Archimedes, but for a long time there was no documented support for this assertion.

In 1899 the Swiss historian Heinrich Suter, working in the libraries of Berlin and Cambridge, found an Arabic manuscript with fragments of a treatise entitled

Figure 2
Hen.

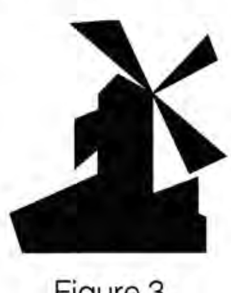


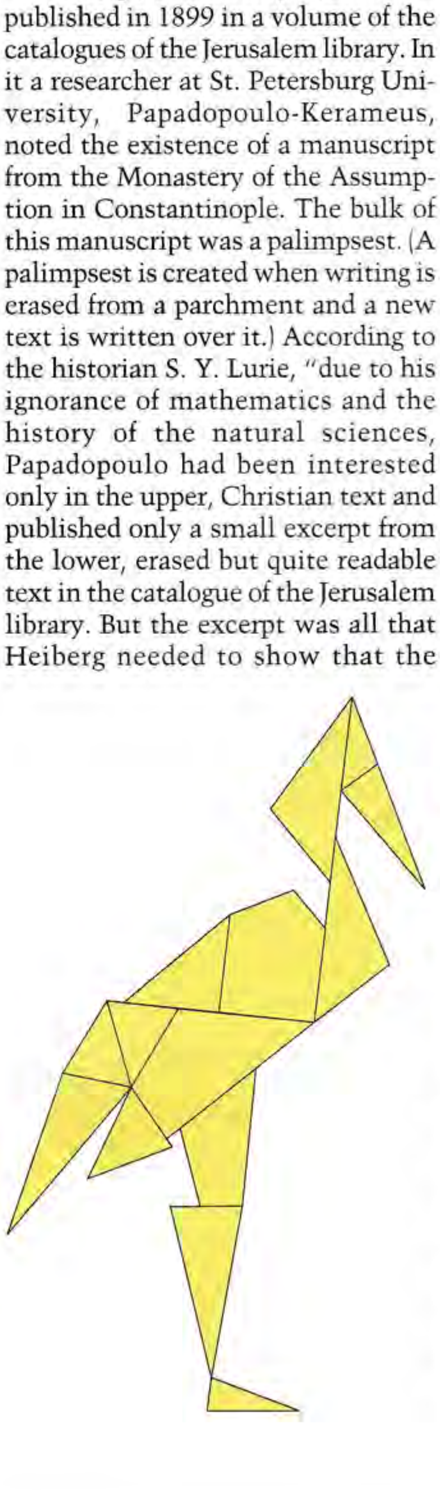
Figure 3
Windmill.

Figure 4
Rooster.

[1]See the Toy Store in the September/October 1992 issue.—*Ed.*

"Archimedes's Book on the Division of the Stomachion Figure into Fourteen Pieces that are in Rational Proportions to It." But some historians expressed skepticism that the work belonged to Archimedes. All doubts were swept away when the well-known Danish historian of mathematics I. Heiberg found the Greek original of some fragments of Archimedes's work—and he made the discovery without leaving his own study.

Heiberg was drawn by a short item published in 1899 in a volume of the catalogues of the Jerusalem library. In it a researcher at St. Petersburg University, Papadopoulo-Kerameus, noted the existence of a manuscript from the Monastery of the Assumption in Constantinople. The bulk of this manuscript was a palimpsest. (A palimpsest is created when writing is erased from a parchment and a new text is written over it.) According to the historian S. Y. Lurie, "due to his ignorance of mathematics and the history of the natural sciences, Papadopoulo had been interested only in the upper, Christian text and published only a small excerpt from the lower, erased but quite readable text in the catalogue of the Jerusalem library. But the excerpt was all that Heiberg needed to show that the erased portion was a text by Archimedes."



It wasn't until 1906 that Heiberg was able to take a look at the original manuscript. It was impossible to see any trace of the original (underlying) text on 29 of 177 sheets. On nine other sheets the earlier text was hopelessly damaged when erased and only individual words can be read. But the underlying text is quite legible on the rest of the sheets.

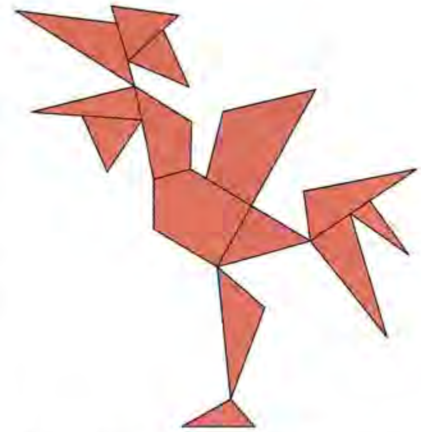
Investigating the original manuscript and a photograph of it, Heiberg managed to establish the content of the main body of the manuscript.

Heiberg's scholarly achievement rescued the following works of Archimedes from virtual oblivion: "On the Sphere and Cylinder" (the greater part), "On Spirals" (almost complete), "On Measuring the Circle" and "On the Equilibrium of Plane Figures" (fragments), and "On Floating Bodies" and "Message to Eratosthenes" (significant portions). Among the restored texts were two theorems from Archimedes's book on the stomachion.

By the way, Heiberg's discovery put an end to the long controversy about the correct name of this "Archimedean game": *stomachion* ("infuriating"), *ostomachion* ("battle of bone pieces"), or *synthemachion* ("collection of scraps"). The word "stomachion" appears in the fragment found by Heiberg.

Fragment from "Stomachion"

"Because the so-called stomachion contains a number of investigations about rearranging the figures it comprises, I deemed it necessary first of all to explain the number of pieces and show which figure each of them is similar to. Then I determined which pairs of angles are needed to create a right angle. Further, I intended to exhaustively examine all conceivable instances in which pieces of the stomachion can abut one another and determine whether the sides of the abutting pieces lie on a single line or whether the deviation from the straight line is so small as to be unnoticeable. These questions require rigorous proofs, and if the deviation from a straight line is small and therefore cannot be seen by the eye, then such incomplete abutment is considered admissible.



"Because any piece can be replaced with another that is equal to it and has the same angles, it is possible to compose many different figures . . . . Sometimes two figures taken together are equal and similar to one figure, or two figures taken together are equal and similar to two figures. This enables us to construct many figures by interchanging pieces of the stomachion."

Archimedean problems

The division of the stomachion figure proposed by Archimedes (fig. 1) has this property: the areas of all 14 pieces are rational fractions of the area of the initial figure.

Problem 1. Prove that the areas of the separate pieces of the initial parallelogram, expressed in 1/48th parts of its area, have the values of the red numbers in figure 1.

The following problem is taken from *Geometrical Puzzles and Paralogisms* (1912) by Emile Fourret.

Problem 2. Group the stomachion pieces such that the areas of the newly composed figures can be expressed in 1/48th parts of the initial parallelogram (a) by three equal integers; (b) by three consecutive integers; (c) by integers from 1 to 8 and by the integer 12.

All three parts of problem 2 can be answered without tearing apart the initial parallelogram—one need only draw the borderlines between the pieces.

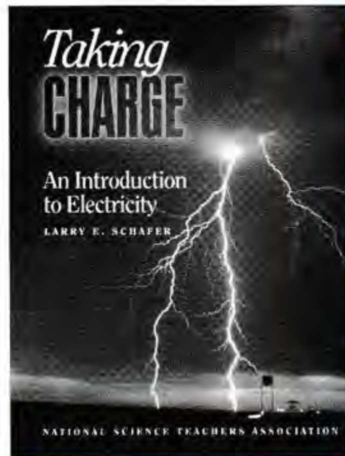
ANSWERS IN THE NEXT ISSUE

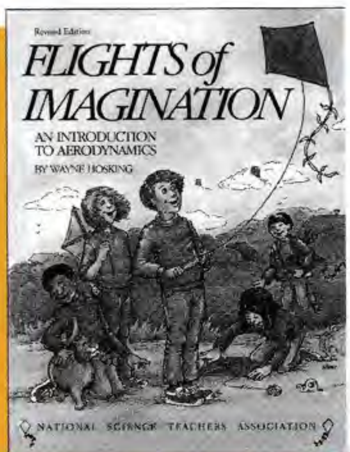
Soar into Science with NSTA!

Taking Charge

An Introduction to Electricity
by Larry E. Schafer

Why do rubbed balloons stick to the wall? What is a short circuit? Spark your interest in electricity with this new book from NSTA. Explore static electricity in the first module before seeing that static electricity can move as current electricity, investigated in the second module. These 25 hands-on activities use readily available materials and make *Taking Charge* a fun exploration of basic electricity. The unique historical approach of the first module shows how we developd our modern ideas of electricity by introducing the ideas of early explorers of electricity like Ben Franklin. In the second module, discover how light bulbs work, and use bulbs and batteries to construct and analyze simple circuits. (grades 5–10)
#PB–96, 1992, 155 pp. $18.95

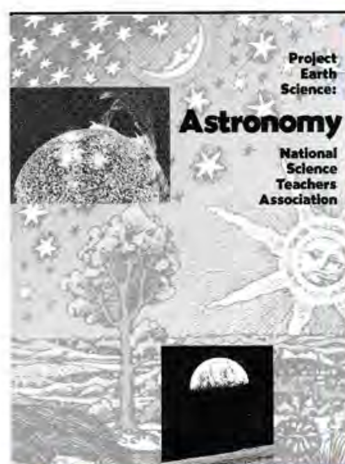




Flights of Imagination

An Introduction to Aerodynamics
Revised Edition
by Wayne Hoskings

Go fly a kite! The clear instructions in these 18 projects transform trash bags, dowels, and tape into high-flying lessons in aerodynamics. Let *Flights of Imagination* add a springtime lift to you study of flight. (grades 5–12)
#PB–61, 1990, 56 pp. $10.50



Project Earth Science: Astronomy

by P. Sean Smith

Why does Earth have seasons? What is a light year? Why does the Moon have phases? The hands-on, teacher-tested activities of *Project Earth Science: Astronomy*, bring the concepts of astronomy down to Earth. Background information, supplementary readings, and suggestions for further study and integrating other disciplines provide a framework to launch a successful study of astronomy. The guiding theme of this book is Earth's uniqueness among the planets of the Solar System. But instead of stating this, *Astronomy* presents concepts and then uses that foundation to examine Earth in the context of our Solar System, as just another planet. In this light, Earth's unique qualities shine through. (grades 5–10)
#PB–90, 1992, 160 pp. $18.50

To order, call 1–800–722–NSTA
MasterCard, Visa, or purchase order

Ad code: Quant Jan