QUANTUM

JANUARY/FEBRUARY 1996

$5.95

Springer

Lion Being Taught by Cupid to Sing (reverse) (1444) by Antonio Pisanello

ONE CAN ALMOST HEAR THE DELIGHTED LAUGHter of 500 years ago, as Antonio Pisanello presents his medal commemorating the wedding of Lionello d'Este and Maria of Aragon. The front of the medal is a portrait of the "little lion," Lionello—a man to be reckoned with. Turn the medal over, and you see what is depicted above. "Music has charms to soothe a savage breast," as Congreve wrote, and when Cupid himself is the music teacher, we can feel confident that Love will indeed prevail.

Although Pisanello also produced frescoes and portrait paintings, his contemporary fame rested largely on his medals. Apparently he learned his craft by studying ancient Greek and Roman numismatic portraits—he had virtually no recent predecessors in this field.

Pisanello's place in the history of art is ensured in part by his drawings. According to the *Encyclopaedia Britannica*, "this is the only instance in which the drawings of a 15th-century workshop have been preserved virtually intact." In addition to providing insight into the drafting techniques of the period, the drawings reveal Pisanello's broad range of interests and his powers of observation. He was one of the first artists of the 15th century to draw from life, rather than copying the work of others in the medieval manner.

Sharp-eyed readers may have noted the date on the medal. Does Pisanello deserve a slap on the wrist for not following the correct style for Roman numerals? Isaak Yaglom discussed the Roman system of notation along with others in the July/August 1995 issue. In this issue, Steven Schwartzman plops us in the middle of a Texas imbroglio involving—yes, Roman numerals! Turn to page 4 to see how justice is meted out.

QUANTUM

JANUARY/FEBRUARY 1996 VOLUME 6, NUMBER 3

Cover art by Leonid Tishkov

He said he wanted "a room with a view." But did he expect a room with a view that looks back? Probably not, but he'll try to make the best of the situation.

"Just one more?" he asks his deciduous friends. "One more little one?" He's obviously pushing the limit. But what is the limit? Say a person wanted a view with trees and a view *through* the trees: what's the maximum number of trees this view-loving, tree-loving person could plant in a given plot of land?

That's the question posed by the article beginning on page 16, "The Orchard Problem."

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

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

So what's the joke?

There's nothing funny about computer viruses

AT ROUGHLY 4 P.M. PACIFIC Standard Time on Monday, December 10, 1995, while I was working on a micro-unit for NSTA's reform project for high school science at my home in Nevada, my PC displayed a message saying that my master boot record had been suddenly changed. Not knowing what had happened, I tried to get some more information. I had "lost" my other hard drive (drive D), and I couldn't gain access to my CD-ROM drive.

The computer seemed to indicate that a SCSI adapter card was disabled, so I ran out and bought a new one and installed it. When I tried to start up my PC again, it crashed completely. With the help of my son Richard, an engineer and computer expert, who stayed "with me" on the telephone for seven hours, until midnight, we were able to determine that the problem was a computer "virus" (which shall remain nameless). My computer had not been protected by antivirus software, so I rushed out and bought some. We managed to use a boot disk and this software to destroy the virus and rebuild my boot sector. Then, using some driver software to restore the CD-ROM drive, we were able to re-install a certain well-known operating system (yes, the one that was upgraded with such fanfare back in August and is named for the year of its rebirth).

The D drive, which contained all of my work files—some 600 megabytes—had been repartitioned into gibberish. Fortunately we could reformat that drive and restore everything from a tape backup I had made in September. I also had a December 2 tape backup of my reform project files and could therefore get everything except what I had produced during those missing eight days.

Careful analysis of various diskettes I had used within a day or so of this viral infection showed that this particular virus was indeed present on four diskettes sent to me from one of my project centers and on one diskette sent to me from NSTA headquarters in Arlington, Virginia. It's not clear which one gave my computer the virus, but naturally I have eradicated the virus from all my diskettes and am *now* running an antivirus program on my computer at all times.

One can draw several conclusions from this terrible experience. First, and most obvious, one should always protect one's computer against these nasty bits of code so aptly nicknamed "viruses."

Another conclusion is more troubling and less amenable to a quick and easy fix. It seems there is a deep-seated urge in human beings—or is it human males (since most if not all of the perpetrators are boys or men)?—to destroy things gratuitously. Where does this strange, anonymous animosity come from? Will the human genome project find a special "destructive" gene that causes a person to do nasty things to strangers? Will it be a rare mutation of a normal gene, or will it be a perfectly "healthy" part of our genetic makeup?

In the meantime, I can't help but wonder how anyone could get pleasure or satisfaction from creating and broadcasting a computer virus that does this sort of damage. I was giving all my time and effort to create materials to help young people learn science. These materials make science much more interesting and accessible to more students. Whoever is responsible for this virus caused a serious disruption in the creation of materials that some 6,000 high school students were waiting to receive. What possible benefit could someone gain from impeding this process?

Of course, this person had no idea who would suffer from his act. Perhaps a hospital somewhere lost its computer system and several people died as a result—who knows? That's what makes computer viruses so unnerving: the victims are as anonymous to its creator as the creator is to his victims. It's a kind of random, faceless evil that seems peculiar to our times.

Bright young people like challenges. I can understand this. It takes a good mind and a clever strategy to devise a powerful computer virus. But obviously it's not enough to be smart and clever. There is a moral dimension to science that some choose to ignore. Many who worked on the creation of nuclear weapons lived to regret it. Why? They learned that just because it's *possible* to do something doesn't make it right to *do* it. In the not so distant future it may be possible for a high school student to create a *real* virus, one that injures or kills *real* people. How could we prevent it? Where does one acquire the moral strength it takes to resist the temptation to show off how cleverly destructive one can be? Is it possible to "teach" this kind of virtue?

I would hope readers of *Quantum* can find ways to use their intelligence constructively—for instance, to create computer vaccines that protect computers from viruses before they even have a chance to enter a PC's RAM or storage media, let alone replicate and destroy data. Computers are powerful devices that can extend your mind and vastly increase your efficiency. They can also separate you from the real world in undesirable ways. I suspect that people who spend their time creating computer viruses are among those who have become addicted to the computer and to "virtual reality" and to that form of "hacking" that might better be termed computer vandalism. Use the PC as a tool, not a crutch. Who needs the "glory" of having one's virus named in an antivirus program? Make your mark in the *real* world, and let *your* name be attached to everything you do.

—Bill G. Aldridge

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

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

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

Advertising:

MPA

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

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

Subscription Information:

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

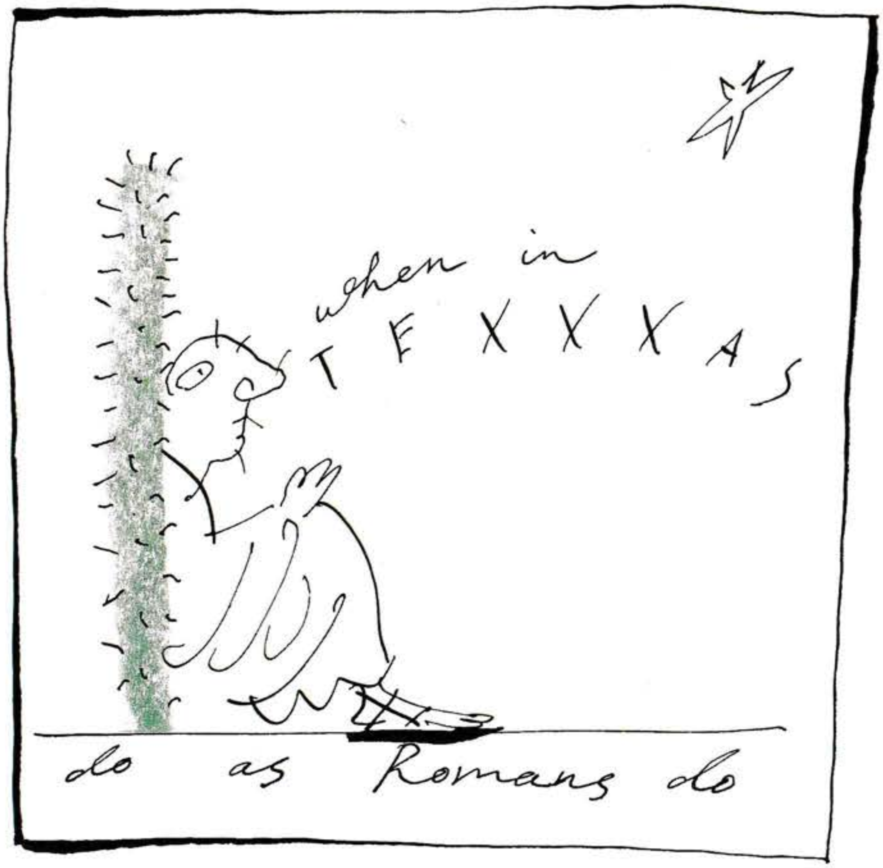
All Countries Outside North America: Personal rate is US$28 including surface delivery charges; institutional rate is US$39; airmail delivery is US$13 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Who owns Roman numerals?

And who knows how to write them correctly? (What is *"correct," anyway?)*

by Steven Schwartzman

THE *AUSTIN AMERICAN–Statesman*, the daily newspaper of the Texas capital, used to carry a column written by Ellie Rucker. Her column, which appeared several times a week, usually published letters asking how to locate a hard-to-find product or service, how to recycle an item, and so on. A letter that appeared in print on March 13, 1990, however, was different. It said that 24,000 people in San Marcos, a town 30 miles south of Austin, were wondering how to write the year 1990 in Roman numerals. Presumably the year was to be carved in stone on a building or monument, and the good townspeople of San Marcos wanted to be sure they got it right.



Art by Sergey Ivanov

The response that appeared directly beneath the letter said that the shortest way to write 1990 is MCMXC, but that the year can also be written MDCDLXXXX, or even MCDCXC [sic—see below]. The column concluded with a note attributing those answers to researchers at the Perry–Castañeda Library, the main library at the University of Texas in Austin.

The original letter and the three answers triggered a barrage of follow-up letters far out of proportion to the seeming simplicity of the question. Those letters revealed that many people today have little knowledge of how Roman numerals were actually written in ancient times. The letters also revealed that many people have strongly held and often contradictory opinions about how Roman numerals are most commonly written now or "should" be written now.

The majority opinion seemed to be that brevity is the soul not only of wit but of Roman numerals as well. Here are some quotations in that vein that appeared in Ellie Rucker's follow-up column on March 26:

> The correct way is the shortest way, period. I learned this when I was a Latin student. The shortest way is MXM.

I saw this on the Charles Karalt [sic] show three weeks ago.

* * * * *

The purpose of the Roman numeral system is to have the shortest amount of numbers which would be MXM.

* * * * *

Ellie, you have to retract your Roman numeral column. The shortest way to write 1990 is MXM and you would never have four X's in a row.

Other people felt that the longest way is the best. Two writers considered both possibilities, and one of them also corrected the newspaper's apparent typo:

Ellie, I'm a Latin teacher. The Romans would say the longest way of writing 1990 is the best. But most of the time we go with the easiest. The correct way is MXM.

* * * * *

MXM would be the shortest way to write it, MDCCCCLXXXX, the longest. MCDCXC, near as I can tell, equals 1590 (I think you meant MDCDXC).

One letter took a different approach:

It frightens me (being European) how misinformed people are here about things children in Europe master in the second grade. There is only one way to write 1990 in Roman numerals—the first one cited by you: MCMXC. Not because it is the shortest, but because there is as much logic in Roman numerals as in Arabic numerals. . . . [Y]ou *never* write 9 as VIIII or IIIIIIIII, but always IX.

This was followed by paragraph after paragraph of explanations.

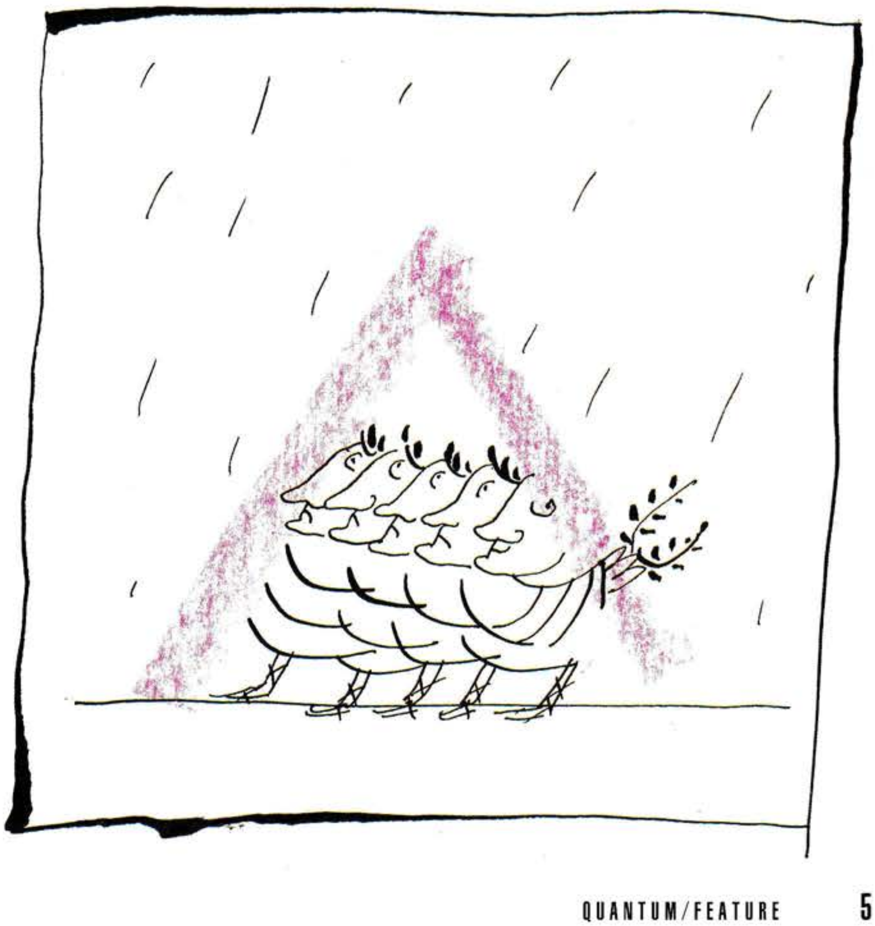
Interestingly, the most popular opinion—that the best way to write a number in Roman numerals is the shortest way—not only conflicts with other opinions but isn't even consistent with itself. Almost certainly no one putting forth that opinion would write IIX for 8 or XXC for 80, even though IIX and XXC are indisputably shorter than the now universally accepted VIII and LXXX. Similarly inconsistent is the opposing view: it's hard to imagine anyone claiming that the best way to write 100 is a string of one hundred I's. Perhaps what people with the longest-is-best view meant is that, of the three answers given in Ellie Rucker's column, the longest one was the best.

Proponents from both camps would probably be surprised to learn that the ancient Romans were inconsistent in the way they wrote their numerals. Sometimes they really did write IIX for 8 and IIXX for 18. To represent 9 they were as likely to write VIIII as they were to write IX. They usually wrote 4 as IIII (a practice, by the way, that Renaissance clock makers continued and that even many modern clock makers observe).

The development of Roman numerals

Given the confusion in many people's minds, let's take a look at how Roman numerals developed in ancient times and how the system was gradually standardized. As with so many scripts around the world, the basic unit was represented by a single stroke. For the Romans, unlike the Chinese, that stroke was a vertical one. To represent two units the Romans wrote two strokes, for three units three strokes, and for four units four strokes. Those vertical strokes could easily be scratched on the ground or carved on wood or stone. When modern Americans who are keeping a running tally advance from four strokes to five, they frequently place the fifth stroke diagonally across the other four. Historians like Menninger and Ifrah hypothesize that the ancient Romans followed a similar practice, indicating a collection of ten tally marks by two diagonally crossing lines: X. Half of that symbol would then have given rise to the Roman numeral for five. Historically speaking, the ancient Italic peoples sometimes represented five with the bottom half of the X, and at other times with the top half, which eventually won out and gave rise to the familiar V. It is also possible that the V symbol originally represented the notch between the thumb and forefinger of a five-fingered hand that was held vertically.

Karl Menninger has proposed that the Romans, who like other Indo-European tribes were acutely aware of the first few powers of ten, initially represented the number 100 by a doubly crossed vertical stroke indicating ten tens: ⋇. The upper half of that symbol, ⩛, would



then have come to represent 50. Given the inconsistency (some would say creativity) of the way cultures often use written symbols, the symbol for 50 appears in ancient carvings as ↆ and ⊥, which most likely evolved at last to the standardized L. The presumed original symbol for one hundred, Ж, was later replaced by C, which was the first letter of the Latin word for one hundred—*centum*.

The earliest Roman symbol for one thousand seems to have been something like ϕ, and may in fact have been the Greek letter phi, which represented a sound that didn't exist in Latin and could therefore be used as a nonphonetic symbol. Whatever its origin, that early symbol for one thousand came to be written in many ways, some of which scarcely resembled the original. Georges Ifrah shows two dozen variants. One variant, ↀ, was in use as early as the third century B.C., as evidenced by stone inscriptions of the period. It's easy to see how the right half of that symbol could have given rise to the use of D to represent 500. In archaic Latin, the D had been preceded by Ð, so that the original symbol for 1,000 may actually have been something like ⊕.

Several of the variant symbols for one thousand bore a sort of "loosely topological" similarity. Among them were ϕ, (|), CIↃ, /|\, and CIↃ. The underlying pattern is that of a vertical line flanked by two symmetric strokes that "bulge away from" the central line. The Romans extended that pattern by adding one extra "layer" for each higher power of ten. A panel found in the ruins of Pompeii, which was buried by an eruption of Mt. Vesuvius in A.D. 79, represents the number 10,000 with the symbol ↂ. Similarly, the Romans occasionally used ↈ for 100,000. An inscription on the so-called Columna Rostrata, which was erected in Rome in commemoration of a naval victory over the Carthaginians in 260 B.C., showed a slight variant of that symbol thirty-four times in a row. The symmetric outermost strokes were extended to the center to produce a symbol that looked like this: ((((|)))).

The fact that the symbol was repeated so many times indicates that the Romans had no symbol for a million. It isn't hard to understand why. Extending the multiplicative pattern tenfold by adding another "layer" to the symbol for 100,000 would have resulted in a symbol with so many strokes that it would have been difficult to interpret quickly. Also, similar symbols with different numbers of strokes could easily have been confused. What's more, the Latin language had no single word meaning "a million." When the Romans needed to refer to one million, they said "ten times a hundred thousand." To refer to two million they said "twenty times a hundred thousand," and so on.

Since 100,000 became the stepping stone to all higher numbers for the ancient Romans, they developed a simplified symbol to represent it. The new form may have developed from the overarching outer layer of the symbol ((((|)))), or it may have been freshly created. In any case, when what we might call a rectangle without a bottom was placed around a Roman numeral, the value of that numeral was multiplied by 100,000. So, for example, $\overline{|\mathrm{II}|}$ meant 200,000, $\overline{|\mathrm{X}|}$ meant 1,000,000, and $\overline{|\mathrm{C}|}$ meant 10,000,000.

Menninger reports that the ancient Romans almost always used ϕ or one of its variants to represent 1,000. On rare occasions they might write IIM to mean 2,000, where the M was the first letter of *mille*, the Latin word for 1,000. Only in the Middle Ages did the earlier symbols for 1,000 begin to be replaced by the now familiar M, and the process wasn't completed until much later. When Descartes published his famous *Discourse on Method* in Leyden in 1637, the date appeared on the title page as CIↃIↃCXXXVII.

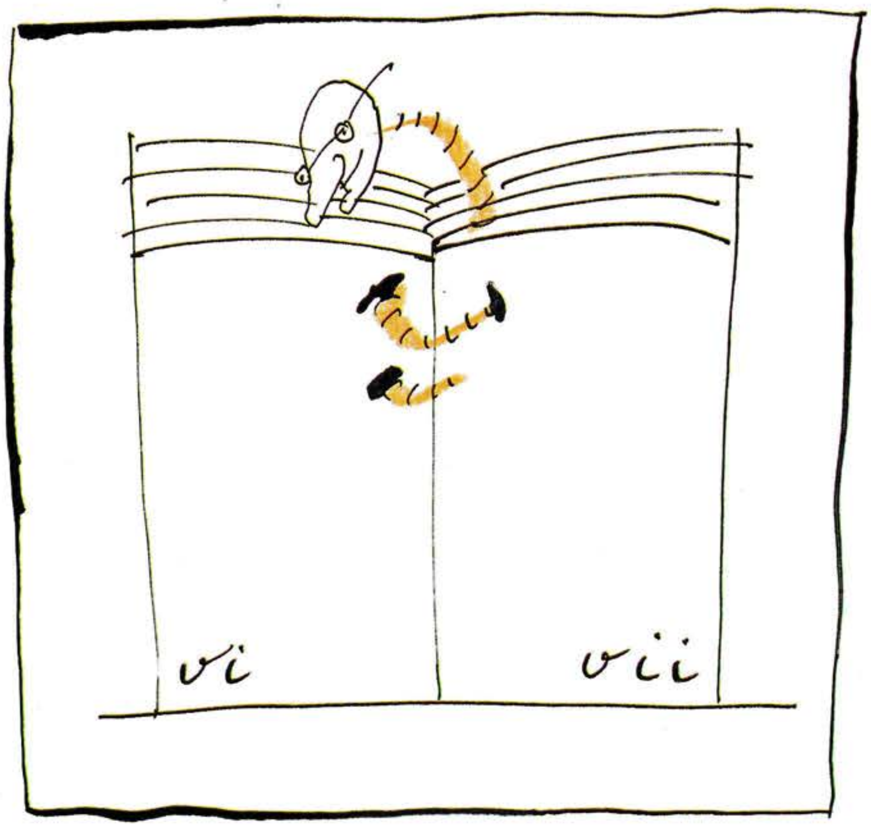
The Middle Ages and Renaissance also saw a continuation of the ancient Roman practice of placing a bar over a numeral or part of a numeral to indicate that it was to be multiplied by 1,000. For example, $\overline{\mathrm{VI}}$XXXIX represented the number 6,039. (That bar was known in Medieval Latin as a *titulus*, which is why our borrowed word "title" refers to a caption placed over the text of an article.)

In contrast, the Romans, as well as Medieval and Renaissance writers, sometimes placed a bar over one or more letters not to indicate multiplication by 1,000 but to indicate that the characters were to be read as a numeral rather than as a word. After all, an I could be the number one or the letter I; a V could be the number 5 or the letter V; and so on. Also, many writers weren't fastidious about leaving sufficiently large (or even any) spaces between words. The overbar that was used to distinguish numerals from letters gave rise to our relatively recent habit of adding an accompanying underbar, so that we now often write a number like 15 as $\underline{\overline{\mathrm{XV}}}$. As greater numbers of people learned to write in the Middle Ages and Renaissance, there was an increasing tendency to write Roman numerals in lower case rather than as traditional capitals, which the ancient Romans used exclusively. The publishing industry often still follows that tradition by numbering the pages of a book's preface with lower-case Roman numerals.

Varying interpretations

The system of Roman numerals once had a natural constituency—the people of Rome and the Roman Empire. But the numeration system was no more regulated by any sort of governing authority than was the Latin language itself. Once the Roman Empire disintegrated, the natural constituency for Roman numerals disappeared. Just as there had been conflicting ways of writing numerals among the ancient Romans, there were competing methods of writing Roman numerals in the Middle Ages, the Renaissance, the Age of Enlightenment, the nineteenth century, and even today.

As an example of that lack of uniformity, consider *The New University Arithmetic*, written by Charles



Davies and published in New York in 1858. In a table of Roman numerals on page fifteen, the numbers 4 and 9 are given as IV and IX, and 40 and 90 appear as XL and XC; yet 400 and 900 are given as CCCC and DCCCC. The asymmetry mimics that of the ancient Romans themselves, who regularly wrote CCCC and DCCCC rather than CD and CM.

Even today statements appear in textbooks that, while not necessarily false, are worded in such a way as to leave students with the wrong impression. For example, in *Mathematical Ideas* by Miller and Heeren (1986), we find the statement that "9 was written IX (10 minus 1) rather than VIIII (5 plus 4)." The implication is that the Romans always wrote IX for 9, when in fact they were as likely to write VIIII. The next sentence in this book is a little better: "The same method was often used to write 4 as IV rather than IIII." Still, it would be more accurate to replace the word "often" with "sometimes."

Miller and Heeren go on to say that "[t]he Romans saved space by using subtraction." While the subtractive principle certainly can save space, many readers interpret the quoted sentence to mean that the Romans used the subtractive principle in order to save space. If the Romans had been concerned about saving space, they wouldn't so often have written numbers like VIIII for 9, XIIII for 14, XXXX for 40, and LXXXXV for 95, all of which are found in ancient inscriptions. What is more, the Romans never wrote a number like 95 as **VC**, which would have been much shorter than the attested LXXXXV. (Here and in the text that follows, I have put incorrect or unattested forms in boldface.)

A plausible explanation for the adoption of the subtractive principle is the natural tendency to quantify something by referring to the closest "landmark." For instance, we are more likely to say that the time is "five to nine" than we are to say it's 8:55 (although that may change with the proliferation of digital timepieces). The *Encyclopaedia Britannica* (1910 ed.) also cites the limited ability of the human brain to directly apprehend the number of objects in a group:

> It has been suggested that as many as six objects can be seen at once; but this is probably only the case with few people, and with them only when the objects have a certain geometrical arrangement. The limit for most adults, under favourable conditions, is about four. Under certain conditions it is less. Thus IIII, the old Roman notation for *four*, is difficult to distinguish from III, and this may have been the main reason for replacing it with IV.

Whatever its origins, the Romans made intermittent use of the subtractive principle. The Latin word for 20 was *viginti*. During the height of Roman civilization the words for 18 and 19 were *duodeviginti* (literally "two from 20") and *undeviginti* ("one from twenty"), although at other periods in their history the Romans used *octodecim* for 18 and *novemdecim* for 19. Nevertheless, although the Latin word for 10 was *decem*, the Romans never wrote ***duodedecem*** for 8 or ***undedecem*** for 9.

In the MCMXC's

My curiosity was piqued by people's apparent unfamiliarity with Roman numerals, so I went to the Perry–Castañeda Library (the same one Ellie Rucker referred the original question to) and looked at some of the arithmetic books currently under adoption in Texas. The series I found, all copyrighted 1991, were *Mathematics in Action* (McGraw Hill), *Mathematics* (MacMillan), and *Exploring Mathematics* (Scott, Foresman—written by a whopping 25 authors). All the books are bright and colorful, all are filled with lots of pictures, and not one bears the word "arithmetic" in its title.

Roman numerals appear in all the series in either the third or fourth grade; only the Macmillan series continues to present Roman numerals through junior high school. Most revealingly, and perhaps most disturbingly, all three series seem to treat Roman numerals as an optional topic, as evidenced by the large word "enrichment" printed at the top of each pertinent page.

The MacMillan series distorts ancient history: "Notice that the

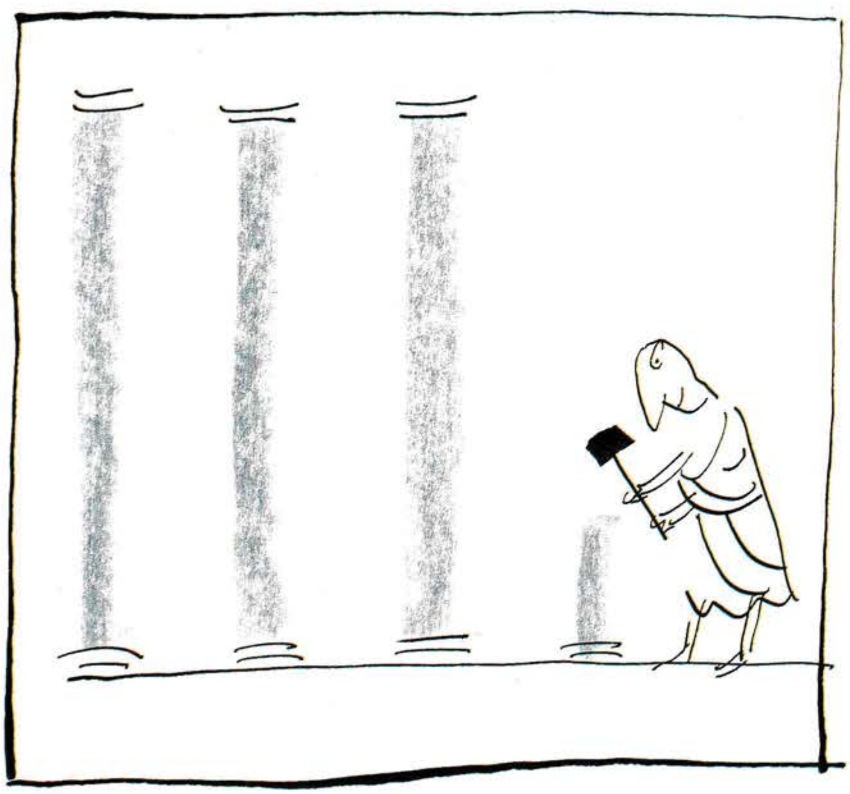
Romans never wrote any letter more than three times in a row" (Grade 4, p. 45). "The Romans never wrote any letter more than three times in a row" (Grade 6, p. 37). A better statement, but still not the whole truth, is: "The Romans did not have to write any letter more than three times in a row" (Grade 7, p. 32). They didn't have to write IIII for 4, but they usually did. The MacMillan series also repeatedly but falsely claims that "the Romans used letters to name numbers." The Romans used letters, as well as other symbols, to write numbers. Presumably they always "named"—that is, pronounced—the Latin words corresponding to the numerals they wrote.

The Scott, Foresman series is more accurate, but that accuracy is reserved for its teacher's editions. One of them (Grade 4, p. 64) says: "You might wish to tell students that the ancient Romans usually used addition only to find the value of a number. For example, the number 4 was almost always written as IIII. Subtraction came into use much later as a means of expressing Roman numerals. By using subtraction, Roman numerals can be written in less space." Unlike the other two series, the Scott, Foresman books don't explain to students when to add and when to subtract within a Roman numeral. Instead, teachers are encouraged to elicit the rule from the students.

Among the three sets of books, the McGraw Hill series devotes the least amount of space to Roman numerals. The book for grade 4 gives a full page to the topic. The grade 5 book gives it half a page. In both books the topic is preceded by the paradoxical heading "Enrichment for all": if all students are to be exposed to the topic, why is it "enrichment" and not part of the prescribed curriculum? The grade 6 book offers only a box occupying about a third of a page. The box bears the redundant heading "Logical Reasoning."

Standardization

Given the precarious position of Roman numerals in America today, I suggest that we follow the lead of the book industry. Publishers have adopted a system of International Standard Book Numbers, or ISBN's, so that every book printed anywhere in the world will have a unique number. In a similar way, I propose that we use the designation ISRN, or International Standard Roman Numerals, to designate the system that Americans used to learn in school, and which the correspondent quoted earlier maintains that Europeans still learn. In such a system a given number (at least up into the millions) will always be written as a unique Roman numeral. Although many of us have internalized the rules of ISRN by example, stating the rules isn't as easy as one might think. Before reading further, you might want to see if you can write a set of rules that allows every form that should be allowed but screens out all unwanted forms.



Here is a set of proposed rules:

1. The acceptable characters, in ascending order, are I = 1, V = 5, X = 10, L = 50, C = 100, D = 500, and M = 1,000.

2. No character can appear more than three times in a row.

3. In general, characters appear in a monotonically decreasing sequence. The values of the characters are added.

4. By exception, a single character representing a power of 10 can appear immediately before a nonrepeated character whose value is no greater than the next highest power of 10, and in such a case the value of the pair will be obtained by subtracting the value of the lesser character from the value of the greater character. (In other words, I can appear immediately before V or X; X can appear immediately before L or C; C can occur immediately before D or M.)

5. A half-step character (V, L, D) cannot be repeated two characters later. (In other words, 9 is represented by IX rather than **VIV**, 90 is represented by XC rather than **LXL**, and 900 is represented by CM rather than **DCD**.)

6. A bar may be placed over a well-formed character or group of characters, beginning from the left end of a numeral; the bar multiplies the value

of the characters beneath it by 1,000.

In practice, it might be easier to teach by example. The chart below contains all the elements needed to write Roman numerals up to 3,999. Since ISRN's are normally used only for page numbers and years, 3,999 is certainly sufficient for the next couple of millennia. For 4,000 and above the system requires the use of an overbar.

Someone wanting to write a number like 2,937 in Roman numerals would rewrite the number using place-value notation: $2 \times 1{,}000 + 9 \times 100 + 3 \times 10 + 7$. The equivalent of each term is chosen from left to right in the table below: MM + CM + XXX + VII. Those values are then strung together to form the final Roman numeral: MMCMXXXVII.

Notice the parallel structure exhibited in the table. In a given row, each entry is written with the same number of characters. As you move from one column leftward to the next, each character is replaced by the character worth ten times as much, and the new characters remain in the same order as those they replace.

Conclusion

The notion of the correct Roman numeral being the briefest isn't true, either historically or in ISRN notation. Although most Americans seem to feel that the shortest way of representing an integer by a Roman numeral is the best or only correct way, they are mistaken: for every instance of something like an IX being shorter than the non-ISRN-compliant VIIII, there is an occurrence of a VIII being longer than the non-ISRN-compliant IIX (which we have seen the Romans themselves sometimes wrote).

Something akin to the ISRN system was once taught in American public schools as part of the standard curriculum (I remember learning it in elementary school in the 1950s), but I can confirm from my experience as a teacher over the last 25 years that many recent American high school graduates don't know the system. The letters quoted at the beginning of this article, and the fact that the topic of Roman numerals is currently treated as "enrichment" in arithmetic books, offer further confirmation that the system is no longer as widely known as it once was. We need to ensure that our students learn Roman numerals in elementary school and that they review the subject periodically in secondary school.

Finally, although the people who wrote to Ellie Rucker in 1990 proposed various ways of representing 1990 in "Roman" numerals, the surprising truth is that an ancient Roman would probably not have known how to decipher any of them! ◘

number	thousands	hundreds	tens	ones
1	M	C	X	I
2	MM	CC	XX	II
3	MMM	CCC	XXX	III
4		CD	XL	IV
5		D	L	V
6		DC	LX	VI
7		DCC	LXX	VII
8		DCCC	LXXX	VIII
9		CM	XC	IX

Steven Schwartzman *teaches mathematics part-time at Austin Community College and is the author of* The Words of Mathematics, *published by the Mathematical Association of America.*

A walk on the sword's edge

Ancient wisdom meets modern science

by V. Meshcheryakov

ONCE THE POPULAR RUSSIAN newspaper *Moskovsky Komsomolets* published an intriguing essay by Alexander Pogonchenkov, "Somersault on a Sword's Blade." It told the story of a sixty-year-old wise man with the body of an athlete and his twenty-year-old wife. The gentleman, as it happens, was an expert in the martial arts, and he put on a marvelous display for the reporter. Here's an excerpt: "He stands on the blade and, holding his young assistant's hand, begins to walk along the sword. One step . . . then another . . . and another . . . Now he turns . . . takes a step back . . . and jumps! One expects to see blood spurting from the soles of his feet. But nothing of the sort happens."[1]

What's going on here? The answer isn't all that simple, and it's not devoid of interest. We can make preliminary estimates of the area of contact, the pressure, the number of supporting atoms. But can we hope to approach such exquisite harmony by way of prosaic numbers? You be the judge.

First we'll try to examine the physical side of this unique experiment by taking a rather simple approach. Measuring the area S_0 of the sole of a person's foot (say, by drawing its outline on graph paper), we get $S_0 \sim 10^{-2}$ m^2. If the mass of a human being is $M \sim 10^2$ kg, we get a force of $F = Mg \cong 10^3$ N, which is the external load applied to the area S_0, if the person is standing on a flat surface. In this case the pressure is

$$P_0 = \frac{F}{S_0} \sim 10^5 \text{ N/m}^2 = 1 \text{ atm.} \qquad (1)$$

When a person stands on the blade of a sword, the same load is applied to a far smaller area of contact. To estimate this area, recall the fact that when metals are polished and ground, one can decrease the height of surface irregularities down to fractions of a micron. So let's assume that sharpening the blade results in a blade width of $b \sim 10^{-7}$ m. Taking the length of a person's sole to be $a \sim 10^{-1}$ m, we get the area of the sole that meets the blade: $S_1 \sim 10^{-8}$ m^2. Therefore, the pressure is

$$P_1 = \frac{F}{S_1} \sim 10^6 \text{ atm.} \qquad (2)$$

This estimate falls within the range of pressures that would be encountered, for instance, at the Earth's core, beneath many kilometers of crust. This results sets us back on our heels. Can human beings really stand on a blade that's "razor-sharp" without slicing the muscle tissue in their soles? Or is it just a trick? Maybe the perpetrator has glued on thin, extremely tough soles, or maybe it's some sort of mass hypnosis.

Let's try another approach. We know that human beings are composed of atoms. So it's appropriate to ask: how many atoms does a person stand on? To answer this question, let's first estimate the number of atoms in the contact areas S_0 and S_1.

The characteristic volume corresponding to 1 atom in condensed states of matter is about $\Omega = 10^{-29}$ m^3. Thus, one surface atom makes an elementary surface with an area $s \sim \Omega^{2/3} \sim 10^{-20}$ m^2. So the number of atoms in a surface area S_0, corresponding the area of a person's footprint on the ground, is

$$n_0 = \frac{S_0}{s} \sim 10^{18}, \qquad (3)$$

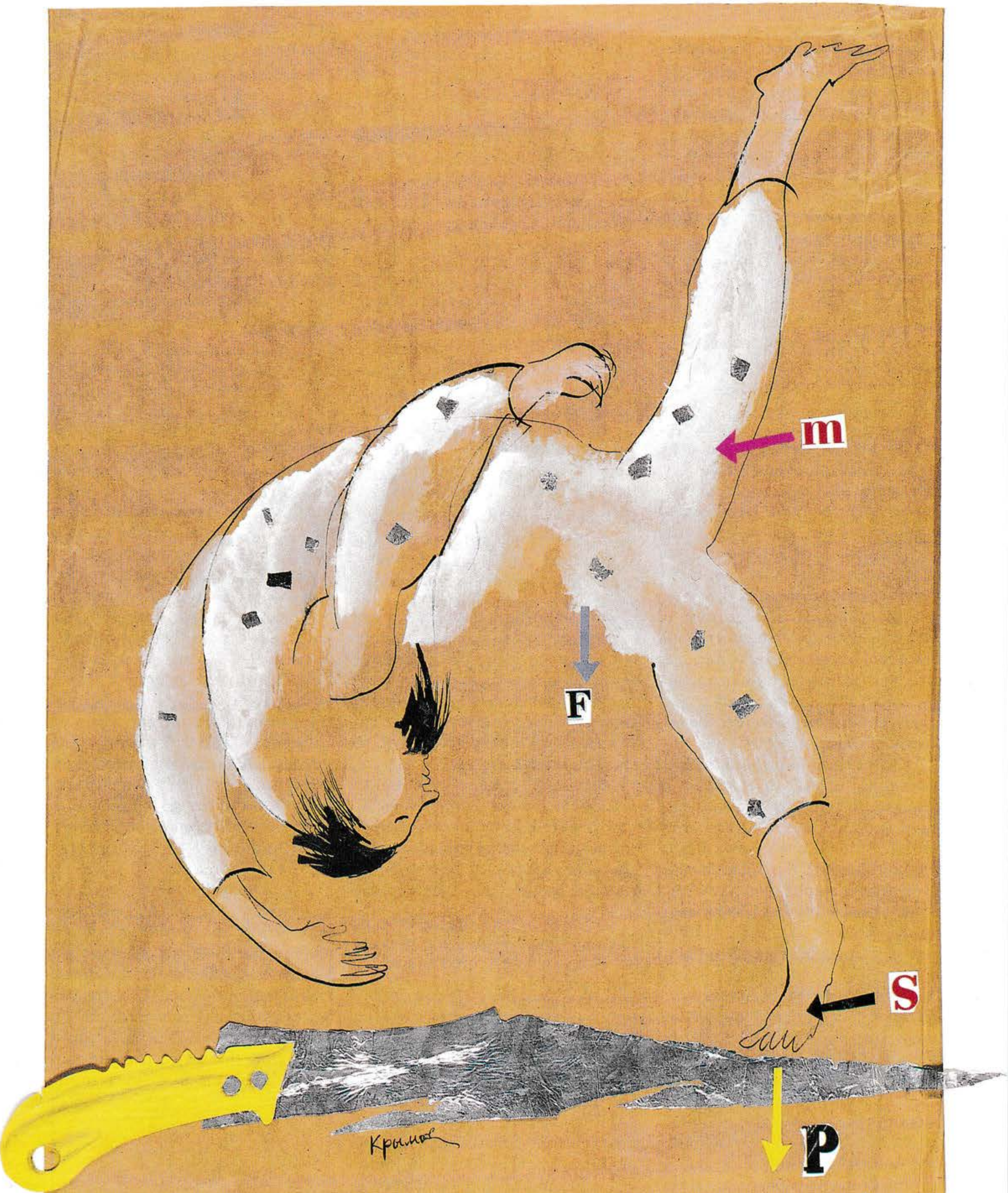
while the number for a person standing on a blade is much smaller:

$$n_1 = \frac{S_1}{s} \sim 10^{12}. \qquad (4)$$

We see that stepping from the ground onto a sword is equivalent to replacing 10^{18} support atoms with a mere 10^{12}. This is a little disappointing.

[1]Need we say it? Don't try this at home!—*Ed.*

Art by Dmitry Krymov


m
F
S
P
Крымов

The difference between these two numbers is hard to imagine. But do n_0 and n_1 really reflect the number of support atoms? Estimates (3) and (4) correspond only to the outline of the contact surfaces. To emphasize this, I described how to draw an outline of the foot on graph paper, at the risk of irritating the savvy reader. But where is the guarantee that all the atoms within the outline are support atoms?

Estimates (3) and (4) don't tell us how far the number n_1 can be decreased. In other words, what are the critical values of the number of atoms n_1 (and, correspondingly, the blade width b) beyond which it is impossible for a person to walk on a sword?

To settle the question, let's consider an idea proposed by the English physicist E. Tomlinson in 1929. The gist of the idea is that contact between bodies is made by a limited number of atoms, depending on the applied stress. Tomlinson called these atoms "caryatids," after the Greek columns carved in the shape of female figures. The number of caryatids is the basic characteristic of contact between solid bodies, and we should estimate it and compare the result with the numbers n_0 and n_1. To do this, let's consider the following physical model.

Imagine that a solid body with an arbitrary surface microrelief is brought up to a static, absolutely rigid support such that the body "touches" it with one of its protruding "soft" atoms. Acting on a solid body with a small force F directed perpendicular to the support area, one can produce an elastic deformation of the surface projection containing the caryatid atom. When this force F exceeds a certain critical value f (depending on the character of the interatomic bonds), the atom makes a diffusion transition. In other words, this atom will be "squeezed out" from the outer (first) atomic layer of the contacting surface. The diffusion disruption of the first atomic layer, and of the subsequent layers, will occur until the number of caryatids is equal to

$$n = \frac{F}{f}. \qquad (5)$$

The critical force f must not exceed the value of the energy E needed to move the atom divided by the interatomic distance:

$$f = \frac{E}{\Omega^{1/3}}. \qquad (6)$$

The value E is also known as the energy of diffusion. For a rough estimate of E, we can assume that it is equal to the electrostatic energy for a single atom:

$$E \sim \frac{ke^2}{r} \sim \frac{ke^2}{\Omega^{1/3}}.$$

A characteristic value for metals is $E \sim 10^{-18}$ J. Inserting formula (6) into formula (5) yields a formula for estimating the number of caryatids:

$$n = \frac{F\Omega^{1/3}}{E}. \qquad (7)$$

It is important that n does not depend on the area of the contact surface. It agrees with Tomlinson's reasoning, which has been confirmed experimentally by a linear dependence of frictional force on external load for the interaction of two solid bodies ($F_{fr} = \mu N$).

Formula (7) makes it possible to determine the real contact area S_r, which should be equal to the product of the number of caryatids n and the elementary area s. Using formula (7) we get

$$S_r = ns = \frac{F\Omega}{E}. \qquad (8)$$

Clearly S_r (just like n) is determined only by the atomic properties of the body deformed on a rigid platform and by the external load.

And now the time has come for the author himself to step on the sword's blade, so to speak—to do the quantitative estimates and draw the qualitative conclusions from them. Using formula (7) and the characteristic values for the atomic volume Ω, external force F, and energy E, we obtain an estimate of the number of caryatids:

$$n \sim 10^{11}. \qquad (9)$$

Thus, the weight of a person is balanced by the force due to the contact of caryatids—about 10^{11} of them. This number depends neither on the shape nor the area of the support, and it does not depend on which part of the human body is touching the Earth. This value is seven orders of magnitude less than the number of atoms contained in the outline of a person's sole:

$$\frac{n_0}{n} \sim 10^7. \qquad (10)$$

In other words, formula (10) shows that only one of every 10^7 atoms of the sole's surface "works" as a caryatid. But perhaps the most interesting thing here is that this number of caryatids isn't established all at once but gradually, as the surface layers of skin are destroyed by the rearrangement of their atoms. Imagine that you press your finger against a table with a force of 1 N. The critical force needed to tear an atom on the finger's surface away from its neighbors is $f \sim 10^{-8}$ N, which follows from formula (6) and the assumed values for E and Ω. Evidently, this force of 1 N breaks the interatomic bonds catastrophically, which eventually leads to the establishment of an equilibrium number of caryatids. This number is $\sim 10^{-8}$ for the case we're examining.

Comparing formula (9) with the number of atoms n_1 inside the support outline on a sword's blade, we can say that the blade is wider than a blade corresponding to a tight arrangement of the equilibrium number of caryatids (from formula (9)). We reduce the width of the blade to $b_r = S_r/a$ without violating the condition of formula (9). Using formula (8) we get

$$b_r = \frac{F\Omega}{Ea} \sim 10^{-7} \text{ to } 10^{-8}\,\text{m}. \qquad (11)$$

From these estimates it follows

that static equilibrium of a person on a sword's blade is possible without catastrophic deformation of the muscle tissue if the width of the blade's edge is no less than 10^{-8} m. In this case, stepping from the floor onto a sword is equivalent to replacing caryatids chaotically distributed in the plane by the same number of caryatids tightly packed and aligned in a row such that the number of atoms across the row is $n_\perp b_r/\Omega^{1/3} \sim 10^2$ and the number of atoms along its length $n_\parallel = a/\Omega^{1/3} \sim 10^9$. The product $n_\perp \cdot n_\parallel = n \sim 10^{11}$ corresponds to the estimate in formula (9).

I hope you liked this simple, prosaic explanation of an ancient mystery. However, many people, the author included, consider the harmony of formulas a kind of magic . . . It's a shame, though, that our last conclusion is at odds with nature. In reality, the thermal motion of the atoms erodes the flat areas of the surface of a crystal, leading to the formation of monatomic steps and the disappearance of the ridges, which makes it impossible to sharpen a blade until you get an ordered, faceted line with a width of a hundred interatomic distances. Apparently it's this natural widening b_r that allows us not only to neglect the sharpening of the sword, but also to do a somersault on the extra support area.

Some concluding remarks

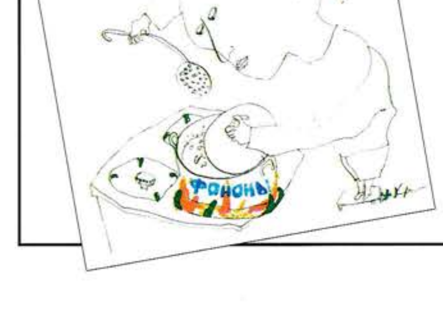
1. The author hasn't heard of any direct experimental measurements that might corroborate or refute the estimates obtained. The method of atomic-force microscopy, which is most relevant to the problem we've examined and is based on the controlled contact of a solid surface with a monatomic needle, still doesn't allow us to study the dynamics of the formation of multiatomic contact.

2. I must admit that the estimates presented are rather approximate. (All the more reason to resist the temptation to try this feat!) First, they're based on the properties of an "average" element in Mendeleyev's table, and so they don't take into account the elementary structure of human skin. Second, due to rounding errors, the values obtained by raising the atomic volume Ω to the 1/3 or 2/3 power may differ from the exact value by an order of magnitude. Because of the circumstances described in point 1 above, it's pointless to try to improve the accuracy of the estimate.

3. One shouldn't infer from this article that a person can safely support just any part of the body on the blade of a sword. Different areas of the skin and underlying muscle tissue differ in their elasticity and durability, which we didn't take into account. Perhaps the skin on the sole of the foot has some advantage over other tissue in this risky experiment conducted at the edge of human capacities (so to speak).

4. Changes in crystal facets that are dependent on temperature caught the interest of such luminaries as Paul Ehrenfest and Lev Landau, but they were described quantitatively only during the last 10 years. Maybe a specialist will read this article and enrich it with an analysis of this phenomenon.

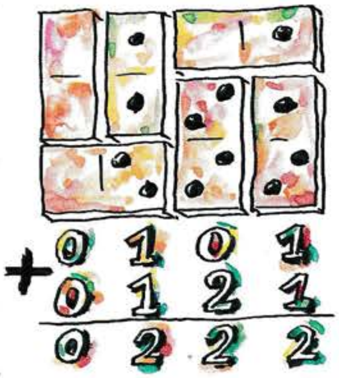
5. The problem of muscle tearing under the action of external forces is much more complicated than the treatment given here indicates. Clearly it's related to the transition from static to sliding friction, an enigma that modern physics has only lately begun to approach. (See the article "Squeaky Doors, Squealing Tires, and Singing Violins" in the November/December 1992 issue of *Quantum*.)



Just for the fun of it!

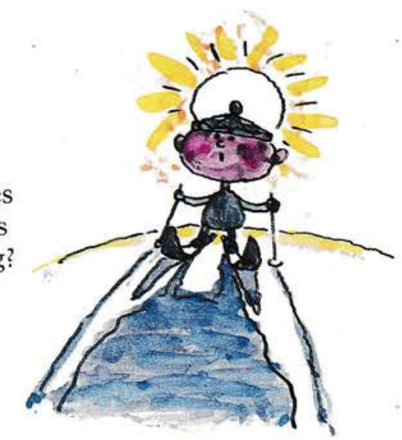
B161

Two-legged, three-legged, and four-legged. There are several three-legged stools and four-legged chairs in a room. If a person is seated on every stool and chair, the total number of legs in the room is 39. How many stools and how many chairs are there?



B162

Domino arithmetic. A 4×3 rectangle is made of six domino tiles such that the dots on the dominoes form a correct example of addition (see the figure for 101 + 121 = 222). Try to find a similar rectangle made from six dominoes (perhaps different from those in the figure) such that the sum—that is, the number in the bottom row—is the smallest. (Tran Quang Dat [Vietnam]).



B163

Shady matters. Everyone knows that the shadow cast by the Sun changes during the day—it's longest at sunrise and sunset and shortest at noon. Is there a place on Earth where a shadow stays the same length all day long?

B164

Numismatic arithmetic. Russian "copper" coins of 1, 2, 3, and 5 kopecks (obsolete nowadays because of inflation) used to weigh as many grams as their denomination. It's known that one of four different copper coins is defective—it differs in weight from a normal coin. How can this coin be singled out using just a pan balance? (G. Tartakovsky)

B165

Tables and napkins. A cafe is furnished with round tables and square tables. The tables of the same shape are the same size; each round table is completely covered by four square napkins equal to each other in size; and each square table is completely covered by four equal round napkins. Show that the diameter of a round napkin is not less than half the diagonal of a square table, and that the side length of a square napkin is no less than the radius of a round table. (V. Proizvolov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 48

Art by Pavel Chernusky


Деревья
растут
ночью,
когда
спят
Тишков 95

The orchard problem

How to plant the most trees and still have a view

by Vladimir Janković

ONCE UPON A TIME THERE lived a mathematically minded gardener named Martin. One day Martin planted an orchard around his small (negligibly small!) house. Following his natural inclination, he planned it very precisely: he marked out a unit square grid with his house at one of its nodes and planted the trees at each of the other nodes within a circle centered at the house. He picked out the trees very carefully, too: they were all of the same species, size, and age, and so they grew at the same rate. At first Martin the gardener could look out his window and enjoy a virtually unobstructed view of the peaceful landscape beyond his orchard. But as time passed the trees got thicker, and one morning Martin woke up to find that the orchard blocked the view completely. Now the question is: given the radius R ($R > 1$) of the orchard, *what is the greatest radius r of the trees*[1] *that still allows Martin to see anything beyond the orchard*? It should be explained that a ray of sight here is considered to be blocked by a tree only if it intersects the tree. So the "last" unblocked rays from the house will touch the sides of at least two trees (of the maximum radius r). An additional question is to determine the directions of these last unblocked rays.

[1]The trees are assumed to be vertical circular cylinders of the same radius.

Art by Leonid Tishkov

This problem has appeared in a number of books—see, for instance, *Geometric Inequalities and Problems of Maxima and Minima* by D. Shklarsky, N. Chentsov, and I. Yaglom (Moscow, 1970, in Russian), *Mathematical Gems I* by R. A. Honsberger (Mathematical Association of America, Washington, D.C., 1973), or *Problems in Plane Geometry*, part 2, problem 24.9 by V. V. Prasolov (Moscow, 1991, in Russian). But they give only a partial answer to the first of the questions posed above. You must prove that for, say, $R = 50$, trees of radius $1/50$ (or, in general, $1/R$ for any R) block the view, whereas trees of any smaller radius do not.

Below I give a complete solution to the Orchard Problem in its general form.

We can think of the "orchard grid" as the grid of integer points (that is, points with integer coordinates) in the coordinate plane with the gardener's house at the origin O. Choose *the integer points outside the orchard closest to the origin whose two coordinates are relatively prime numbers.* Let l be the distance from the origin to any of these points. Then we claim that the largest tree radius that still allows the gardener to see through the orchard is equal to $1/l$, and the last unobstructed rays are those from the origin to the chosen points.

More formally, if l^2 is the smallest number no less than R^2 representable as the sum of the squares of two relatively prime integers, then we will show that

1. For any $r > 1/l$, any ray from the origin O intersects a circle of radius r whose center is at an integer point within the circle of radius R centered at O;
2. For $r = 1/l$, the only rays that do not intersect these circles are those that meet the circle with center O and radius l at points with relatively prime integer coordinates (fig. 1).

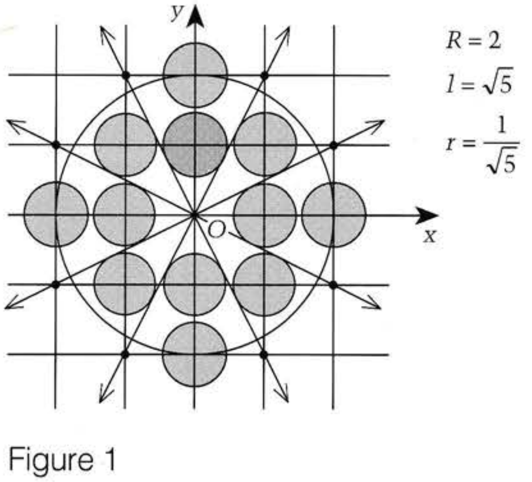


Figure 1

For example, if $R = 50$, we have $R^2 = 2{,}500 < 2{,}501 = 50^2 + 1^2$, so $l = \sqrt{2{,}501}$. It can be shown that there are exactly two representations of 2,501 as the sum of the squares of two relatively prime positive integers: $2{,}501 = 50^2 + 1^2 = 49^2 + 10^2$. Thus, an orchard of radius 50 completely blocks the view from the origin for any $r > 1/\sqrt{2{,}501}$, and for $r = 1/\sqrt{2{,}501}$ there are exactly sixteen directions in which the land beyond the orchard can be seen—they pass through the points $(\pm 50, \pm 1)$, $(\pm 1, \pm 50)$, $(\pm 49, \pm 10)$, and $(\pm 10, \pm 49)$ (where the signs are chosen independently).

Proof. Consider an arbitrary ray h drawn from the origin. By the symmetry of the problem we can assume that it lies in the first quadrant of the coordinate plane. We want to determine the values of r for which it is blocked by a certain tree.

If this ray passes through the center of a tree, it is blocked by this tree for any $r > 0$. So in what follows we'll assume that *the ray has no integer points within the radius of the orchard.*

We can ignore the trees outside the first quadrant, because before they are thick enough even to reach the quadrant ($r = 1$), all trees will touch their neighbors ($r = 1/2$), block everything, and stop growing.

On each side of the ray h in quadrant 1, let us now choose a point of the orchard closest to h. Denote the two points U and V (fig. 2). Clearly, our ray is blocked if and only if it's blocked by one of the trees at U and V. Notice that there are no integer points inside or on the border of the triangle OUV except its vertices, because otherwise point U or point V (or both) wouldn't be the closest to ray h on its side of h. (Recall that we have assumed that ray h has no integer points on it inside the orchard.) This is the crucial point of the proof: it will allow us to estimate the distances from h to U and V.

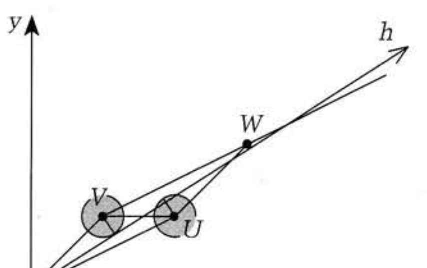


Figure 2

We'll need one additional construction. Let's complete the parallelogram $OUWV$ (fig. 2). Its fourth vertex W is also an integer point (like O, U, and V), since its coordinates are the sums of the corresponding coordinates of U and V.

The property of triangle OUV stated above implies the following three facts:

1. Parallelogram $OUWV$ has area 1.
2. Point W is outside the orchard.
3. The coordinates of W are relatively prime numbers.

Proof of fact 1:

Indeed, the area of $OUWV$ is twice the area of triangle OUV. But it's known that the area of a triangle with integer vertices and no other integer points inside it or on the border is equal to 1/2. (This is the basic lemma for deriving Pick's theorem for the area of an arbitrary polygon with integer vertices—see, for instance, "Chopping Up Pick's Theorem" in the January/February 1994 issue of *Quantum*, or "Suggestive Tilings" in the July/August 1994 issue.[2])

Proof of fact 2:

Further, suppose for definiteness that ray h intersects the side UW of the parallelogram (and not VW). Then W is closer to h than V, because W lies between V and the intersection point of rays h and VW. (You may show in addition that the distance from W to h equals the difference of the distances from V and U to h.) By the choice of V, this means that W doesn't belong to the orchard. This conclusion remains true if ray h passes through W, since it doesn't have integer points within the orchard.

Proof of fact 3:

Let d be the greatest common divisor of the coordinates a and b of W. Then the point $(a/d, b/d)$ is an integer point on OW. If we had $d \geq 2$, this point would lie in the triangle OUV, which is impossible, as we know. So $d = 1$, and we're done.

Now we're ready to finish the solution of the Orchard Problem.

Denote by w the length of the side OW of triangle OUW. Since the area of this triangle is 1/2 (fact 1), its height dropped from U—that is, the distance from U to OW—equals $(2 \cdot 1/2)/w = 1/w$. The distance from V to OW is, of course, the same. But by facts 2 and 3 and the definition of l, $w \geq l$. Therefore, for any $r > 1/l$ both trees at U and V block the ray OW and, therefore, any ray in the angle UOV—in particular, our ray h. So *for any $r > 1/l$, the view from the origin is completely blocked*.

Not only that, the same argument shows that our ray h will be blocked even for $r = 1/l$ if it doesn't coincide with OW, or if it coincides with OW, but $w > l$. So it remains to examine the case when h is drawn through a point L with positive relatively prime integer coordinates (a, b) at a distance of $OL = l$ from the origin.

Consider an arbitrary integer point $M(x, y)$ in the first quadrant. It doesn't lie inside the segment OL, because in that case we'd have $y/x = b/a$ or $ay = bx$ and, since a and b are relatively prime, x and y would be divisible by a and b, respectively. Thus, $x \geq a$, $y \geq b$, and $OM \geq OL$. Now, if N is the integer point closest to OL in the triangle OLM, then the area of triangle OLN is 1/2 by the argument applied above to triangle OUV. So the area of OLM is no less than 1/2 and the distance from M to OL is no less than $1/l$.

It follows that *no tree of radius $r \leq 1/l$ can block the ray OL*, which completes the proof of our general statement.

[2]However, Pick's theorem can be proved without this lemma; then the lemma becomes a simple particular case of the theorem.

Challenges in physics and math

Math

M161

Estimating the GCD. Positive integers a and b are such that the number $(a+1)/b + (b+1)/a$ is also an integer. Prove that the greatest common divisor of a and b does not exceed $\sqrt{a+b}$. (A. Golovanov, E. Malinnikova)

M162

Perfect square emerges. The sum of three integers a, b, c is zero. Prove that the number $2a^4 + 2b^4 + 2c^4$ is the square of an integer. (S. Fomin, L. Kurlyandchik)

M163

A one-checker game. A checker is placed on a corner square of an $n \times n$ chessboard. Then two players take turns moving it one square horizontally or vertically so as not to hit any previously visited square. The player who can't make a move according to this rule loses. (a) Prove that for even n the player who makes the first move can force a win, and that for odd n the second player can always win. (b) Who wins if the checker is placed initially on a square adjacent to a corner square (along a side)? (N. Netsvetayev)

M164

Double inscription. An n-gon of area A is inscribed in a circle of radius R. On each of its sides a point is marked. Prove that the perimeter P of the n-gon whose vertices are the marked points is no smaller than $2A/R$. (V. Dubrovsky)

M165

Isosceles and monochromatic. R vertices of a regular $(6n+1)$-gon are colored red, and all the remaining vertices are colored blue. Prove that the number of isosceles triangles with vertices of the same color depends only on the number R, but not on the arrangement of colors. (D. Tamarkin)

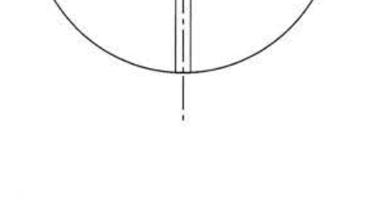
Physics

P161

A body on the table. When the mass of a body placed in the center of a square table exceeds a certain value m, the table's legs break. Find the set of points on the table's surface where a body of mass $m/2$ can be placed without breaking the table. (O. Batishchev)

P162

Lunar metro. A straight channel connects points A and B located on the surface of the Moon at an angular distance of 90° (see the figure below). The channel is filled with air at 0°C. The pressure at the middle of the channel is $P_c = 10^5$ Pa. Find the air pressure in the channel at the lunar surface. Consider the Moon to be a homogeneous sphere of diameter D = 3,480 km. The acceleration due to gravity for the Moon g_M at the surface is one sixth the corresponding value for the Earth. (A. Stasenko)



P163

Ice in a tube. A vertical tube is partially filled with water at a temperature t_0 = 0°C to a height H = 20 m. By how much does the height of the tube's contents change if the temperature decreases to t_1 = –0.01°C? The latent heat of fusion for ice is L = 335 J/g, the density of ice is ρ_i = 0.92 g/cm^3. It is known that a change in the external pressure ΔP causes a change in the melting temperature of ice of ΔT and that $\Delta T/T = (1/\rho_w - 1/\rho_i)\Delta P/L$, where T is the temperature of the ice–water mixture and ρ_w is the density of water. (W. Ovchinkin)

P164

Two large plates. Two large square nonconducting plates (each of area S) are placed parallel to each other and separated by a small distance d. Electric charges Q and $-Q$ are distributed uniformly over their surfaces. Find the potential difference between the center and a corner of one of the plates. (O. Savchenko)

P165

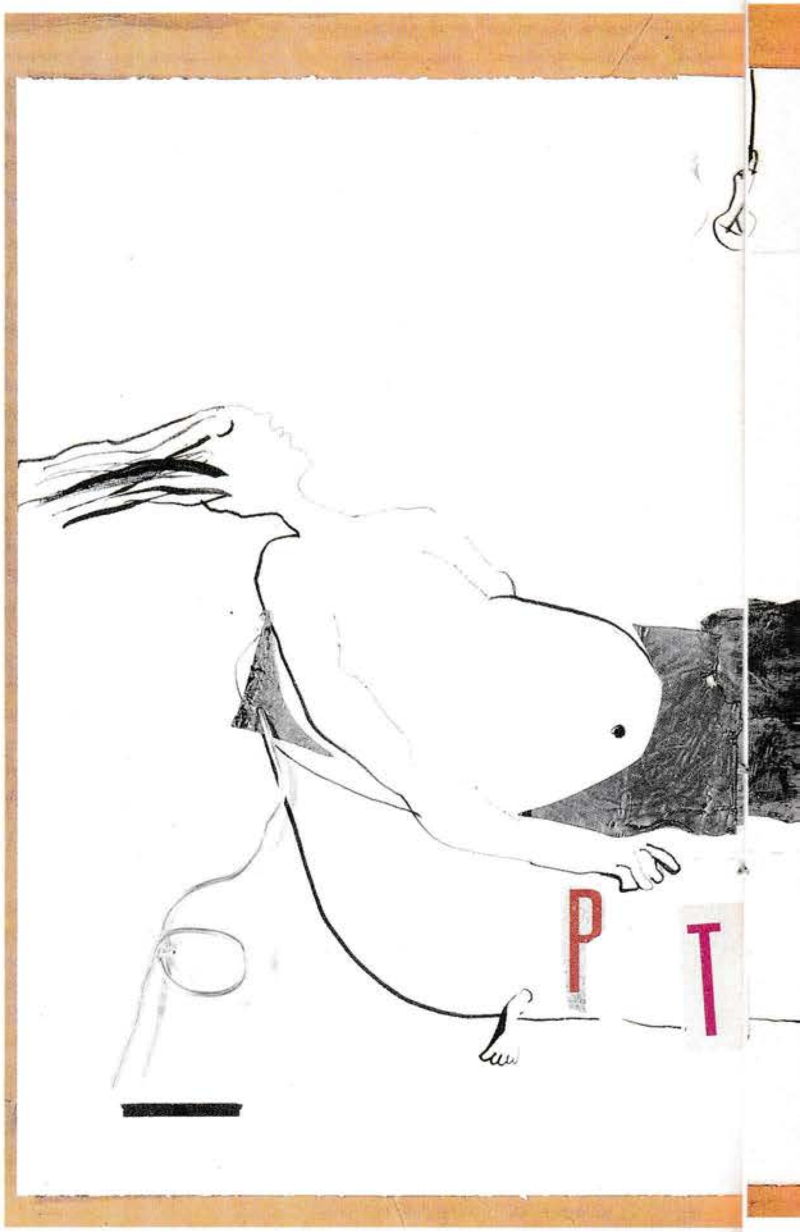
Cup of mercury. A cylindrical cup filled with mercury rotates about the vertical axis with an angular velocity ω. In so doing, the mercury's surface forms a parabolic mirror. Find the focal length of this mirror. The density of mercury is ρ and the acceleration due to gravity is g.

ANSWERS, HINTS & SOLUTIONS ON PAGE 44

The wind in the quicksilver

Explaining the backward flow of ions in mercury amalgams

by Ivan Vorobyov



SINCE THE END OF THE LAST CENTURY, EXPERIMENTS have been carried out to study the conductivity of liquid metals. Under normal conditions, you hardly ever come across metals in the liquid phase, but many of them can be mixed with mercury to form liquid solutions, or "amalgams." The interesting thing about the conductivity of liquid metals is that, in addition to electrons, ions must contribute to the electric current. And where you have ion motion, you have a transfer of mass.

Let's try to imagine qualitatively what will occur in a cuvette with pure mercury when it is connected to a voltage source. Electrically, mercury is a mixture of electrons and positive ions. At first glance everything is clear: under the action of the electric field, the electrons will move from the negative to the positive electrode, while the positive mercury ions will move in the opposite direction. But while the electrons can travel along the entire electrical circuit, things are a little different for the ions. Arriving at the negative electrode, it seems they must collect there, raising the level of the mercury at this point.

But nothing of the sort occurs! Experiments showed that the level of pure mercury is practically the same in the cuvette, and that if there is any small increase in height, it's not near the positive electrode but at the negative. In some amalgams the ions move as they should—that is, in the direction of the field; but the ions of other metals force their way in the opposite direction—against the field! It's rather surprising—the ions move in a direction opposite to the electrical force.

What causes this "strange" ion behavior? Well, we've omitted something up till now: the constant bombardment of the ions by the electrons. Let's try to determine what effect the mutual collisions of electrons and ions have on the movement of the electrons and ions in the presence of an external electric field.

To help us visualize what's going on, we can take an ion to be a heavy elastic ball that is struck by a stream of lighter particles—that is, electrons (the mass of an ion is several thousands of times greater than that of an electron). Let the stream of particles have an average velocity **u** before the collision and a corresponding nonzero momentum. After the collision the stream scatters uniformly in all directions (fig. 1).[1] (A beam of light is scattered by a reflecting spherical mirror in much the same way.) This means that immediately after the collision, all directions of electron motion are equally probable, and so the total momentum of the electrons is zero.

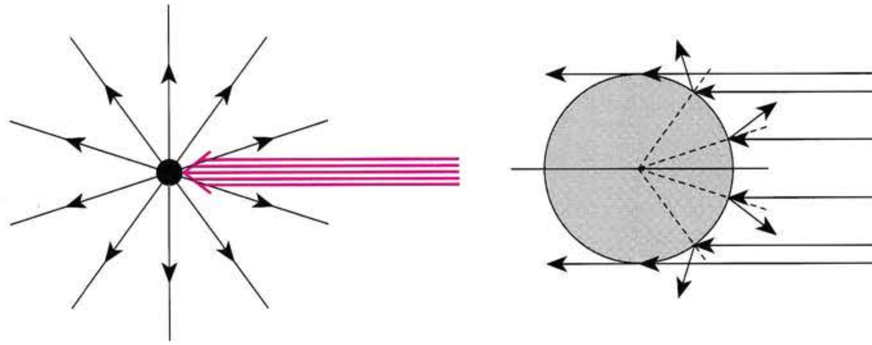


Figure 1

[1]The strict solution of this problem will be found in a course on theoretical physics.

If the electron stream's momentum changes due to the collisions with the ions, this means that the ions exert a force on the electrons. Let **u** be the average velocity of the electrons in the stream, and let $m\mathbf{u}$ be the average momentum of an electron. Let the time between collisions of an electron with the ions be τ, which we can consider to be the time it takes for the average momentum $m\mathbf{u}$ of each electron to be lost. The average force acting on one electron is thus

$$\mathbf{f} = -\frac{m\mathbf{u}}{\tau}.$$

The loss of momentum by the electrons *taken as a whole* due to collisions with the ions is described by the average braking force **f**. This force need not act on any *particular* electron, but it is this average force that is responsible for the movement of the stream as a whole.

It's true that, in addition to the directed motion with mean velocity **u**, the electrons also participate in chaotic thermal motion. But this doesn't change the value of the braking force. In the midst of this thermal motion, the total momentum of a large number of electrons is always equal to zero, so the change in the total momentum is related only to the loss of momentum in the directed motion.

When there is an external electric field **E**, the "scattered" electrons are "picked up" by this field, which organizes their motion by imparting some directed momentum to them. This momentum is lost in the next collision, then replenished again by the external field, and so on. Thus, one might say that some constant average velocity is established for the stream of electrons. In other words, the sum of the forces acting on the electron stream is equal to zero. For a single electron, this can be expressed as

$$e\mathbf{E} + \mathbf{f} = 0.$$

Indeed, if the force eE of the field exceeds the braking force f, the average velocity of the electrons increases, and so does the braking force f (which is proportional to the velocity), until the total force is again zero. If, on the other hand, the velocity by chance becomes too great and the braking force f exceeds the tractive force eE, the stream will be slowed until the force f again becomes equal (in magnitude) to eE. Thus, for motion

Art by Dmitry Krymov

with an established average velocity (when the total force is zero), the braking force of an ion calculated per electron is

$$\mathbf{f} = -e\mathbf{E} \qquad (1)$$

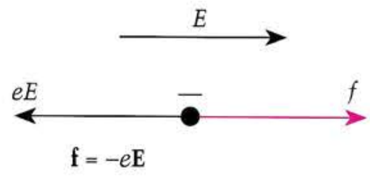


Figure 2

(see figure 2).

Why don't we take collisions between electrons into account? Because they don't affect the average velocity of the electrons. Indeed, consider the collision of two electrons. Let their velocities before an impact be $\mathbf{v}_1$ and $\mathbf{v}_2$, and after it $\mathbf{v}_1'$ and $\mathbf{v}_2'$. The law of conservation of momentum yields

$$m\mathbf{v}_1 + m\mathbf{v}_2 = m\mathbf{v}_1' + m\mathbf{v}_2',$$

from which we get

$$\frac{\mathbf{v}_1 + \mathbf{v}_2}{2} = \frac{\mathbf{v}_1' + \mathbf{v}_2'}{2}$$

—that is, the average velocity before the collision is equal to that after it (fig. 3).

And what forces act on the ions? Let's consider a pure metal whose ions are all alike, and each atom has contributed Z conduction electrons. The electric field acts on each ion with a force $Z|e|\mathbf{E}$, where $Z|e|$ is the ion's charge (the absolute value of the electron charge is used because the ion has a positive charge). According to Newton's third law, on average each electron applies a force

$$\mathbf{f}_1 = -\mathbf{f}.$$

If the number of ions is N, then each ion is affected by one electron with a force $\mathbf{f}_1/N$. Thus the force applied to an ion from all the electrons (which total ZN) is equal to $Z\mathbf{f}_1$. However, the sum of the forces

$$\mathbf{F} = Z|e|\mathbf{E} + Z\mathbf{f}_1$$

is zero (fig. 4)! To convince yourself of this, insert the value of the force $\mathbf{f}_1$ from equation (1), which differs from $\mathbf{f}$ only in its sign:

$$\mathbf{F} = Z|e|\mathbf{E} + Ze\mathbf{E} = 0.$$

Thus, the forces from an electric field acting on the ions in a pure metal are compensated by the forces of the "electron wind" (due to the moving electrons). The fact that the total force is zero means either a state of rest or of uniform motion. However, the flow of liquid in a vessel gives rise to viscous friction (caused in the final analysis by interaction with the vessel's wall), and it is this friction that ensure the rest state of the liquid relative to the vessel when all other forces acting on the liquid are balanced.

But what if there are some "alien" ions (impurities) in the metal? Let the charge of the alien ions be $|e|Z'$. Since the size of the alien ion is different, the number of impacts with the electrons changes. If the cross section of an alien ion is σ' and that of a "native" ion is σ, the alien ion will be subjected to σ/σ' times more collisions than a native one. The force from the electrons acting on an alien ion will change in the same proportion. So the sum of the forces of the electric field and the electric wind is equal to

$$\mathbf{F}' = Z'|e|\mathbf{E} + Z\frac{\sigma'}{\sigma}\mathbf{f}_1$$

(see figure 5). Substituting $-|e|\mathbf{E}$ for $\mathbf{f}_1$ yields

$$\mathbf{F}' = |e|\left(Z' - Z\frac{\sigma'}{\sigma}\right)\mathbf{E}.$$

If $Z' > Z(\sigma'/\sigma)$, the alien ions will move in the direction of this force, and their velocity will be directed along the field. If $Z' < Z(\sigma'/\sigma)$, the ions will move counter to the field.

So how do we explain the fact that in pure mercury the ions move in the direction of the electron wind? We have just shown that the ions must be immobile in a pure metal! Here's the catch: we've assumed that all the ions are identical, but this isn't exactly true. Although most ions are indeed in the same "normal" energy state, some ions always have a higher energy. These are the so-called "activated" ions. The probability of an electron–ion collision increases with the energy of the ions. Thus, the activated ions are "aliens"—they have the same charge as a normal ion, but they have a larger cross section. These ions will be picked up by the electron wind and carried to the positive electrode of the voltage source.

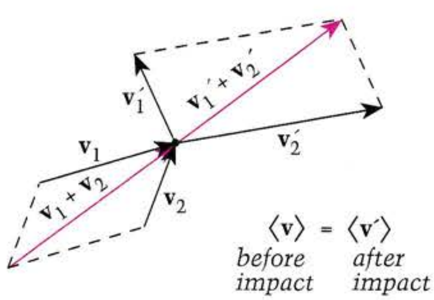


Figure 3

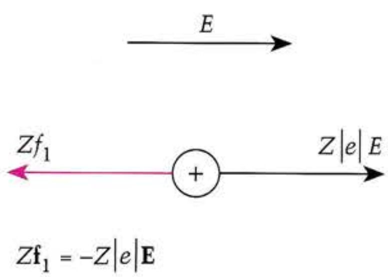


Figure 4

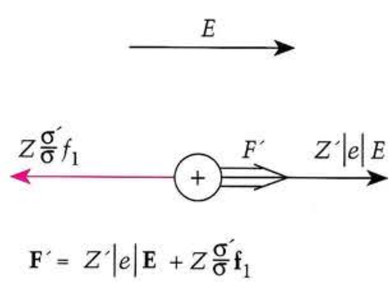


Figure 5

This phenomenon underlies the separation of mercury isotopes in an electric field. The ions of different isotopes have practically the same charge and cross section, but due to their smaller mass the activated ions of the light isotopes are carried by the electron wind more quickly than the activated ions of the heavy isotopes. This results in an increased concentration of light isotope near the positive electrode—that is, if mercury is drawn off near the positive electrode, it will be enriched by the light isotopes.

This "paradoxical" transfer of ions in mercury amalgams was first noted in 1907. At that time the concept of mutual friction between ions and electrons was introduced into physics. However, subsequent theories, which remained in circulation until 1959, were actually a step backward—they considered the motion of the electrons and ions as independent of one another. This can partially be explained by the fact that the experimental results were unconvincing. In 1953 the separation of mercury isotopes in an electric field was discovered. This was followed by research on the electrical transfer of ions. The first theoretical papers explaining the physical mechanism of the electron wind appeared at the beginning of 1959.

Many important and interesting phenomena are related to the electron wind in both liquid and solid metals and also in semiconductors.

MATHEMATICAL SURPRISES

The magic of 3 × 3

The \$100 question: can you make a magic square of squares?

by Martin Gardner

PERHAPS THE OLDEST OF ALL combinatorial problems about numbers is the task of placing the first nine counting numbers in a 3 × 3 matrix so that each row, column, and main diagonal has the same sum. It turns out that, not counting rotations and reflections as different, there is only one solution:

2	9	4
7	5	3
6	1	8

Legend has it that in the 23rd century B.C. the mythical King Yu saw the pattern as spots on the back of a sacred turtle in the River Lo. Many modern scholars doubt that *lo shu*, the most common name in China for the magic square, is that ancient. They believe that the pattern is not older than the tenth century A.D. At any rate, the name means Lo River writing. The Chinese identify it with their familiar yin-yang circle of light and dark regions. The four even digits (shown shaded above) are identified with the dark yin. The Greek cross made of the five odd digits is identified with the light yang. For centuries the *lo shu* has been used as a charm on jewelry and other objects. Today large passenger ships frequently arrange their shuffleboard numbers in the *lo shu* pattern.

How can we prove that the pattern is unique? The simplest way known to me is first to note that the nine digits add to 45. If each of the three rows (or columns) has the same sum, then the sum must be one third of 45, or 15. We next list the eight possible triads of digits that add to 15:

9 + 5 + 1
9 + 4 + 2
8 + 6 + 1
8 + 5 + 2
8 + 4 + 3
7 + 6 + 2
7 + 5 + 3
6 + 5 + 4

The square's center digit belongs to four lines of three. Five is the only digit to appear in four of the triads; therefore, 5 must go in the center. Digit 9 is in only two triads, so it must go in a side cell, and the digit at the other end of the same line must of course be 1. Digits 3 and 7 are also in just two triads. For symmetry reasons it does not matter how they are placed in the other two side cells. This gives us

	9	
7	5	3
	1	

The remaining empty cells can now be filled with even digits in only one way to complete the magic.

Various mnemonic methods allow one to form the *lo shu* without having to memorize the entire pattern. For me the easiest is to place the even digits in the corners, in sequence from left to right, top to bottom. The odd digits, also in sequence starting with 9 and going backward, form the zigzag pattern show below:

2	9	4
7	5	3
6	1	8

Note that 1, 2, 3 and 7, 8, 9 mark the corners of two isosceles triangles, with 4, 5, 6 along a main diagonal.

In *The Mathematical Gazette* (December 1970, p. 376) R. Holmes pointed out a surprising property of the *lo shu*. Take each row, column, and diagonal (including the four "broken" diagonals) as a three-digit number to be read both forward and backward. The following identities hold:

orthogonals

$$(294 + 753 + 618)^2 = (492 + 357 + 816)^2$$
$$(276 + 951 + 438)^2 = (672 + 159 + 834)^2$$

diagonals

$$(654 + 132 + 879)^2 = (456 + 231 + 978)^2$$
$$(852 + 174 + 639)^2 = (258 + 471 + 936)^2$$

The identities are unaffected if the middle digits of the numbers are deleted, or any two corresponding two digits.

Imagine the *lo shu* to be toroidally connected—that is, imagine the pattern wrapped around both vertically and diagonally, as it would be if drawn on a torus divided into nine cells. I was amazed to discover years ago that if you add the four digits in each 2×2 square, the sums are the nine consecutive numbers from 16 through 24.

Of course, an infinity of 3×3 magic squares can be constructed with other numbers, not necessarily in counting sequence, and including all real numbers. To exclude trivial examples, we assume that no two numbers are alike. For the matrix to be magic, however, the nine numbers must fall into three triads, in each of which the numbers are in the same arithmetic progression (that is, with the same differences). Moreover, the smallest numbers of the three triads must also be in arithmetic progression, though not necessarily with the same differences as the triads. We can express these rules algebraically as follows:

$a + x$	$a + 2y + 2x$	$a + y$
$a + 2y$	$a + y + x$	$a + 2x$
$a + y + 2x$	a	$a + 2y + x$

Note that each line of three cells has a sum of $3a + 3y + 3x$, or $3(a + y + x)$, proving that the magic constant must be a multiple of 3, and that the center number is one third of the constant.

With the square's algebraic structure in mind, it is possible to construct a fascinating variety of magic squares based on given restraints. The box on the next page shows a sampling of such squares. Each is the simplest square meeting the restraints, defining "simplest" as the square with the lowest magic constant. Do you see why the number 2 cannot be used in a magic square of primes?

In 1987 I offered \$100 to anyone who found a 3×3 magic square made with consecutive primes. The prize was won by Harry Nelson of Lawrence Livermore Laboratories. He used a Cray computer to produce the following simplest such square:

1,480,028,201	1,480,028,129	1,480,028,183
1,480,028,153	1,480,028,171	1,480,028,189
1,480,028,159	1,480,028,213	1,480,028,141

Martin LaBar, in *The College Mathematics Journal* (January 1984, p. 69), asked if a 3×3 magic square exists with nine distinct square numbers. (Such squares exist with eight distinct squares plus a zero.) Neither such a square nor a proof of impossibility has been found. Nelson believes it exists, but beyond the reach of any of today's supercomputers running a reasonable amount of time.

I here offer \$100 to the first person to construct such a square. If it exists, its numbers are sure to be monstrously large. John Robertson has shown that the task is equivalent to finding an elliptic curve of the form $y^2 = x^3 - n^2x$ with three rational points, each the double of another

Advertisement from *Time* magazine (1959). From the author's files.

$c = 27$	$c = 30$	$c = 24$
15 1 11 / 5 9 13 / 7 17 3	16 2 12 / 6 10 14 / 8 18 4	14 0 10 / 4 8 12 / 6 16 2
odd integers	*even integers without zero*	*even integers with zero*

c = 27 (*odd integers*)

15	1	11
5	9	13
7	17	3

c = 30 (*even integers without zero*)

16	2	12
6	10	14
8	18	4

c = 24 (*even integers with zero*)

14	0	10
4	8	12
6	16	2

c = 111 (*primes including 1*)

67	1	43
13	37	61
31	73	7

c = 177 (*primes excluding 1*)

71	89	17
5	59	113
101	29	47

c = 3177 (*primes in arithmetic progression*)

1669	199	1249
619	1039	1459
829	1879	409

c = 354 (*consecutive composites*)

121	114	119
116	118	120
117	122	115

c = 54 (*composites in arithmetic progression*)

27	6	21
12	18	24
15	30	9

c = 636 (*all numbers and c are palindromes*)

222	101	313
303	212	121
111	323	202

Three-by-three magic squares of special types with lowest constant (c). The palindrome square was constructed by Rudolph Ondrejka of Linwood, New Jersey.

rational point on the curve, having x-coordinates in arithmetic progression.

Henry Ernest Dudeney, in *Amusements in Mathematics* (pp. 124–25) and in his article on magic squares in the fourteenth edition of the *Encyclopaedia Britannica*, defines magic squares based on subtraction, multiplication, and division, and gives 3 × 3 examples of each kind. There are also 3 × 3 *antimagic* squares with the property that no two sums of three lines are alike. If certain provisos are met, a variety of interesting antimagic combinatorial problems result. These variants, however, will have to be the topic of another article.

Surely the most fantastic 3 × 3 magic square ever discovered is one constructed by Lee Sallows, a British electronics engineer who works for the University of Nijmegen in Holland:

5	22	18
28	15	2
12	8	25

It would be hard to guess its amazing property. For each cell, count the number of letters in the English word for its number, then place these counting numbers in the corresponding cell of another 3 × 3 matrix. For example, "five" has four letters, so 4 goes into the top left corner of the new matrix. Here is the result:

4	9	8
11	7	3
6	5	10

Not only is it another magic square, but its integers are in consecutive order! Sallows calls the first square the *li shu* (*li* for his first name Lee), and the second square its alphamagic partner. His computer investigations of alphamagic squares in more than twenty languages are reported in his two-part article "Alphamagic Squares" in *Abacus* (Vol. 4, 1986, pp. 28–45, and 1987, pp. 20–29, 43).

Now for three easy puzzles to work on before the answers are given in the next issue.

1. Construct a 3 × 3 magic square with nine consecutive positive integers that has a magic constant of 666—the notorious number of the Beast in the Bible's Apocalypse.

2. Place 1 in a corner cell of the 3 × 3 matrix, then fill the remaining cells with consecutive nonnegative digits to form a magic square.

3. Arrange nine playing cards like this:

1	2	3
3	1	2
2	3	1

All rows and columns have a sum of 6, as well as one main diagonal, but the other diagonal has a sum of 3. Change the positions of three cards to make the square completely magic.

Martin Gardner *does his magic in Hendersonville, North Carolina, and writes nonstop. Springer-Verlag will soon publish* The Universe in a Handkerchief, *a book on the mathematical recreations of Lewis Carroll.*

FEEDBACK

Bell curve? What bell curve?

Students are capable of more than we give them credit for

by Paul Horwitz

> *Editor's note*: In the May/June 1995 issue, *Quantum*'s publisher, Bill Aldridge, expressed his views on how the National Science Education Standards were being interpreted. He rejected the notion that "real science and math—something of significant depth—is beyond the reach of most American young people." Aldridge stressed the importance of proper sequencing of the material: experience → words → symbols → numbers → concepts → relationships. If properly presented over an extended period of time (kindergarten through twelfth grade), science is accessible to all students, not just the gifted.
>
> Some of the feedback from Aldridge's editorial was published in the July/August issue. He recently received this commentary from Paul Horwitz, and we deemed it worthy of our readers' attention.
>
> For the sake of our younger readers, we should explain the title. The "bell curve" is a reference to the controversial book by Herrnstein and Murray (1994) and is a nickname for the so-called "normal distribution" in statistics. The bell curve was once a popular tool used in evaluating student performance. Teachers grading "to the curve" would *necessarily* give the great majority of students a grade of "average" and a few students grades of either "excellent" or "poor." The bell curve in effect doomed a portion of the population to "failure."

LUC HAS BEEN LEARNING Einstein's Theory of Relativity. Using a computer, he and his classmates at Boston High School have built "thought experiments" that simulate objects moving at close to the speed of light. They observe these in one frame of reference and try to imagine what they would look like in another—out the window of a moving train, for instance.

Some people claim that human intelligence is largely determined by DNA. They say that IQ scores are indicators of an immutable genetic inheritance. Education, they imply, is irrelevant. Luc and his friends are quietly proving them wrong.

I have observed them for several months now as part of a National Science Foundation project that is exploring the use of computers for teaching math and science. Most of the time they work in small groups, oblivious to the teacher and to me. They build pictures in their heads, try to communicate them through gestures, then use the computer to transform them into live, interactive demonstrations. Every so often they seem to breach some mental barrier reef and come spluttering out on the other side, clinging proudly to a new idea. Their smiles at these times clutch at the heart. They do not smile in school very often, these inner-city kids.

These introductory physics students—all but one Haitian or African American—have advanced steadily from simple, everyday problems to the notoriously difficult Twin Paradox, according to which someone who embarks on a round trip at high speed returns younger than her twin sister. They have followed the same path, encountered the same frustrations, and overcome the same obstacles as any students I have ever observed—and they have progressed at about the same rate. Nevertheless, nearly every member of the class shows signs of severe educational deprivation, particularly in mathematics. Their knowledge of the decimal system is spotty and unreliable; they are easily confused by numbers bigger than one million; and they have the greatest difficulty creating or understanding simple graphs. It is hardly surprising that they perform so poorly on the tests society uses to evaluate them.

Years of neglect have left these students perilously at risk, but modern technology, combined with a new approach to learning, is having a remarkable effect on them.

Rather than teaching them facts, the computer is literally changing the way they think. It provides them with a manipulable, visual medium within which they can construct scenarios, puzzle over them, and alter them to answer

CONTINUED ON PAGE 52

Fluids and gases

"Everyone will readily agree that the the
if only it were not created against Minerva's will, is

QUESTIONS OF THE MOTION of liquids and gases or the movement of various solid bodies in fluids (primarily in water and air) are considered within the framework of special disciplines—hydrodymanics and aerodynamics. It's reasonable to study both media in the same course in school (and unite them by the single phrase "the mechanics of fluids") because they have many features in common.

Bernoulli proclaimed nature's truth: that this science is far from simple or trivial. As for its usefulness, who will argue against it? From ancient times people have dealt with problems of water flowing in channels and pipes; they have constructed mills driven by water and wind; and in modern times people have become interested in problems related to travel on and under water, through and beyond the atmosphere.

As time went on, practical needs made ever greater demands on theory to find the laws that "regulated" the motion of fluids. Among those enlisted to solve the most intricate problems of fluid mechanics were such glorious names as Newton, Euler, Thomson, Maxwell, Prandtl, Zhukovsky, and Chaplygin. The law of fluid motion formulated more than 250 years ago by Daniel Bernoulli in his famous *Hydrodynamics* (and named after him) is still used in modern practice.

Try to apply the laws of fluid motion yourself in the following problems.

Questions and problems

1. Can sails and a rudder be used to control the flight of a hot air balloon?

2. A helicopter initially at rest on the ground lifts off and hovers at a low altitude. When is the force on the ground the largest?

3. How does a kite fly? Why does it need a tail?

4. What will happen if you blow air between two burning candles standing side by side?

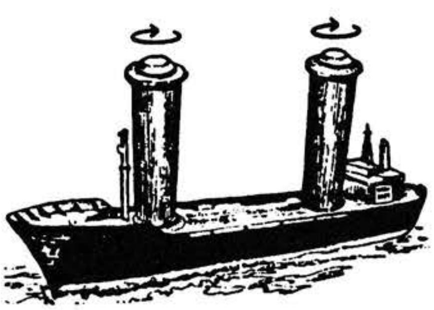
5. Why are there vents in the roofs of houses?

6. Why is it dangerous to stand near the edge of a platform when an express train goes by at high speed?

7. If a piece of corrugated tubing (like the hose of a vacuum cleaner) is waved in a circle overhead, it produces a sound. Why?

8. What is the purpose of the massive flywheels attached to the shafts of high-speed windmills?

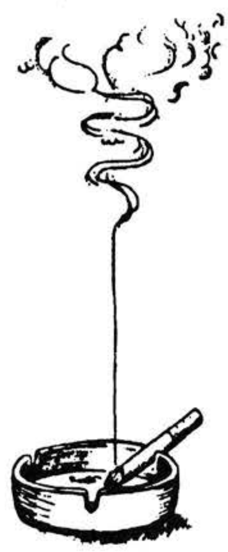
9. How can the rotating vertical cylinders move the ship shown in the figure below?



10. Why is there is a hole in the center of a parachute?

11. If you blow on a candle's flame through the stem of a funnel, the flame tilts toward the funnel. Why?

12. Why does the smoke from a cigarette sitting in an ashtray rise first as a uniform stream and then begin to curl?



13. Why does a stream of water flowing uniformly from a faucet become thinner as it travels downward?

es on the move

theory of forces and motions in fluids,
l, is neither useless nor trivial."—Daniel Bernoulli

14. If you block an open water faucet with your finger so as to leave a small opening, the water will spurt from it with a higher velocity than when it flowed from the fully open faucet. Why?

Microexperiment

Glue a piece of thread to a Ping-Pong ball. Suspend the ball near an open faucet and let it touch the stream of water. Pull the ball away from the stream. Why does the ball seem to stick to the water?

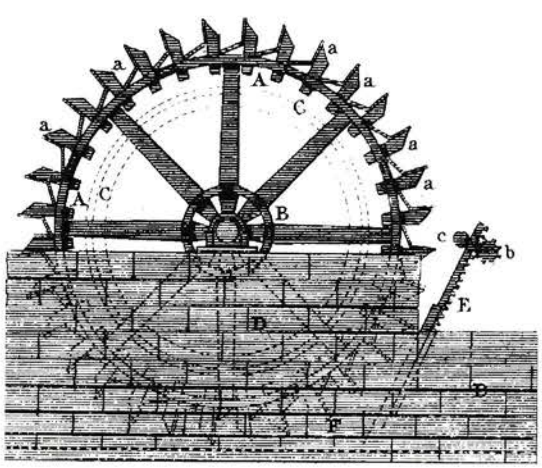
It's interesting that . . .

. . . the force of water striking an obstacle increases sharply with the stream's speed. This is responsible for the creation of ravines, riverbeds, and valleys and for the erosion of riverbanks and seacoasts. The huge scale of the work performed by these forces can be seen in the amount of alluvia carried off by the water, which is hundreds of millions of tons per year.

. . . living beings obey the laws of hydrodynamics. For example, birds flying in V-formation save their strength during their long-range migrations. And it has been estimated that fish swimming together as a school increase their endurance by more than 100%.

. . . the advantage of using the wind for transportation was noted by an ancient Chinese proverb: "A thousand oars and ten thousand puntpoles are not worth a single sail."

. . . the outstanding Dutch scientist and engineer Simon Stevin, known as one of the fathers of hydrostatics, managed to build a sail-driven car. This wind vehicle, nicknamed the "Hague Wonder," developed an appreciable speed and carried twenty passengers. Just like a sailboat, it also could turn and run against the wind.



. . . human beings have long used aspects of the movement of rotating bodies in air—for instance, when throwing a boomerang. It is only recently, however, that athletes paid any real attention to them. As a result, distances in discus throwing have increased, and soccer players routinely give the ball a spin to send it along a curved trajectory.

. . . vortices form when air flows past a body. Detaching from one side and then the other, in alternation, they rock the body. The amplitude of these oscillations can become so large as to destroy the object. This has happened to suspension bridges, antenna masts, and oil derricks.

. . . the water in the jet of a monitor (a hydraulic excavator used in mining) is ejected with such tremendous speed that it's more like an artillery projectile, bursting the soil and throwing large clods up into the air.

. . . by creating compressed and rarefied areas in a tube, one can transfer loads along it. This phenomenon resulted in the invention of pneumatic mail at the beginning of the 20th century. The same principle was used in the pneumatic railway that carried passengers in New York in the 1870s.

. . . turbines are an example of the most effective application of the jet pressure of a liquid or gas. Nowadays practically all power stations in the world use water or steam turbines, and the turbojet engine is the workhorse of the aviation industry.

. . . supersonic airplanes will be economical for long-range flights, according to calculations that have been made. The speed of such planes can reach 12,000 km/hour at a cruising altitude of 36–50 km. Due to different flying and landing conditions, these planes must have a special shape—in particular, their wings must have a variable geometry.

—Compiled by L. Leonovich

ANSWERS, HINTS & SOLUTIONS ON PAGE 51

PHYSICS CONTEST

Focusing fields

"We love to overlook the boundaries which we do not wish to pass."—Samuel Johnson

by Arthur Eisenkraft and Larry D. Kirkpatrick

HOW UTTERLY DELIGHTFUL is the Lorentz force. When we think of forces, we usually imagine a push or a pull. Such a push or pull is assumed to be in the direction of the line connecting the pusher and the object pushed. Not so with the Lorentz force. A magnetic field acting on a moving charge pulls in a direction perpendicular to the velocity of the particle and perpendicular to the magnetic field. If the particle moves parallel to the field, there is no force. If the particle comes to rest, there is no force. Finally, if the particle were to lose its charge, there would be no force. We can express the Lorentz force mathematically as $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$, where $\mathbf{F}$ is the force, q is the charge, $\mathbf{v}$ is the velocity of the charge, and $\mathbf{B}$ is the strength of the magnetic field. The multiplication sign tells us that this is a vector cross product, which simply means that the force vector is perpendicular to both the velocity vector and the magnetic field vector.

A most important research use of a magnetic field is to control a particle beam. In the cyclotron, the magnetic field causes the charged particle to move in a circle. The magnitude of the Lorentz force acting on the particle is $qvB = mv^2/R$. This can be simplified to an expression that relates the momentum of the particle to the charge, magnetic field, and radius of the particle's path:

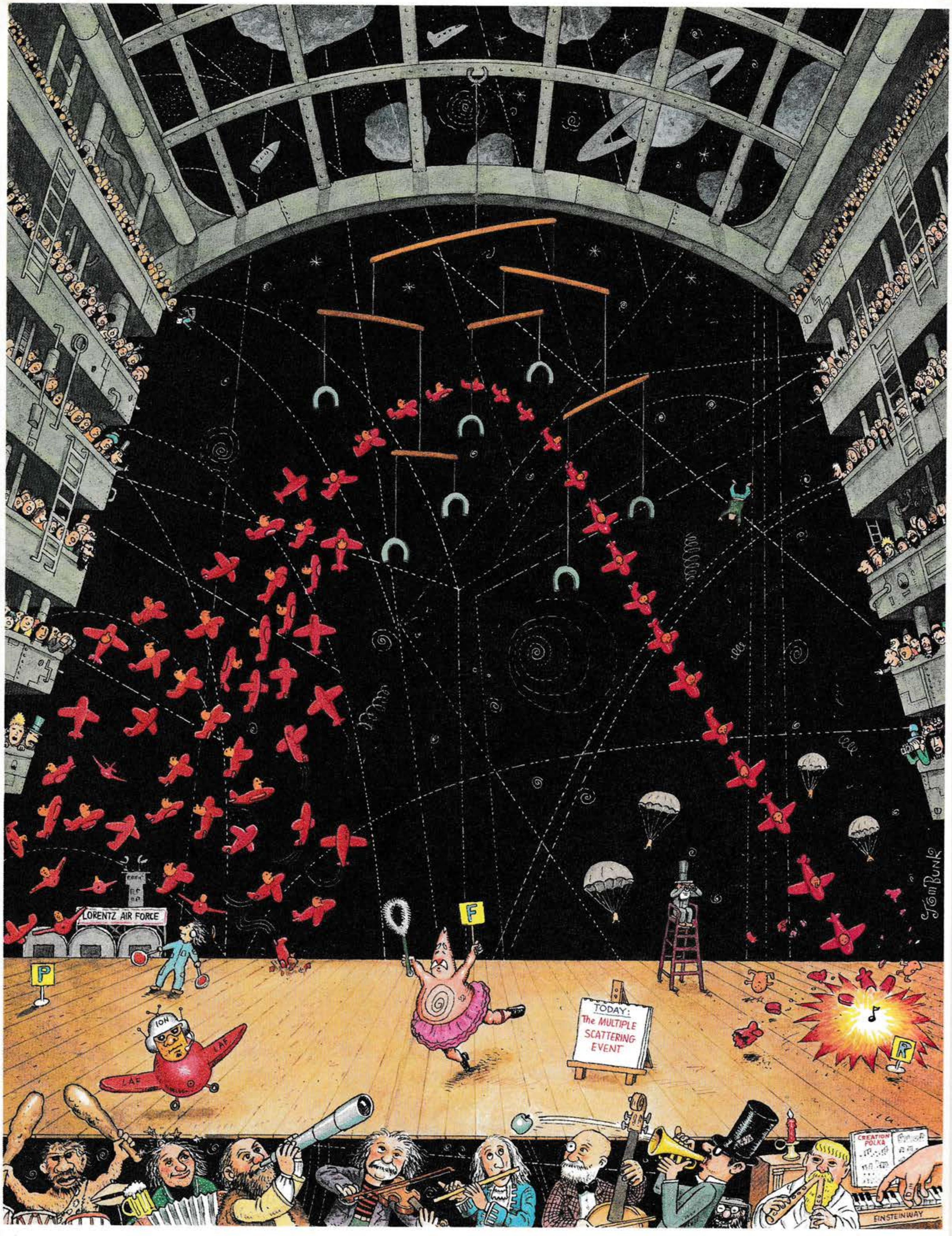
$$mv = qBr.$$

This nonrelativistic treatment provides the basic physics. Particles in the cyclotron or a bubble chamber travel at extremely high speeds, and relativistic equations must be used where we replace the momentum mv with the relativistic momentum γmv ($\gamma = 1/\sqrt{1-v^2/c^2}$).

Figure 1 is a display of a bubble chamber photograph. An electron is the only charged particle that will make a complete circle in a bubble chamber. Protons, pions, and mesons will trace out circular arcs, but can't trace out the complete circle due to their large momenta and the limited size of the chamber. By investigating a path of an electron, we can find its initial momentum and notice that it loses momentum (spirals in) during its interaction with the hydrogen in the detector. Since the electron is traveling counterclockwise, we can also determine

Figure 1

Art by Tomas Bunk


LORENTZ AIR FORCE
P
F
R
ION
LAF
TODAY:
The MULTIPLE SCATTERING EVENT
CREATION POLKA
EINSTEINWAY

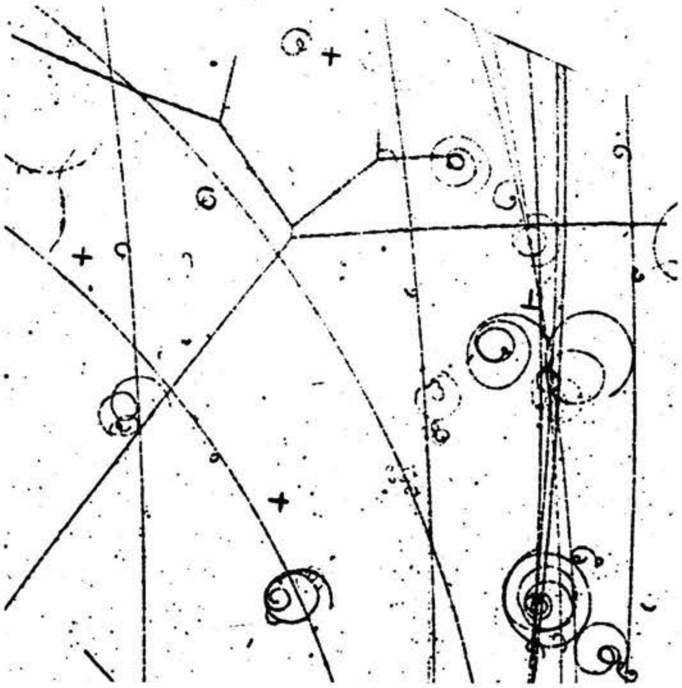


Figure 2

the direction of the external magnetic field.

Similar analyses are able to determine the momenta of other particles and to discover new particles as they interact with known particles. An example of a multiple scattering event that can be analyzed is shown in figure 2.

The problem for this month is from the VIII International Physics Olympiad that was hosted by East Germany in 1975. It requires us to find a magnetic field that can focus charged particles.

Ions of identical mass m, charge q, and speed v diverge from a point P (fig. 3). A uniform magnetic field B perpendicular to the plane of the page focuses them to a point R located at a distance $PR = 2a$ away from P. Their trajectories have to be symmetrical to the axis that is the perpendicular bisector of PR. Determine the boundaries of the magnetic field.

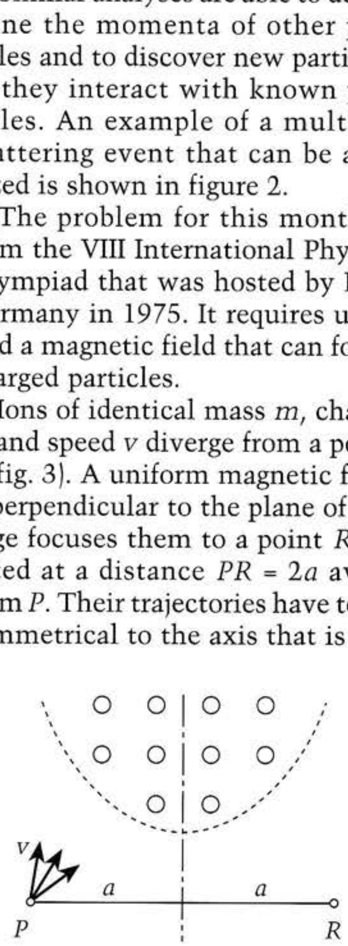


Figure 3

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

Pins and spin

The Contest problem in the July/August issue asked you to try bowling by laying a spinning cylinder on the alley rather than throwing a spherical ball. Correct solutions were submitted by Noah Bray-Ali from Venice High School in Los Angeles and Canh Nguyen from Little Canada, Michigan, although they had different interpretations of part E.

A. As with the bowling ball, the net force on the cylinder is the usual force of friction $f = \mu mg$. Therefore, Newton's second law tells us that

$$f = ma = \mu mg,$$

and the linear acceleration is

$$a = \mu g.$$

The frictional force also exerts a torque $\tau = rf$ on the cylinder about its center. According to Newton's second law for rotation,

$$\tau = I\alpha = -rf = -r\mu mg,$$

with $I = ½mr^2$. The resulting angular acceleration is

$$\alpha = \frac{-2\mu g}{r}.$$

Using the kinematic equations for translational and rotational motion, we have

$$v = v_0 + at = \mu gt$$

and

$$\omega = \omega_0 + \alpha t = \omega_0 - \frac{2\mu gt}{r}.$$

Using the condition for rolling without slipping, $v = \omega r$, we solve for the time when this first occurs:

$$\mu gt = r\omega_0 - 2\mu gt,$$

$$t = \frac{r\omega_0}{3\mu g}.$$

B. We can now plug this time into the equation for v,

$$v = \mu gt = \frac{1}{3}r\omega_0,$$

to find the speed when rolling occurs without slipping.

C. We can now obtain the corresponding distance in a variety of ways. We choose

$$d = \frac{1}{2}at^2 = \frac{r^2\omega_0^2}{18\mu g}.$$

D. The initial kinetic energy is only rotational and is given by

$$KE_i = \frac{1}{2}I\omega_0^2 = \frac{1}{4}mr^2\omega_0^2,$$

but the final kinetic energy is a combination of translational and rotational energy:

$$KE_f = \frac{1}{2}mv^2 + \frac{1}{2}I\omega^2,$$

with $v = r\omega$. Therefore,

$$KE_f = \frac{3}{4}mr^2\omega^2 = \frac{1}{12}mr^2\omega_0^2,$$

and the change in kinetic energy is

$$\Delta KE = -\frac{1}{6}mr^2\omega_0^2.$$

This gives a fractional loss of 2/3.

E. The work-energy theorem says

that the loss in kinetic energy is equal to the work done by the frictional force. However, in calculating the work, we must use only the distance slipped, not the linear distance traveled. We begin by calculating the total angle through which the cylinder rolls up until the time it quits slipping:

$$\theta = \omega_{\text{ave}} t = \frac{2}{3}\omega_0 t = \frac{2r\omega_0^2}{9\mu g}.$$

If the cylinder had not slipped, this angle would correspond to a linear distance of $r\theta$. The distance slipped is just the difference between this distance and the actual linear distance traveled:

$$d_{\text{slip}} = \left(\frac{2}{9} - \frac{1}{18}\right)\frac{r^2\omega_0^2}{\mu g} = \frac{r^2\omega_0^2}{6\mu g}.$$

The work done by the frictional force is then

$$W = -fd_{\text{slip}} = -\frac{1}{6}mr^2\omega_0^2,$$

which is equal to the change in the kinetic energy calculated above.

Noah Bray-Ali agreed with our interpretation for part E, but decided to calculate the loss in rotational kinetic energy and the gain in translational kinetic energy to get the total change in the kinetic energy. Both Noah and Canh Nguyen showed that the fractional loss in the rotational kinetic energy by itself is 8/9 and that this is due to the rotational work performed by the torque.

Talk back to us

We actually like it! We want to know what you think about *Quantum*. So

drop us a line at
Quantum
1840 Wilson Boulevard
Arlington VA 22201-3000

or

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

The discriminant at work

A universal tool for elementary algebra

by Andrey Yegorov

MANY OF THE PROBLEMS that you'll find below will seem completely baffling to an untrained person.[1] However, they can be solved using some surprisingly simple ideas. I'll try to demonstrate here how much information can be extracted from the well-known conditions for the solvability of quadratic equations and inequalities in real numbers. (We do not consider complex solutions in this article.)

Equations and systems

The main idea behind the method used below is illustrated by the following problem.

Problem 1. Solve the equation

$$5x^2 - 2xy + 2y^2 - 2x - 2y + 1 = 0.$$

Solution. You may be lucky enough to think of rewriting this equation in the form

$$(x + y - 1)^2 + (2x - y)^2 = 0.$$

Then, equating both terms on the left side to zero, you'll get a simple linear system that immediately yields the answer. But it isn't easy to come up with this transformation.

[1]And this is exactly how they were intended to look by their creators. Most of the problems in this article were taken from entrance exams for various Moscow universities and colleges. So now our readers can get some idea of what Russian students are up against!—*Ed.*

So let's try another approach. Look at the given equation as a quadratic equation for x with coefficients depending on y:

$$5x^2 - 2(y + 1)x + 2y^2 - 2y + 1 = 0.$$

To be solvable, this equation must have a nonnegative discriminant D. Let's compute it—or rather, $D/4$, which is easier here:

$$\begin{aligned}\frac{D}{4} &= (y+1)^2 - 5\left(2y^2 - 2y + 1\right)\\ &= -9y^2 + 12y - 4\\ &= -(3y-2)^2.\end{aligned}$$

It follows immediately that $y = 2/3$ (otherwise, $D < 0$). Then

$$x = -\frac{-2(y+1)}{5} = \frac{1}{3}.$$

Thus, the answer is $(1/3, 2/3)$.

Now, a problem with two simultaneous equations.

Problem 2. Solve the system of equations

$$\begin{cases} x^2 - 2xy + 2y^2 + 2x - 8y + 10 = 0, \\ 2x^2 - 7xy + 3y^2 + 13x - 4y - 7 = 0. \end{cases}$$

Solution. It's hard to handle this problem with standard methods. So let's tackle it as we've already done before: let's solve the first of the two equations for x. Its discriminant equals

$$D = -4(y - 3)^3.$$

So a real solution exists only for $y = 3$. Substituting it into the first equation, we find that $x = 2$. Then we only have to check that $(2, 3)$ satisfies the second equation as well.

Next is a problem with more unknowns than there are equations.

Problem 3. Solve the system of equations

$$\begin{cases} x + y + z = \sqrt{3}, \\ x^2 + y^2 + z^2 = 1. \end{cases}$$

Solution. Express x in terms of y and z using the first equation, substitute into the second, and rework it to get

$$y^2 + \left(z - \sqrt{3}\right)y + z^2 - \sqrt{3}z + 1 = 0.$$

We get a quadratic equation in y with the coefficients depending on z. Calculate its discriminant:

$$D = -3z^2 + 2\sqrt{3}z - 1 = -\left(\sqrt{3}z - 1\right)^2.$$

It follows that $z = 1/\sqrt{3}$. Now we can do some more calculations to find that $x = y = 1/\sqrt{3}$, too. But in fact this is clear from the symmetry of the problem and the uniqueness of the value of z.

Problem 4. Solve the equation

$$\cos x + \cos y - \cos(x+y) = \frac{3}{2}.$$

Solution. Using well-known trigonometric formulas for $\cos\alpha + \cos\beta$ and $\cos 2\alpha$, we arrive at the following form of the equation:

$$4\cos\frac{x+y}{2}\cos\frac{x-y}{2} - 4\cos^2\frac{x+y}{2} - 1 = 0,$$

or

$$4t^2 - 4\left(\cos\frac{x-y}{2}\right)t + 1 = 0,$$

where $t = \cos[(x+y)/2]$. The condition for solvability yields

$$\frac{D}{4} = 4\left(\cos^2\frac{x-y}{2} - 1\right) \geq 0,$$

and therefore

$$\cos^2\frac{x-y}{2} = 1.$$

Now the problem is reduced to two systems of equations

$$\begin{cases}\cos\dfrac{x-y}{2} = 1,\\ \cos\dfrac{x+y}{2} = \dfrac{1}{2}\end{cases} \quad \text{or} \quad \begin{cases}\cos\dfrac{x-y}{2} = -1,\\ \cos\dfrac{x+y}{2} = -\dfrac{1}{2}\end{cases}$$

that quickly lead to the following two series of solutions (x, y):

$$(\pm\pi/3 + 2(k+n)\pi,\ \pm\pi/3 + 2(k-n)\pi)$$

and

$$(\pm 2\pi/3 + \pi + 2(k+n)\pi,\ \pm 2\pi/3 - \pi + 2(k-n)\pi),$$

where k and n are arbitrary integers.

Problem 5. Determine the values of the parameter a for which there exists a unique pair (x, y) of real numbers satisfying the equation

$$ax^2 + (3a+2)y^2 + 4axy - 2ax + (4-6a)y + 2 = 0.$$

Solution. Notice that for $a = 0$ the equations turns into

$$2y^2 + 4y + 2 = 0,$$

which has the unique solution $y = -1$. Therefore, in this case any pair $(x, -1)$ is a solution, so $a = 0$ does not satisfy the condition.

For $a \neq 0$, the given equation is quadratic in x:

$$ax^2 + 2a(2y-1)x + (3a+2)y^2 + (4-6a)y + 2 = 0.$$

For this equation,

$$\frac{D}{4} = a(a-2)(y+1)^2.$$

Now, if $a(a-2) \geq 0$, then our equation in x will have at least one solution for *any* y; therefore, as an equation in x and y, it will have infinitely many solutions. But if $a(a-2) < 0$, then the only possible value of y is $y = -1$, and the only corresponding x is $x = 3$. So the answer is $0 < a < 2$.

Inequalities

Recall that the inequality

$$ax^2 + bx + c \geq 0$$

(with $a > 0$) is true for all x if and only if

$$D = b^2 - 4ac \leq 0.$$

And the inequality

$$ax^2 + bx + c < 0$$

has a solution (for $a > 0$) if and only if the corresponding equation $ax^2 + bx + c = 0$ has two distinct real roots—that is, for $D > 0$.

The solution of the next problem, which is rather widely known, serves as a good example of how these properties of quadratic inequalities can be applied.

Problem 6. Prove the inequality

$$a^2 + b^2 + c^2 \geq ab + bc + ca,$$

where a, b, and c are arbitrary real numbers.

Solution. Rewrite the given inequality as a quadratic in a:

$$a^2 - (b+c)a + b^2 + c^2 - bc \geq 0.$$

Its discriminant D equals $-3(b-c)^2$. Since it is nonnegative, the inequality is true for all a, b, and c. In addition, we see that $D = 0$ only for $b = c$, so exact equality is attained only for $a = b = c$.

The next problem is similar, but a bit more difficult.

Problem 7. Prove the inequality

$$6a + 4b + 5c \geq 5\sqrt{ab} + 7\sqrt{ac} + 3\sqrt{bc},$$

where $a \geq 0$, $b \geq 0$, $c \geq 0$.

Solution. Substitute t for $\sqrt{a}$:

$$6t^2 - \left(5\sqrt{b} + 7\sqrt{c}\right)t + 4b + 5c - 3\sqrt{bc} \geq 0.$$

Again, compute the discriminant of the quadratic trinomial in t on the left side:

$$\left(5\sqrt{b} + 7\sqrt{c}\right)^2 - 24\left(4b + 5c - 3\sqrt{bc}\right)$$
$$= -71\left(\sqrt{b} - \sqrt{c}\right)^2 \leq 0.$$

This proves the inequality in question and shows that it turns into an equality only for $a = b = c$.

Problem 8. For what values of a does the inequality

$$-1 < \frac{ax^2 + x + 2}{x^2 + 1} < 3$$

hold for all x?

Solution. The given inequalities can be rewritten as

$$\begin{cases}ax^2 + x + 2 < 3x^2 + 3,\\ ax^2 + x + 2 > -x^2 - 1.\end{cases}$$

The first of them, $(a-3)x^2 + x - 1 < 0$, holds for all x if $a - 3 < 0$ and the discriminant $D = 1 + 4(a-3) = 4a - 11 < 0$ —that is, for

$$\begin{cases}a < 3,\\ a < \dfrac{11}{4}.\end{cases}$$

or $a < 11/4$. Similarly, for the second inequality we get $a > -1/3$, which combines with the first result to yield the answer: $-1/3 < a < 11/4$.

Problem 9. Find all values of a such that

$$|3\sin^2 x + 2a\sin x\cos x + \cos^2 x + a| \leq 3$$

for any x.

Solution. Before we can apply our method, the given inequality must be properly treated.

If $\cos x = 0$, it takes the form $|a + 3| \leq 3$. So the possible values of a can be confined to $-6 \leq a \leq 0$.

If $\cos x \neq 0$, divide both sides of the inequality by $\cos^2 x$ and substitute $t = \tan x$. In view of the fact that

$1/\cos^2 x = 1 + \tan^2 x$, this results in

$$|(a+3)t^2 + 2at + a + 1| \le 3(1 + t^2).$$

This inequality must hold for all t. It's equivalent to the pair of inequalities

$$-3 - 3t^2 \le (a+3)t^2 + 2at + a + 1 \le 3 + 3t^2.$$

They are obviously true for all t if $a = 0$. Finally, in the case $a \ne 0$, we have to write out and resolve for a the conditions that the two corresponding discriminants are nonpositive, as we did before. The answer is $-12/5 \le a \le 0$.

Ranges of functions: maxima and minima

By definition, a number a belongs to the range of a function $f(x)$ if and only if the equation $f(x) = a$ has a solution. When this equation reduces to a quadratic, our method may prove simpler and more convenient than analysis of the function by means of calculus.

Problem 10. Find the range of

$$y = \frac{x}{(x-1)^2}.$$

Solution. The problem is equivalent to the following question: for what values of a does the equation $x/(x-1)^2 = a$, or $ax^2 - (2a+1)x + a = 0$, have a solution?

If $a = 0$, a solution $(x = 0)$ exists. For $a \ne 0$, we have a quadratic inequality, and the problem boils down to solving the inequality $D = 4a + 1 \ge 0$. The answer here is $[-1/4, \infty)$.

Now let's consider some problems that are a little more difficult.

Problem 11. Find the minimum value of the function

$$y = x(x+1)(x+2)(x+3).$$

Solution. Since

$$y = (x^2 + 3x)(x^2 + 3x + 2),$$

let's substitute $t = x^2 + 3x$. Then the function takes the form $y = t(t+2)$. Applying our method again, we find that the equation $t(t+2) = a$ has solutions for $a \ge -1$, and its left side attains its minimal value $a = -1$ at $t = -1$. It remains to verify that the equation $x^2 + 3x = -1$ has a solution. Thus, the answer is -1.

Problem 12. Find the greatest of the values z such that for certain x and y the following equation holds:

$$2x^2 + 2y^2 + z^2 + xy + xz + yz = 4.$$

Solution. Consider this equation as an equation in x. Its discriminant must be nonnegative:

$$D = (y+z)^2 - 16y^2 - 8yz - 8z^2 + 32 \ge 0, \quad (1)$$

or

$$15y^2 + 6yz + 7z^2 - 32 \le 0.$$

This can be viewed as a quadratic inequality in y and it has a solution if and only if $9z^2 - 105z^2 + 15 \cdot 32 \ge 0$—that is, if $z^2 \le 5$. So z cannot be greater than $\sqrt{5}$. If $z = \sqrt{5}$, then it turns out that inequality (1) in y has a unique solution $(y = -1/\sqrt{5})$, and for these values of y and z the original equation has a solution in x (also unique). So the answer is $z = \sqrt{5}$.

The next problem can be reduced to the previous one, although it looks different from the outside.

Problem 13. The numbers x, y, and z satisfy the condition $x^2 + 2y^2 + z^2 = 2$. What is the greatest value that the expression $2x + y - z$ can take?

Hint. Put $t = 2x + y - z$ and substitute $z = 2x + y - t$ into the problem. We'll get a problem that differs from problem 12 only in its numerical coefficients. We leave it to the reader to complete the calculations. The answer is $t_{\max} = \sqrt{11}$.

Now it's time for you to try using the "power of the discriminant" yourself.

Exercises

1. Solve the following equation and system of equations:

(a) $5x^2 + 5y^2 + 8xy + 2x - 2y + 2 = 0$;

(b) $\begin{cases} x + y + z = 4, \\ 2xy - z^2 = 16. \end{cases}$

2. Prove the inequality $x^2 + 2xy + 3y^2 + 2x + 6y + 3 \ge 0$.

3. Find the minimum value of the function

$$y = \frac{2x^2 + 9x + 11}{3x^2 + 11x + 12}.$$

4. For what values of a does the inequality

$$\frac{x-2}{ax^2 - 2x + a - 2} < 1$$

hold for all x?

5. Find the minimum value taken by the expression $x + 5y$ if $x > 0$, $y > 0$, and $x^2 - 6xy + y^2 + 21 \le 0$.

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

IN THE LAB

Behind the mirror

How can we measure the thickness of the reflecting layer?

by N. M. Rostovtsev

WHEN MIRRORS ARE MADE, a polished glass surface is covered by a thin metal layer that reflects light very effectively. One widely used technique is to precipitate silver onto glass by chemical means.

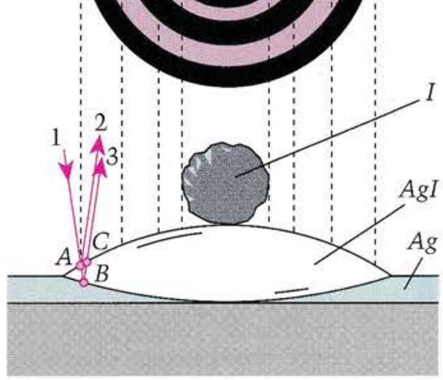
But, as you know, silver is a rather expensive metal. As in any production process, you need to know how much of each material is being used per item. How can you measure the thickness of the layer of silver on the back of a mirror? We'll, you'd have the answer already if you knew about the method invented in 1861 by the French physicist Armand Fizeau. This method is based on the chemical interaction of iodine and silver, and it is still widely used today.

To determine the thickness of the silver layer, a small grain of iodine is placed on it. Almost immediately the color of the silver changes near the grain. In 5 to 10 minutes (depending on the layer's thickness) a number of colored rings are formed around the grain and the mirror becomes transparent where the grain was placed. If the mirror is illuminated by a red light (or examined through a red filter), alternating red and black rings will be seen around the grain (the upper portion of the figure at right). What's happening here?

Iodine vapor (I_2) spreads from the grain and interacts with the silver (Ag), turning it into silver iodide (AgI)—a transparent, solid substance. Near the grain of iodine a layer of AgI is formed, which has a variable thickness and is similar to a double-convex lens (the bottom portion of the figure). The thickness of this "lens" decreases to zero at its rim. Light interference is responsible for the appearance of the rings.

When a light wave falls on the mirror (beam 1 in the figure), it is partially reflected from the external surface of the AgI layer (point A, beam 2), and it passes partially into the layer, where it is reflected from its internal surface—that is, from the Ag–AgI boundary (point B). Then the wave emerges from the mirror (point C, beam 3). The resulting reflected waves (beams 2 and 3) are coherent and so they interfere. The result of the interference depends on the difference δ in the path lengths of beams 2 and 3, which is

$$\delta = AB + BC.$$



If the incident angle of the light is approximately zero, $AB = BC$, and

$$\delta = 2AB = 2d, \tag{1}$$

where d is the thickness of the "lens" where the beam strikes its surface. The thickness d increases gradually toward the center of the lens, as does the path difference δ. When the mirror is illuminated by a monochromatic light (say, by red light), maxima (red rings) will be seen where the beams' path difference equals an integer number of wavelengths λ of the light in the silver iodide. The dark rings (minima) are formed where the path difference is an odd number of half-wavelengths of the light in this substance:

$$\delta = (2k-1)\frac{\lambda}{2}, \tag{2}$$

where $k = 1, 2, 3 \ldots$ is the number of the dark ring. The first number corresponds to the outermost dark ring.

Relationships (1) and (2) lead to a formula that describes the thickness of the AgI layer where dark ring number k is formed:

$$d = (2k-1)\frac{\lambda}{4}. \tag{3}$$

If the last dark ring (number k_{max}) coincides with the edge of the transparent spot at the center, the maximum thickness d_{max} of the layer is

$$d_{max} = (2k_{max}-1)\frac{\lambda}{4}. \tag{4}$$

Usually, however, the last dark ring does not coincide with the edge of the transparent spot. It's generally located some distance from it (as shown in the figure). So d_{max} is actually somewhat larger than the value given by equation (4). Let's estimate the magnitude of this discrepancy. The difference in the thickness of the "lens" at places corresponding to two adjacent dark rings is one half of the wavelength. Thus, the error in determining d_{max} doesn't exceed $\lambda/4$. So we can conclude that the maximum thickness is approximately equal to

$$d_{max} = (2k_{max} - 1)\frac{\lambda}{4} + \frac{\lambda}{4} = k_{max}\frac{\lambda}{2}. \qquad (5)$$

The wavelength λ of light traveling in a medium is related to its wavelength in a vacuum λ_0 by the relationship $\lambda = \lambda_0/n$, where n is the refractive index of the medium. Taking this into account, we get

$$d_{max} = k_{max}\frac{\lambda_0}{2n}. \qquad (6)$$

It turns out (see the problem at the end of the article) that the thickness h of the silver layer in a mirror is linked with the maximum thickness of the silver iodide d_{max} by the following formula:

$$h = \frac{d_{max}}{4} = k_{max}\frac{\lambda_0}{8n}. \qquad (7)$$

Inserting the refractive index of silver iodide ($n = 2.15$) into equation (7) yields the formula for the thickness of the silver layer:

$$h = k_{max}\frac{\lambda_0}{17.2}.$$

Now let's imagine we look at the object in the upper part of the figure through a red filter (the wavelength of red light is $6.2 \cdot 10^{-5}$ cm). In this case,

$$h = \frac{3 \cdot 6.2 \cdot 10^{-5}\,\text{cm}}{17.2} \cong 10^{-5}\ \text{cm}.$$

If the rings are observed in white light, k_{max} is related to the number of colored rings and λ is taken to be $5.6 \cdot 10^{-5}$ cm (the wavelength of green light, to which the human eye is most sensitive).

As we noted above, the accuracy of such measurements of the thickness of the AgI layer is $\lambda/4 = \lambda_0/4n$. Since the thickness of the Ag layer is 1/4 that of the AgI layer, the accuracy of the measurement of the Ag layer's thickness is $(1/4)(\lambda_0/4n) = \lambda_0/34.4$.

The simplest way to perform experiments aimed at measuring the thickness of the reflecting layer is to use a glass Christmas tree ornament. The ornament should be carefully broken into pieces about 1 × 1 cm in size. You can also use a small mirror. The silver layer on a mirror is usually covered with paint, but you can easily wipe it off with a cotton ball soaked in turpentine or another solvent. You can also do the experiments with sheets of mica coated with a thin layer of silver. Such sheets can be taken from 100–1,000-pF mica capacitors.

You can find grains of iodine in your school's chemistry lab. For our purposes, grains about the size of a poppy seed or a bit larger are quite suitable. If such grains are difficult to find, you can use a tincture of iodine. Soak a match in this solution and then cut off a small piece about 1–1.5 mm long. This "grain" can be used to measure the thickness of the silver layer.

If the rings are too close to one another, use a magnifying glass. You'll have an easier time seeing the rings and counting them if you look at them through a red filter.

Exercise

Show that the silver layer's thickness is about 1/4 that of the AgI, given that the density of silver is $10.5 \cdot 10^3$ kg/m^3, the density of silver iodide is $5.73 \cdot 10^3$ kg/m^3, the atomic weight of silver is 107.8, and the molecular weight of silver iodide is 234.7.

Note: Write down the reaction for the production of silver iodide and compare the volume occupied by the silver participating in the reaction with that of the resulting silver iodide. The mean horizontal cross-sectional areas of both layers are assumed to be equal.

MATH INVESTIGATIONS

Nonstandardly continued fractions

Yet another infinite process worthy of attention

by George Berzsenyi

SOME YEARS AGO, WHEN I was asked to initiate the Problem Section of *Math Horizons*, it seemed reasonable to appeal to many of the known problemists, and so I circulated a "Call for Problems." Some of the responses were most gratifying, and they provided me with enough ammunition for my two years as the editor of that section. In fact, I still have a few gems from that collection and will share some of them with my readers in the present column.

The three problems featured below were sent to me by a retired professor, Edwin O. Buchman, of Fullerton, California. They are his creations, and they had not appeared in print previously. As he indicated in his letter to me, it may be inappropriately difficult for my readers to **prove that each actually has a value** (in the sense that the sequences formed by truncating the expressions successively further to the right converge to a limit), but it might be more engaging to **determine the value of each, given that they do exist**. I would also like to recommend that you **create your own nonstandard continued fractions** and submit your challenges for possible inclusion in this column. By the way, the answers to the three problems are $\sqrt{3}$, $\sqrt[3]{2}$, $\sqrt[3]{12}$ in some order.

I am most grateful to Prof. Buchman for his contribution and promise to share more of his very different nonstandardly continued fractions with my readers in a future column. For the general theory of (standardly) continued fractions, the reader may wish to consult a book on elementary number theory—most of them devote at least a section to that topic.

$$\cfrac{3+\cfrac{3+\cfrac{3+\cfrac{3+\cdots}{1+\cdots}}{1+\cfrac{3+\cdots}{1+\cdots}}}{1+\cfrac{3+\cfrac{3+\cdots}{1+\cdots}}{1+\cfrac{3+\cdots}{1+\cdots}}}}{1+\cfrac{3+\cfrac{3+\cfrac{3+\cdots}{1+\cdots}}{1+\cfrac{3+\cdots}{1+\cdots}}}{1+\cfrac{3+\cfrac{3+\cdots}{1+\cdots}}{1+\cfrac{3+\cdots}{1+\cdots}}}}$$

$$1+\cfrac{1+\cfrac{1+\cfrac{1+\cfrac{1+\cdots}{3+\cdots}}{3+\cfrac{5+\cdots}{1+\cdots}}}{3+\cfrac{5+\cfrac{1+\cdots}{3+\cdots}}{1+\cfrac{5+\cdots}{1+\cdots}}}}{3+\cfrac{5+\cfrac{1+\cfrac{1+\cdots}{3+\cdots}}{3+\cfrac{5+\cdots}{1+\cdots}}}{1+\cfrac{5+\cfrac{1+\cdots}{3+\cdots}}{1+\cfrac{5+\cdots}{1+\cdots}}}}$$

$$2+\cfrac{2+\cfrac{2+\cfrac{2+\cfrac{2+\cdots}{6+\cdots}}{6+\cfrac{11+\cdots}{4+\cdots}}}{6+\cfrac{11+\cfrac{2+\cdots}{6+\cdots}}{4+\cfrac{11+\cdots}{4+\cdots}}}}{6+\cfrac{11+\cfrac{2+\cfrac{2+\cdots}{6+\cdots}}{6+\cfrac{11+\cdots}{4+\cdots}}}{4+\cfrac{11+\cfrac{2+\cdots}{6+\cdots}}{4+\cfrac{11+\cdots}{4+\cdots}}}}$$

Feedback

I am most thankful to my friend Prof. Murray Klamkin for telling me about further references to the problem area featured in my November/December 1995 column in *Quantum*. They include an article by V. Faber and J. Mycielski (*American Mathematical Monthly* (1986), pp. 796–801), Murray's Editorial Note to a problem by Richard Bellman (Problem 63–9*, in *SIAM Reviews* (1963), p. 274), a reaction by W. Franck to that problem (*SIAM Reviews* (1965), pp. 503–12), and the references cited in these papers. He also indicated that there are some related publications by Anatole Beck.

With respect to the problems featured in my May/June 1995 column in *Quantum*, my colleague John Rickert managed to make some further progress. He proved that $G(5, k)$ is a factor of $3125k^4 + 625k^2 + 1$ and $G(7, k)$ is a factor of $823543k^6 + 6000099k^4 + 12005k^2 + 1$. In particular, it turns out that $G(5, 1) = 341$, $G(5, 2) = 52501$, $G(5, 3) = 258751$, $G(5, 4) = 810001$, and $G(5, 5) = 1968751$. They provide wonderful illustrations for "Guy's Strong Law of Small Numbers" since, for example, $\gcd(n^5 + 5, (n + 1)^5 + 5) = 1$ for $n = 1, 2, 3, \ldots, 533359$, while $\gcd(533360^5 + 5, 533361^5 + 5) = 1968751$.

AT THE BLACKBOARD II

Shall we light a fire in the fireplace?

Or shall we listen to the equation that says it won't make any difference at all?

by Victor Lange

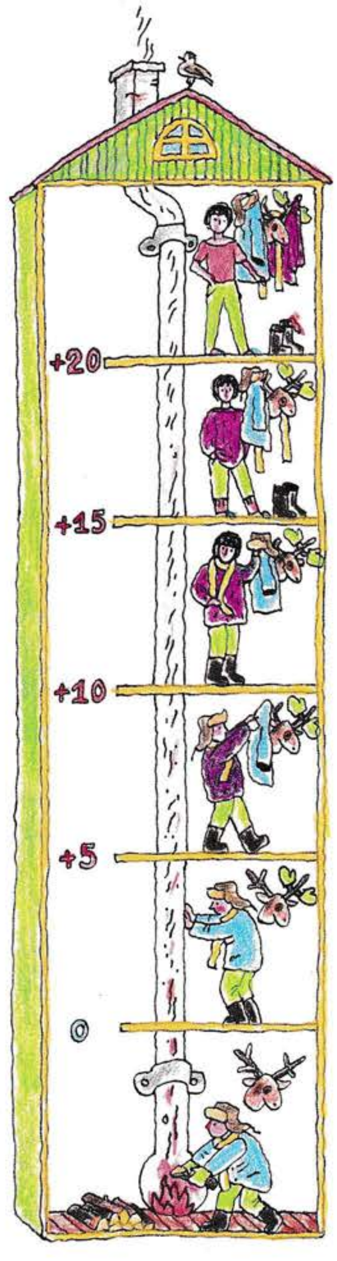
SUPPOSE WE ASK A SPECIAList in heat engineering and a person with no experience in this field the same question: "Why do we turn on our furnaces in the winter?" The specialist's answer probably will be as follows: "To increase the internal energy of the air in the house"; whereas the layperson will simply say: "So that it's warmer in the house." And the nonspecialist will have given the more precise answer.

Improbable as it may seem at first glance, the internal energy of the air in the room remains the same after the firewood has been burned in the potbellied stove.

Suppose you had a cabin in the woods, and you went there in late autumn. The temperature in the rooms is about 0°C, and you decide to light the wood stove in order to get warm. While the column of alcohol in the thermometer is rising to the 20°C mark, let's estimate the change in the internal energy of the air in the cabin.

Under conditions close to normal (that is, a temperature not far from 0°C and a pressure of about 10^5 N/m^2), the air is similar to an ideal gas whose internal energy is proportional to its mass m and its absolute temperature T:

$$U = \alpha m T,$$



where α is a proportionality factor—a constant for any given gas.

When the cabin is heated, the air gets warmer, it expands, and a portion of it escapes through the small openings around the doors and windows and through the pores and cracks in the walls, so that the mass of air in the cabin changes. The volume of the cabin and the air pressure, however, remain constant. So the change in the internal energy after heating can be written as follows:

$$\begin{aligned}\Delta U &= U_2 - U_1\\ &= \alpha m_2 T_2 - \alpha m_1 T_1\\ &= \alpha(m_2 T_2 - m_1 T_1).\end{aligned}$$

To solve the problem we use the ideal gas law:

$$PV = \frac{m}{\mu} RT.$$

Since, as we noted, the pressure in the cabin remains equal to the outside pressure and the volume of the rooms is unchanged, and the only thing that changes is the air mass, we obtain

$$m_2 T_2 = m_1 T_1 = \frac{PV\mu}{R} = \text{constant}.$$

So, after the firewood is burned, the internal energy of the air in the cabin really doesn't change, which means that all the dissipated heat escapes through the cracks to the outside world.

So why do we heat our homes? Well, even though our bodies are not sensitive to the total internal energy of the air in the house, they are very sensitive to temperature, which is determined by the average energy per molecule.

In 1938 the German astrophysicist R. Emden noted that our living quarters must be heated for the same reason that life on Earth is impossible without a constant influx of solar heat. And it has nothing to do with the amount of incident energy, which is reradiated (down to the last iota) to the surroundings, just as a person's mass doesn't change even though food is continually ingested. We require a certain range of ambient temperatures in order to stay alive, and that's why, equation or no equation, we stoke our furnaces and light fires in our fireplaces.

Grab that chain of thought!

Did an article in this issue of *Quantum* make you think of a related topic? Write down your thoughts. Then write to us for our editorial guidelines. 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 target readership of high school and college students.

Send your inquiries to:

Managing Editor
Quantum
1840 Wilson Boulevard
Arlington VA 22201-3000

Bulletin Board

A yearly PROMYS

This year's Program in Mathematics for Young Scientists (PROMYS) will be held at Boston University from June 30 to August 10. 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 the art of mathematical discovery—numerical exploration, formulation and critique of conjectures, and techniques of proof and generalization. More experienced participants may also study other advanced topics. Problem sets are accompanied by daily lectures given by research mathematicians with intensive experience in Prof. Arnold Ross's long-standing Summer Mathematics Program at Ohio State University. In addition, a highly competent staff of 18 college-age counselors lives in the dormitories and is always available to discuss mathematics with students. Each participant belongs to a problem-solving group that meets with a professional mathematician three times a week. Special lectures by outside speakers offer a broad view of mathematics and its role in the sciences.

PROMYS is a residential program designed for 60 students entering grades 10 through 12. Admission decisions will be based on the following criteria: 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.

The estimated cost to participants is about $1,300 for room and board. Books may cost an additional $100. Financial aid is available. PROMYS is dedicated to the principle that no student will be unable to attend because of financial need.

PROMYS is directed by Prof. Glenn Stevens. 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, 1996.

What's happening?

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

What's on your mind?

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

What's our address?

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

Be a factor in the QUANTUM equation!

Shady CyberTeaser

The January/February CyberTeaser (brainteaser B163 in this issue) at *Quantum*'s home page on the World Wide Web drew a wider range of responses than usual. The answers fell into five broad categories. Some gave the correct answer (that is, *our* answer!) with breathtaking succinctness; some gave our answer, but with caveats or elaborations; some gave our answer, but with unnecessary conditions or restrictions; some explicitly rejected our answer; and some thought it was a trick question and supplied tricky answers.

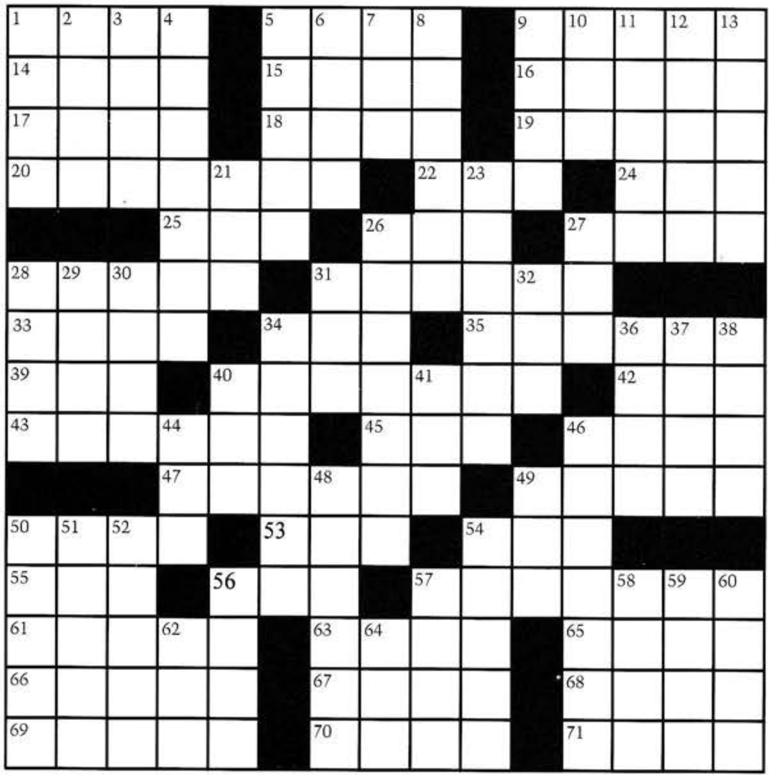
In addition to submissions from many states in the US, we received entries from Canada, Israel, Italy, Mexico, and Singapore. The following entrants were the first ten to submit an answer that satisfied our quirky but earnest CyberJudge:

Roger Khazan (Waltham, Massachusetts)
Ophir Yoktan (Jerusalem, Israel)
Joe Pascoe (Tonawanda, New York)
Oleg Shpyrko (Rochester, New York)
Byron Peterson (Regina, Saskatchewan)
Brian Platt (Woods Cross, Utah)
Sterling Marshall (St. Louis, Missouri)
Matthew Padilla (Lombard, Illinois)
David Craig (Clemson, South Carolina)
Sheena Delgado (Lansing, Michigan)

We thank all those who sent us their answers, and we invite everyone to take a shot at the current CyberTeaser at http://www.nsta.org/quantum.

Criss × cross science

by David R. Martin



Across

1 Pilewort fiber
5 Gazelle
9 Star in Aries
14 1936 medicine Nobelist Sir Henry ___
15 City in Utah
16 Stay
17 1963 medicine Nobelist ___ Hodgkin
18 ___ black (carbon pigment)
19 Poet Audre Geraldin ___ (1934–)
20 Nuclear ___
22 Potential difference: abbr.
24 Sense organ
25 Charged particle
26 ___ Lanka
27 Type of current
28 African
31 $C_{27}H_{46}O$
33 Comply
34 Large snake
35 Highly basic compound
39 Sick
40 Potential difference
42 Christmas month: abbr.
43 1970 Chemistry Nobelist Luis ___
45 ___ Got a Secret
46 10^{-9}: pref.
47 Man to be married
49 1975 Physiol. Nobelist Howard ___
50 Plant part
53 "___ the season"
54 Fundamental physical truth
55 Whir
56 Unit of time: abbr.
57 ___ reality
61 Words for the audience
63 A meson
65 ___ *de choc*
66 Type of boom
67 Great lake
68 Pleasant
69 ___ null
70 Semiconductor devices: abbr.
71 Large plasma ball

Down

1 44,463 (in base 16)
2 Spanish painter
3 Expression of woe
4 Mass/volume
5 Large intestine section
6 Algerian seaport
7 Sleep motion: abbr.
8 Current unit
9 ___-life
10 Blood group system
11 Stuck in mud
12 712,109 (in base 16)
13 Wary
21 Debt paper
23 Optical illusion
26 Branch of mechanics
27 Large deer
28 Vaporize
29 Competent
30 Actress ___ Carter
31 Colloidal suspension
32 Spanish cheer
34 Related to B_2O_3
36 Swedish botanist ___ Afzelius (1750–1837)
37 Film director ___ Riefenstahl
38 Computer image
40 7 to Brutus
41 Street: abbr.
44 On's partner
46 Force units
48 Element 28
49 Distilled coal
50 Tibetan city
51 Chlorinated lime and boric acid
52 Organic compound
54 Shortest paths between points
56 Hyperbolic function
57 Null and ___
58 Specified measure
59 44,490 (in base 16)
60 Smile
62 Type of stick
64 Exist

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE NOVEMBER/DECEMBER PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M161

Let d be the greatest common divisor of a and b. Since the number

$$\frac{a+1}{b}+\frac{b+1}{a}=\frac{a^2+b^2+a+b}{ab}$$

is an integer and ab is divisible by d^2, the numerator $a^2 + b^2 + a + b$ is divisible by d^2. But $a^2 + b^2$ is also divisible by d^2. Therefore, $a + b$ is divisible by d^2, and so $\sqrt{a+b} \geq d$.

M162

Using the condition $a + b + c = 0$, we get

$$\begin{aligned} a^4 + b^4 + c^4 &= a^2(b+c)^2 + b^2(c+a)^2 + c^2(a+b)^2 \\ &= 2(a^2b^2 + b^2c^2 + c^2a^2) + 2abc(a+b+c) \\ &= (a^2+b^2+c^2)^2 - a^4 - b^4 - c^4, \end{aligned}$$

so $2a^4 + 2b^4 + 2c^4 = (a^2 + b^2 + c^2)^2$.

M163

(a) For any even n the $n \times n$ chessboard can be divided into $n^2/2$ "dominoes" of two squares each (fig. 1). The winning strategy for the first player is to always move the checker onto the second square of the domino where it is located. This can be done on the very first move. Then the second player will be forced to move into a new domino, so the first player will again be able to go to the second square of the second domino, which will force the second player to "open" a third domino, and so on. Thus, the first player can always respond to any move of the second player and, therefore, wins.

For odd n, we can divide the $n \times n$ chessboard into dominoes without the initial corner square (fig. 2). After the first move of the first player, the second player can forget about the initial square, apply the strategy described above to the domino-tiled "notched" chessboard, and thus win.

(b) If the game starts from a square next to a corner square, the first player can always force a win.

For even n the strategy remains exactly the same as in part (a). For odd n we notice that the second player can never get onto the corner square, because it's the "wrong" color. (If the starting square is white, the first player always goes to black squares and the second to white, whereas the corner square is black.) So the first player can simply apply the winning strategy on the "notched" chessboard described in part (a) for odd n.

We leave it to the reader to investigate this game for an arbitrary initial position of the checker.

M164

Let $A_1A_2...A_n$ be the larger polygon and let $B_1B_2...B_n$ be the polygon inscribed in $A_1A_2...A_n$ (with B_1 on the side A_1A_2, B_2 on A_2A_3, ...). Suppose first that the center O of the given circle lies *inside* the larger polygon (fig. 3). Then the segments OA_1, OB_1, OA_2, OB_2, ..., OA_n, OB_n divide this polygon into triangles OA_1B_1, OB_1A_2, ..., OB_nA_1 whose total area equals the area A of $A_1A_2...A_n$. Consider the pair of triangles OB_1A_2 and OA_2B_2 (fig. 4). They lie on different sides of their common side OA_2, and it's not difficult to see that the sum of their respective altitudes h_1 and h_2 dropped on OA_2 is no greater than B_1B_2. So the sum of their areas can be estimated as

$$\frac{1}{2}OA_2\cdot h_1+\frac{1}{2}OA_2\cdot h_2 \leq \frac{1}{2}OA_2\cdot B_1B_2 = \frac{1}{2}R\cdot B_1B_2.$$

A similar estimation is true for any

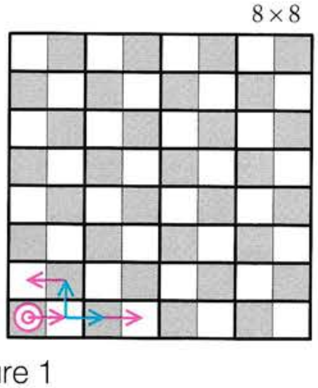


Figure 1

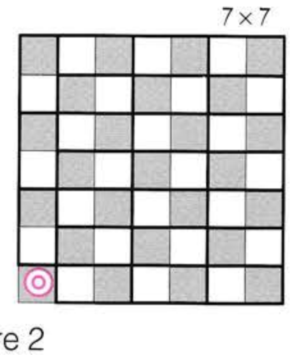


Figure 2

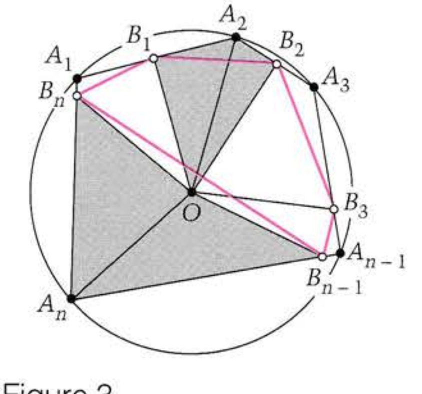


Figure 3

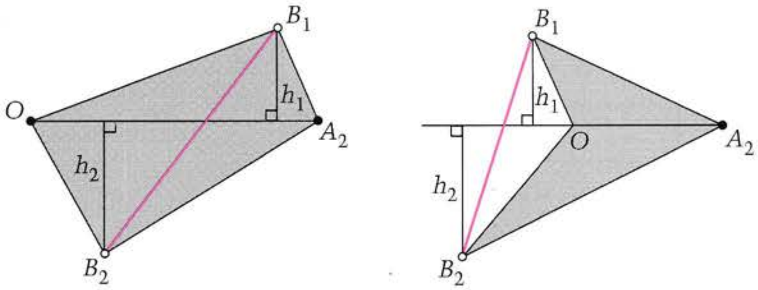


Figure 4

pair of these triangles with a common side OA_k, $k = 1, 2, \ldots, n$. Thus we get

$$\begin{aligned} A &= \left[\operatorname{area}(OB_1A_2) + \operatorname{area}(OA_2B_2)\right] + \\ &\quad \cdots + \left[\operatorname{area}(OB_nA_1) + \operatorname{area}(OA_1B_1)\right] \\ &\le \frac{1}{2}R\cdot B_1B_2 + \cdots + \frac{1}{2}R\cdot B_nB_1 \\ &= \frac{1}{2}R(B_1B_2 + \cdots + B_nB_1) \\ &= \frac{1}{2}RP, \end{aligned}$$

or

$$P \ge \frac{2A}{R}.$$

It is also possible that the center O be *outside* the polygon $A_1A_2\ldots A_n$. In this case one of its sides—say, A_nA_1—separates the other sides from point O (fig. 5), the sum of the areas of triangles OA_1B_2, OB_2A_2, ..., $OB_{n-1}A_n$ gives the area of the polygon $OA_1A_2\ldots A_n$, and we have to *subtract* the areas of triangles OA_nB_n and OB_nA_1 (which make the area OA_nB_1) from this sum to obtain

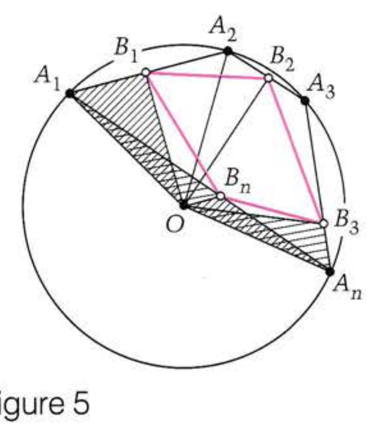


Figure 5

the area A of $A_1A_2\ldots A_n$. However, this case requires only a minor amendment to the proof above. We note that if two triangles with a common base—say, OA_1B_1 and OA_1B_n—are on the same side of this base, then the absolute value of the *difference* of their heights h_1 and h_n is no greater than B_1B_n (see figure 6). So now we have

$$\begin{aligned} &\operatorname{area}(OA_1B_1) - \operatorname{area}(OB_nA_1) \\ &= \frac{1}{2}OA_1(h_1 - h_n) \\ &= \frac{1}{2}R|h_1 - h_n| \\ &\le \frac{1}{2}R\cdot B_1B_n, \end{aligned}$$

and, similarly,

$$\operatorname{area}(OB_{n-1}A_n) - \operatorname{area}(OA_nB_n) \le \frac{1}{2}R\cdot B_{n-1}B_n.$$

Adding these two inequalities to the estimates for the areas of the pairs of triangles OA_1B_1 and OB_1A_2, ..., $OB_{n-2}A_{n-1}$ and $OA_{n-1}B_{n-1}$ obtained above, we get the area A on the left side and ½ $R \cdot P$ on the right, which completes the proof.

It follows from the proof that for a polygon $A_1A_2\ldots A_n$ that contains its circumcenter, our inequality $P \le 2A/R$ becomes an exact equality if and only if the sides of $B_1B_2\ldots B_n$ are perpendicular to the corresponding radii OA_k ($B_1B_2 \perp OA_2$, ..., $B_nB_1 \perp OA_1$). Such a polygon $B_1B_2\ldots B_n$ may not exist (it never exists in the case of figure 5), but if it does, it has the smallest possible perimeter $2A/R$.

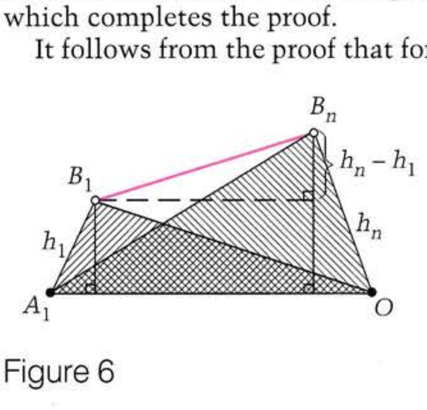


Figure 6

In particular, you can use this remark to prove that *a triangle $B_1B_2B_3$ inscribed in an acute triangle $A_1A_2A_3$ has the smallest possible perimeter if and only if its vertices are the bases of the altitudes of $A_1A_2A_3$.*

M165

For brevity, any side or diagonal of the given $(6n + 1)$-gon will be called a segment. Also, since all the triangles we consider are isosceles triangles whose sides are such "segments," we'll call them simply "triangles."

Denote by N_{bb}, N_{br}, and N_{rr} the numbers of segments that have two blue endpoints, one blue and one red endpoint, or two red endpoints, respectively. Clearly $N_{bb} = B(B-1)/2$, $N_{br} = BR$, and $N_{rr} = R(R-1)/2$, where $B = (6n+1) - R$ is the number of blue points. Note that these numbers do not depend on the arrangement of colors. We'll try to express the number of monochromatic triangles (triangles with vertices of the same color) in terms of these numbers.

Denote by T_3, T_2, T_1, T_0 the numbers of triangles with 3, 2, 1, and 0 blue vertices. Then the number in question is $T_3 + T_0$. Every segment is a side of exactly three (isosceles) triangles (fig. 7 on the next page). (This is true only if the number of sides of the polygon is neither even nor divisible by 3—that is, has the form $6n + 1$ or $6n + 5$.) Therefore, if we count the number of sides with two blue endpoints in all the triangles, we'll count each of the "blue–blue" segments three times, which gives $3N_{bb}$. On the other hand, a triangle with three blue vertices has three such sides, a triangle with two blue vertices has one, and other triangles don't

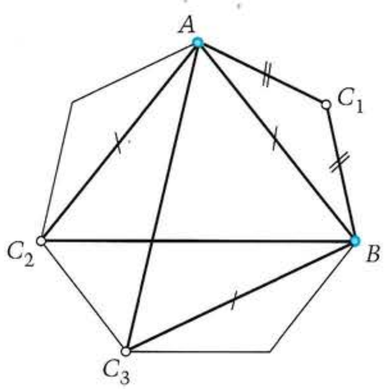


Figure 7

have them at all, so the same number can be written as $3T_3 + T_2$. Thus,

$$3N_{bb} = 3T_3 + T_2.$$

In the same way we can show that

$$3N_{br} = 2T_2 + 2T_1$$

and

$$3N_{rr} = T_1 + 3T_0.$$

Adding the first and the third relations and subtracting the second divided by two, we get the desired expression:

$$3\left(N_{bb} + N_{rr} - \frac{N_{br}}{2}\right) = 3(T_3 + T_0),$$

or

$$T_3 + T_0 = N_{bb} + N_{rr} - \frac{N_{br}}{2},$$

which completes the proof.

Physics

P161

In order not to break the table's legs, the force applied to each leg must not exceed $m/4$, as follows from the statement of the problem.

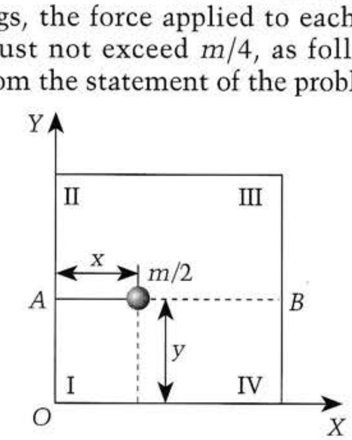


Figure 8

Let's see what force f_1 acts on leg I (fig. 8) when a body of mass $m/2$ lies on the table at point (x, y) (point I is taken as the origin of the coordinates.)

Let f_2, f_3, f_4 be forces applied to legs II, III, IV, respectively. Clearly the effect of forces f_1 and f_2 is equivalent to that of the force $f_{12} = f_1 + f_2$ applied to point A with coordinates $(0, y)$; similarly, $f_{34} = f_3 + f_4$ applied to point B (l, y) is equivalent to the effect of forces f_3 and f_4 (l is the length of the table).

The equilibrium condition yields

$$f_{12} + f_{34} = \frac{mg}{2},$$

$$f_{12}x = f_{34}(l - x).$$

from which we get

$$f_1 + f_2 = f_{12} = \frac{mg}{2}\frac{l - x}{l}. \qquad (1)$$

On the other hand, $f_1 y = f_2(l - y)$, and so

$$f_1 = f_2\frac{l - y}{y}. \qquad (2)$$

Combining equations (1) and (2) we get

$$f_1 = \frac{mg}{2}\frac{(l - x)(l - y)}{l^2}.$$

In order that a leg not be broken, the following condition must be met:

$$f_1 = \frac{mg}{2}\frac{(l - x)(l - y)}{l^2} < \frac{mg}{4}$$

—that is, a mass $m/2$ can be placed at the point characterized by the relationship

$$y > l\left(1 - \frac{l}{2(l - x)}\right).$$

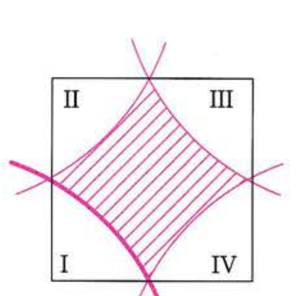


Figure 9

The graph of the function

$$y(x) = l\left(1 - \frac{l}{2(l - x)}\right)$$

is shown by the thick red line in figure 9. For reasons of symmetry it's clear that the set of "safety" points is the shaded curvilinear tetragon shown in figure 9.

P162

The air pressure P_C in the middle of the channel is equal to the pressure in a radial shaft at a distance $a/\sqrt{2}$ from the Moon's center (where $a = D/2$ is the radius of the Moon), and the pressure we seek is equal to P_C—that is, the pressure at the top of the shaft. What we need is to find how the pressure varies with the depth of the shaft.

Let's consider an air column of small height Δy located at a distance y from the Moon's center and write down its equilibrium condition:

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

where ρ is the density of air, g is the acceleration due to gravity at this point, and ΔP is the pressure difference at the heights $(y + \Delta y)$ and y. The air density can be found from the ideal gas law:

$$PV = \frac{m}{\mu}RT,$$

where m is the mass of air, $\mu = 14.5$ g/mole is the molecular mass of air, and the density of the air is

$$\rho = \frac{m}{V} = \frac{\mu}{RT}P,$$

where P is the air pressure at the chosen depth.

Now let's find how the acceleration due to gravity changes as we go deeper into the Moon. To do this, we imagine a sphere of radius $y < a$ inside the Moon, which is itself a sphere of radius a (fig. 10). If a small test mass m_0 is placed at the surface

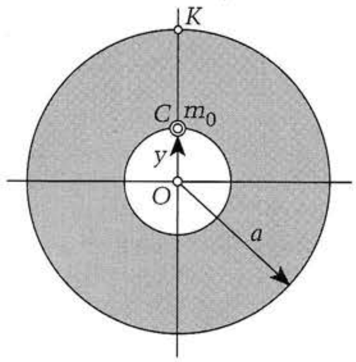


Figure 10

of this small sphere, it will be affected by the entire lunar gravitation as if only the small sphere were present. The sum of all the gravitational forces acting on the test mass from the shaded part of figure 10 is zero. So the acceleration due to gravity g at point C (with coordinate y) is determined by the formula

$$G\frac{m_0 m_y}{y^2} = m_0 g,$$

and at point K by

$$G\frac{m_0 M}{a^2} = m_0 g_M,$$

where m_y and M are the masses of the small and the large spheres, respectively. Dividing one formula by the other and taking into account that the sphere's mass is proportional to its volume (or to the third power of its radius), we obtain

$$g = g_M \frac{y}{a}.$$

Now let's insert the formulas for ρ and g into our equilibrium equation above:

$$\frac{\mu P}{RT}\frac{g_M y}{a}\Delta y = -\Delta P,$$

or

$$\frac{\Delta P}{P} = -\frac{\mu g_M}{RTa} y\,\Delta y.$$

If we replace the increments ΔP and Δy by infinitely small differentials dP and dy, we get

$$\frac{dP}{P} = -\frac{\mu g_M}{RTa} y\,dy.$$

This equation needs to be integrated:

$$\int_{P_C}^{P_K}\frac{dP}{P} = \int_{y_C}^{y_K} -\frac{\mu g_M}{RTa},$$

or

$$\ln P_K - \ln P_C = -\frac{\mu g_M}{RTa}\left(\frac{a^2}{2} - \frac{a^2}{4}\right).$$

It follows that

$$\ln\frac{P_C}{P_K} = \frac{\mu g_M a}{4RT} \cong 9.1,$$

or

$$\frac{P_C}{P_K} \cong 9{,}000.$$

Thus, the pressure in the channel at the lunar surface is less than the pressure at the middle of the channel by a factor of 9,000:

$$P_A = P_B = P_K \cong \frac{P_C}{9{,}000} \cong 11 \text{ Pa}.$$

This result shows that it is theoretically possible for cavities filled with gas to exist inside the Moon.

P163

Let's find the excess pressure ΔP_1 (with respect to 1 atm) that corresponds to a change in the melting temperature for ice of $\Delta t_1 = t_1 - t_0$:

$$\Delta P_1 = \frac{L(t_1 - t_0)\rho_w \rho_i}{T_0(\rho_i - \rho_w)} = 1.41\cdot 10^5 \text{ Pa}.$$

This pressure corresponds to an ice column of height

$$H_i = \frac{\Delta P_1}{\rho_i g} = 15.64 \text{ m}.$$

The ice column will be located in the upper part of the tube, and the water will be under it. The height of the water column H_w can be obtained from the condition of conservation of the total mass of the water and ice:

$$\rho_w H = \rho_i H_i + \rho_w H_w,$$

from which we obtain

$$H_w = H - \frac{\rho_i}{\rho_w} H_i = 5.61 \text{ m}.$$

So the change in the height of the tube's contents is

$$\Delta H = H_i + H_w - H = \frac{\lambda(t_0 - t_1)}{T_0 g}$$

$$= 1.25 \text{ m}.$$

P164

The strength of the electric field between the centers of the plates (indeed, not just the centers but almost everywhere, but not too close to the edges of the plates) is determined as usual:

$$E = \frac{Q}{\varepsilon_0 S}.$$

Thus, the potential difference between the centers of each plate is

$$\Delta\phi_0 = Ed = \frac{Qd}{\varepsilon_0 S}.$$

The perpendicular component of the field strength near a corner is 1/4 times that in the center. This can be shown by adding three pairs of similar charged plates to our "capacitor" so as to place this corner in the middle of the large "capacitor" that results. In so doing, the field strength increases by a factor of four, and the corner ends up at the center. So the potential difference between the plates, if measured near the corner, is

$$\frac{1}{4}\Delta\phi_0 = \frac{Qd}{4\varepsilon_0 S}.$$

Now let's move along the chosen closed circuit: from the center of one plate to its corner, then to the similar corner of the other plate, then to its center, and finally to the starting point. The potential difference $\Delta\phi$ that we seek between the plate's center and its corner has been passed twice (there's no problem with the signs—the charge of the second plate is opposite that of the first plate, but the direction of our path was also opposite). The total work performed by the electric field is zero:

$$2\Delta\phi - \Delta\phi_0 + \frac{1}{4}\Delta\phi_0 = 0,$$

so

$$\Delta\phi = \frac{3}{8}\Delta\phi_0 = \frac{3Qd}{8\varepsilon_0 S}.$$

P165

The free surface of a liquid in a rotating vessel takes the shape of a rotational paraboloid. Consider a small element of mass m located on the surface at a distance x from the axis of rotation (figure 11). This element is affected by the force of gravity mg and the reaction force N from the other parts of the liquid, because the liquid's free surface is a surface of constant pressure.

As the chosen element rotates uniformly along a circle of radius x, the resulting force is horizontal to the axis of rotation and equals $m\omega^2 x$. The parallelogram of forces (fig. 11) yields

$$\tan\alpha = \frac{\omega^2 x}{g}, \qquad (1)$$

where α is the angle of inclination of the tangent to the liquid surface at the chosen point. Notice that it follows from equation (1) that the surface of a rotating liquid is a paraboloid.

Now consider a sphere of radius R and determine the angle of inclination β of the tangent to this sphere at a point at a distance x from its vertical diameter. Figure 12 shows that

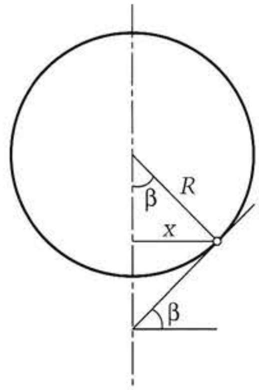


Figure 12

$$\sin\beta = \frac{x}{R}. \qquad (2)$$

Strictly speaking, equations (1) and (2) describe different surfaces. However, if we're interested in surface sites located not very far from the axis (so the corresponding angles of inclination are small enough), the trigonometric functions sine and tangent can be approximated by the values of the angles in radians, and we get approximately

$$\alpha \cong \frac{\omega^2 x}{g}, \quad \beta \cong \frac{x}{R}.$$

This shows that the best approximation for the surface of a rotating liquid is a sphere of radius $R = g/\omega^2$. The focal length of this surface (a concave mirror) is

$$F = \frac{R}{2} = \frac{g}{2\omega^2}.$$

(The focal length of parabolic mirror can be found directly. One should use the canonical equation for a parabola and its focal property.)

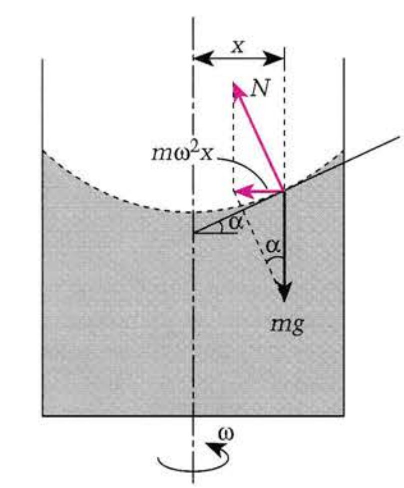


Figure 11

Brainteasers

B161

The unique solution can be easily guessed. Or, we can reason mathematically as follows. If s and c are the numbers of stools and chairs in the room, then for the total number of legs we have the equation $5s + 6c = 39$, which must be solved in nonnegative integers. From this equation we immediately see that s is odd, divisible by 3, and satisfies the inequality $5s \le 39$—that is, $s \le 7$. These conditions uniquely determine the value of s, which is 3. It follows that $c = 4$. (Alternatively, we could derive from the equation that c yields a remainder of 4 when divided by 5, because $5s + 6c \equiv 4 \pmod 5$ and $39 \equiv 4 \pmod 5$. This gives $c = 4$, since $6c \le 39$.) (V. Dubrovsky)

B162

The smallest possible sum is 102. The three significantly distinct arrangements are shown in figure 13 (we can also swap the two upper horizontal tiles in each of them). A direct trial-and-error search will show that there are no other solutions.

B163

At the North and South poles, the Sun is at the same height all day long (except for their respective "polar nights," when the Sun doesn't shine at all). So the shadow of any object at the poles "walks around" the same spot all day, and its length stays the same.

B164

Put the 1- and 2-kopeck coins on the left pan and the 3-kopeck

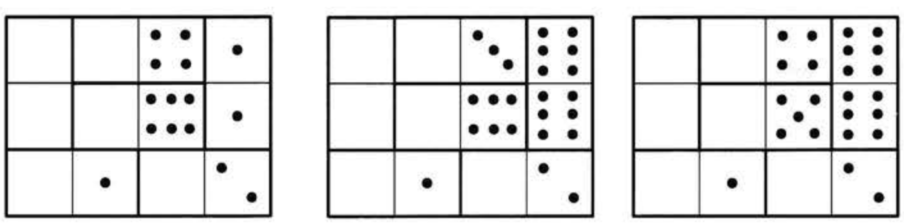


Figure 13

coin on the right pan. If the balance is in equilibrium, the remaining 5-kopeck coin is defective. If not, we put the 2- and 3-kopeck coins on the left pan and the 5-kopeck coin on the right pan. In the case of equilibrium, the 1-kopeck coin is defective. Otherwise, the 1- and 5-kopeck coins are good, and we compare the results of the two weighings. If they are the same (the left pan is heavier both times or lighter both times than the right pan), then the 2-kopeck coin is defective, because it remains on the same pan. Two different results (the left pan is heavier one time and lighter the other time) indicate that the 3-kopeck coin is defective.

B165

In the case of the square table and round napkins, consider the center and the four corners of the table (fig. 14a). At least two of these five points must be covered by the same round napkin. So the diameter of the napkin can't be smaller than the distance between these points, which is the smallest and equal to half the table's diagonal for the center and a corner of the table.

If four square napkins cover a round table, they cover its circumference, so at least one of them covers an arc equal to one quarter of the circumference (fig. 14b). The distance between the endpoints of this arc is $R\sqrt{2}$, where R is the table's radius, and the largest distance between points on a napkin is equal to its diagonal—that is, to $a\sqrt{2}$, where a is the side length of a napkin. Therefore, $a \geq R$.

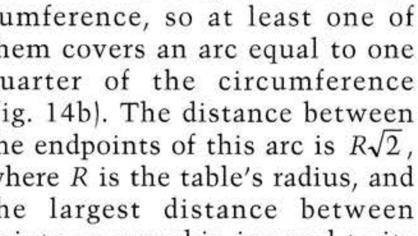


Figure 14

Toy Store

1. The recursive equation can be rewritten as $(t_n + 1) = 2(t_{n-1} + 1)$, from which it follows that $t_n + 1 = 2^{n-1}(t_1 + 1) = 2^n$.

Every symbol added at a certain step of the "binary scale" process to the "word" obtained at the previous step is surrounded by two new symbols at the next step. So the number of added symbols doubles at every step. Since we started with one symbol, the total number after n steps is $1 + 2 + 4 + \ldots + 2^{n-1}$.

2. The rule is very simple: the disks whose numbers are the same parity as n always move clockwise; the others move counterclockwise. This can be proved by induction using the recursive relation for T_n in the article.

3. The first idle peg is B for even n and C for odd n. For a proof, use the recurrence for T_n in the article and mathematical induction.

4. Let $k_n k_{n-1} \ldots k_1$ be the binary representation of the number k, $0 \leq k \leq 2^n - 1$ (if $k < 2^{n-1}$, we complete its representation to n digits with zeros at the left) and let $p_n p_{n-1} \ldots p_1$ be the code of the position created by the kth move of our disks (or the initial position for $k = 0$). It will be convenient to replace the labels A, B, C of the pegs in this code by the numbers 0, 1, 2, respectively. The string of numbers thus obtained will be denoted $q_n q_{n-1} \ldots q_1$. We'll prove that the two strings of numbers are linked by the following relation:

$$k_n = q_n,$$
$$k_{i+1} - k_i \equiv (-1)^{n+1}(q_{i+1} - q_i) \pmod 3$$
$$\text{for } i = 1, \ldots, n-1 \qquad (1)$$

(recall that the notation $a \equiv b \pmod 3$ means that $a - b$ is divisible by 3). This relation gives a very simple rule for finding k from the code $q_n q_{n-1} \ldots q_1$ of a given position: we read the code from left to right and write the binary notation of k in the same order, starting with $k_n = q_n$ and changing the last written digit k_{i+1} (0 to 1, 1 to 0) if and only if there is a change in the same place in the code ($q_i \neq q_{i+1}$). That is, if $q_i = q_{i+1}$, we write $k_i = k_{i+1}$. An example for $n = 5$ is given in the following table:

i	5	4	3	2	1
q_i	1	2	0	1	1
$q_i =/\neq q_{i+1}$		$\neq$	$\neq$	$\neq$	$=$
k_i	1	0	1	0	0

Given position: disk 3 on peg A, disks 5, 2, 1 on peg B, disk 4 on C
Number of moves: $k = 2^4 + 2^2 = 20$

Thus, the number of symbol changes in the string $Ap_n \ldots p_i$ or, equivalently, $0q_n \ldots q_1$ is the same as in $0k_n \ldots k_1$. The first change and every odd change in $0q_n \ldots q_1$ are from 0 to 1; all even changes are from 1 to 0. So if the number of changes is even, then $k_1 = 0$ and k is even; otherwise, $k_1 = 1$ and k is odd. And this is the answer to the second part of the problem.

Relation (1) can be proved by induction over n. For $n = 1$ it's obvious ($k_1 = q_1$). Suppose it's true for the $(n-1)$-disk puzzle ($n \geq 2$). The relation $T_n = \overline{T}_{n-1} n \overline{T}_{n-1}$ (see the article) shows that for $k < 2n - 1$ the arrangement of the top $n - 1$ disks in the n-disk puzzle after k moves can be obtained from the arrangement of the $(n-1)$-disk puzzle after the same number of moves simply by swapping the labels of pegs B and C. In other words, the codes of these two arrangements—$q_{n-1} \ldots q_1$ and $q'_{n-1} \ldots q'_1$, respectively—are obtained from each other by replacing ones with twos and vice versa (zeros stay in place). Since $1 \equiv -2 \pmod 3$, we can write $q_i \equiv -q'_i \pmod 3$ for all $i \leq n - 1$. The nth disk is moved only at the 2^{n-1}th step, so $q_n = 0$. We also have $k_n = 0$ for $k < 2^{n-1}$, so $k_n = q_n$. Then, by the induction hypothesis, for $i = 1, \ldots, n-2$,

$$\begin{aligned} k_{i+1} - k_i &\equiv (-1)^{n-1+i}(q'_{i+1} - q'_i) \\ &\equiv -(-1)^{n+1}(-q_{i+1} + q_i) \\ &\equiv (-1)^{n+1}(q_{i+1} - q_i) \pmod 3, \end{aligned}$$

and for $i = n - 1$ we have

$$\begin{aligned} k_n - k_{n-1} &= -k_{n-1} = -q'_{n-1} = q_{n-1} \\ &= (-1)^{n+(n-1)}(q_n - q_{n-1}). \end{aligned}$$

If the number of moves is 2^{n-1} or greater, represent it as $2^{n-1} + k$, $0 \le k < 2^{n-1}$. Then its binary notation is $1k_{n-1}...k_1$, and the code of the corresponding position is $1(q_{n-1}+2)(q_{n-2}+2)...(q_1+2)$, where $0k_{n-1}...k_1$ is the binary notation of k and $0q_{n-1}...q_1$ is the position code after k moves (check this using $\overline{T}_{n-1}n\overline{T}_{n-1}$). Now it's easy to see that expression (1) is true from the first part of the proof.

5. Denote by t'_m the minimum number of moves needed to transfer the top m elements from notch A to B without placing element k in notch C. For $m \le k$ the old algorithm is good, and $t'_m = t_m = 2^m - 1$. For $m > k$ the argument in the article must be modified: to move m elements to B we first have to move the top $m-1$ elements to B (not to C as before!), then take element m to C, return the $m-1$ elements to A, take element m to B, and move the $m-1$ elements to B again. So $t'_m = 3t'_{m-1} + 2$ for $m > k$. It follows that

$$\begin{aligned} t'_n + 1 &= 3(t'_{n-1} + 1) \\ &= 3^2(t'_{n-2} + 1) \\ &= \\ &\vdots \\ &= 3^{n-k}(t'_k + 1) \\ &= 3^{n-k} \cdot 2^k, \end{aligned}$$

or

$$t'_n = 3^{n-k} \cdot 2^k - 1.$$

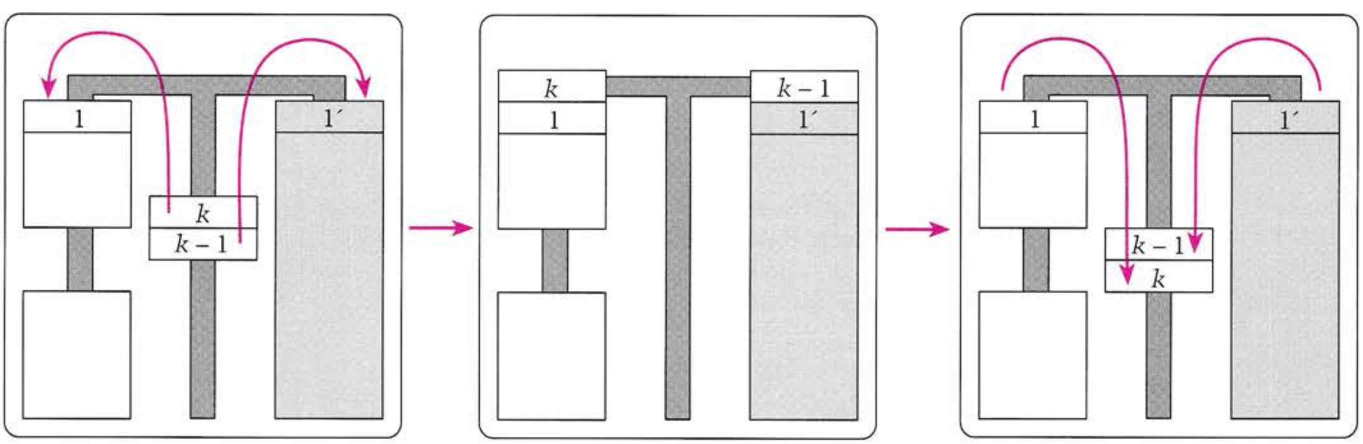
6. We'll use induction over n. The case $n = 1$ is trivial. Suppose the statement is true for any number of disks less than n ($n > 1$) and consider an arbitrary initial arrangement of n disks. Suppose that the nth disk starts on peg A and must be brought (with all the other disks) onto peg B. In the initial position, choose all the disks on A above disk n and all the disks on B. Let m be the number of the largest of them. If there are larger disks on peg C, find the largest of them and add all the disks on it to those chosen before; otherwise, choose all the disks on C. All in all, we'll choose no more than $m \le n-1$ disks, the largest of which (m) is on peg B, and the top unchosen disks on A and C will be larger than disk m. By the induction hypothesis we can collect them on peg C in no more than $2^m - 1$ moves. Now peg B is free and disk n (on A) has no disks on it, so we can move it to B (one move). Since disk $n-1$ is on A or C, we can again apply the hypothesis to all the $n-1$ disks on these pegs and bring them onto B in no more than $2^{n-1} - 1$ moves, completing the solution for this case in no more than $(2^m - 1) + 1 + (2^{n-1} - 1) \le 2^n - 1$ moves.

Now consider the case when the disks are to be stacked on peg A. If disk n initially lies at the bottom of this peg, we can collect all the other $n-1$ disks on it in no more than $2^{n-1} + 2^{n-3} - 1 < 2^n - 1$ moves by the induction hypothesis.

Otherwise, choose the disks on peg A above disk n and all the disks on B and C. There are no more than $n-2$ of them, and they can be brought to one of the pegs B or C in no more than $2^{n-2} - 1$ moves by the hypothesis. The other of these two pegs is thus emptied and we move the nth disk onto it. Now, by the first part of the statement (proved for n disks above), we can collect all n disks on A in no more than $2^n - 1$ moves (because the nth disk is not on A). The total number of moves does not exceed $2^{n-2} - 1 + 1 + 2^n - 1 = 2^n + 2^{n-2} - 1$. A position for which this estimate is attained results from the normally ordered tower on peg A by swapping the two bottom disks n and $n-1$. Check that we can't unswap these disks in fewer than $2^n + 2^{n-2} - 1$ moves.

7. Consider the left piece on the kth level. In its trip to its target location, it must be carried over the $(k-1)$st piece. A simple analysis shows that this exchange can take place (for $k \ge 3$) only in the space formed by the zeroth level and central column (with other pieces in the left and right columns) as illustrated in figure 15. The exchange operation takes four moves and will be denoted by E. Since after this exchange we get the final position of operation D_k, L_k must necessarily begin with this operation. The remaining part of L_k takes the $k-2$ top left pieces to the central column; we denoted it by L_{k-2}.

8. The same argument as in the previous solution shows that S_k must



operation E

Figure 15

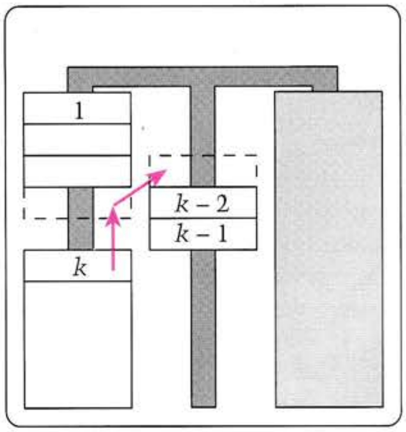
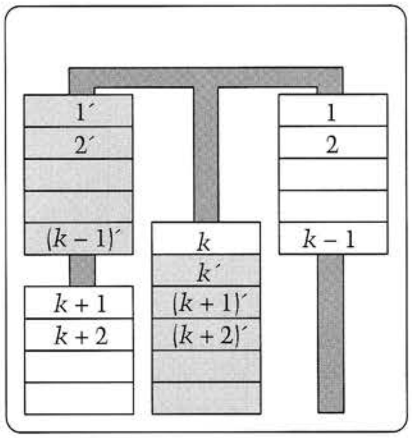


Figure 16



position B_k

Figure 17

begin with D_k. After D_k is performed, we must bring the piece $k-1$ back to the left column, which is done by the inverse of S_{k-1}. So $S_k = D_k S_{k-1}^{-1}$, and $s_k = d_k + s_{k-1}$.

Then, as was explained in the article, to clear the way for piece k, we must remove all the pieces above it. In particular, piece $k-1$ must go to the $(k-1)$st level in the central column. Applying the argument used above for L_k once again, we see that D_k must begin with D_{k-1}. After D_{k-1} is completed (fig. 16), we have to bring piece k to the central column and "push it through" pieces $k-2$ and $k-1$—that is, exchange it with these pieces, one by one. To prepare the first exchange, piece k must be brought (alone) on top of piece $k-2$ (this is done by the operation S_{k-2}); then we exchange them (E), return piece $k-2$ to the left column (S_{k-2}^{-1}), and exchange pieces k and $k-1$ (E). So $D_k = D_{k-1}S_{k-2}ES_{k-2}^{-1}E$, which gives the second equation in the problem, because E takes four moves.

9. The pieces at each level can be exchanged independently from the others. We can move the left piece k into the central column by the operation S_k, then move its right neighbor on top of it by a "mirror" operation, then move the right piece to the left column by S_k^{-1}, and, finally, move the left piece to the right by the "mirror image" of S_k^{-1}. This will require $4s_k$ moves, and $4(s_1 + s_2 + \ldots + s_{10})$ moves for the entire puzzle.

A shorter solution proceeds as follows. First we move the entire right column to the center ($l_{10} + 1$ moves). Then we go through the series of "basic positions" B_k shown in figure 17 in the following way. After the right column is moved, position B_2 is obtained in four moves. To pass from B_k to B_{k+1}, $k \ge 2$, we move piece k to its destination on the right (s_k moves), move piece $k+1$ from the left to the center on top of k' (the piece from the kth level in the right column—this takes sk moves more), exchange pieces $k+1$ and k' (four moves), and move k' to the left (s_k moves). The total number of moves for this transition is $3s_k + 4$. The last step—from B_{10} to the final position—takes $2t_{10}$ moves, and the entire rearrangement takes $l_{10} + 3(t_2 + \ldots + t_9) + 2t_{10} + 37 = 33{,}213$ moves. The order of the exchanges in the top two levels can be improved slightly to save six moves.

Kaleidoscope

1. No, because the velocity of a hot air balloon is equal to that of the wind.

2. The forces are the same. When the helicopter hovers, it exerts a force on the air with a force equal to its weight, and the air merely transfers this force to the ground.

3. The body of a kite divides the incident air flow in such a way that the pressure under the kite is greater than that above it, so lift is created. The kite's tail stabilizes the kite and helps maintain the necessary angle of attack (the inclination of the kite).

4. The flames of the candles tilt toward each other because the pressure in the stream of air is less than that in the surrounding air.

5. In order to equalize the pressure. On a related note, think of how a shock wave from an explosion knocks over the boards of a fence but leaves the fence poles standing.

6. The air carried along by the passing train reduces the pressure, so a force arises that draws a person toward the train.

7. The air pressure at the moving end of the tube is less than that at the (more or less) stationary end. This pressure difference causes the air to flow through the tube, and the air vibrates as it passes along the corrugated walls.

8. The flywheel regulates the rotational frequency of the windmill, preventing abrupt variations due to wind gusts.

9. With proper rotation of the cylinders, the speed of the air will be less at the back surfaces than at the front surfaces of the cylinders. Thus, the air pressure on the cylinders will be greater from the stern than from the prow, resulting in a net force in the forward direction.

10. The hole in the center of a parachute lets part of the incoming air through. This dissipates the vortices that would otherwise form near the edges of the parachute and cause it to rock back and forth.

11. Vortices are generated at the larger opening of the funnel, which form an area of low pressure that draws the candle's flame toward the funnel.

12. Initially the hot cigarette smoke rises rather slowly and forms a laminar (nonturbulent) flow. Then the buoyancy force accelerates it, vortices are generated, and the flow eventually becomes turbulent.

13. Because the water flow is continuous, the volume of water discharged is constant along the entire stream. So as velocity of the falling

water increases, the cross section of the stream becomes thinner.

14. When water flows through a pipe, the initial pressure (of the order of several atmospheres) decreases gradually almost down to atmospheric pressure due to the viscosity of the water. When you block the faucet with your finger, the water almost stops flowing inside the pipe and the pressure near the opening increases drastically. It is this high pressure that imparts a large velocity to the small jet of water.

Microexperiment

The pressure inside the stream is less than the atmospheric pressure, so the air pushes the ball against the stream.

Discriminant

1.(a) (–1, 1); (b) (4, 4, –4). *Hint:* for any fixed z the numbers x and y are the roots of the equation $t^2 - (4 - z)t + (16 + z^2)/2 = 0$, whose discriminant equals $-(z + 4)^2$. One can also eliminate z and factor the resulting quadratic in x and y.

2. *Hint:* The discriminant of the left side as a function of x equals $-8(y + 1)^2$.

3. Answer: $y_{\min} = 7/23$. *Hint:* the unknown value can be found as the smallest number y such that the equation $(3y - 2)x^2 + (11y - 9)x + 12y - 11 = 0$ in x has a solution.

4. Answer: $a \in (-\infty, -3/2) \cup (1 + \sqrt{2}, \infty)$. *Hint:* basically, the problem is solved like problem 8 in the article. But pay attention to the sign of the denominator and the case when it vanishes!

5. Answer: $7\sqrt{3}$. *Hint:* let $t = x + 5y$; then, substituting $t - 5y$ for x, we arrive at the problem of finding the smallest positive t such that a certain quadratic inequality in y has a solution. This can be done in a manner similar to the solution of problem 13 in the article. After that you have to make sure that at least one of the solutions y corresponding to this t is positive together with $x = t - 5y$.

"BELL CURVE!" CONTINUED FROM PAGE 27

"what if" questions. It frees them to think without having to calculate, build without having to describe, plan without having to formalize. It rewards "street smarts" over "book learning."

It is comfortable to believe that something as complex and multifaceted as mental ability can be reduced to a neat set of numbers. It appeals to our national preoccupation with quantifiable statistics while excusing our failure to educate a vast and growing number of our youth. It justifies our creation of filters in the form of the "objective" tests that snuff out hope for the sons and daughters of our poor. But perhaps its most important function has been to insulate us from what might otherwise become the intolerable suspicion that in ceasing to believe that all men are created equal we have placed at risk the very foundation of our democracy.

The other day I asked Luc, who has recently been informed that low SAT scores may have destroyed his chance for higher education, why he liked the Relativity class. He thought for a moment, then looked up at me with a shy smile. "I like it because we are doing things that most college students cannot do."

"How does that make you feel?" I asked.

"It makes me feel intelligent."

The confidence and self-esteem engendered by his success in this class may not prove sufficient for Luc to overcome the formidable obstacles he will face on the way to becoming a productive citizen in a knowledge-based economy.

But it's a start.

Paul Horwitz *is Principal Scientist in the Educational Technologies Department of Bolt Beranek and Newman Inc., Cambridge, Massachusetts.*

Index of Advertisers

Nesting puzzles

Part I: Moving oriental towers

by Vladimir Dubrovsky

MAYBE YOU'VE HEARD about or even seen one of the most popular Russian souvenirs—a toy called the *matryoshka*. It's a family of brightly painted, pear-shaped wooden dolls of decreasing size. Each of them is hollow (except the smallest one), is cut in two around the middle, and contains the next one in order of size (fig. 1). (Nowadays they are often made to caricature historical personalities or political leaders.) The puzzles that we'll talk about in this article are, in a way, similar to this packed (if not action-packed) toy: each of them contains a series of telescoping similar puzzles, and you can't solve it until you solve all the "smaller" puzzles, just like you can't extract the innermost "baby doll" from a matryoshka before you open up, in order, all of its "mother dolls." No wonder these puzzles are often used for demonstrating the power of mathematical induction and recursive equations. In this installment we'll deal with perhaps the most popular of the nesting puzzles and its less familiar versions. This will lead us into the second part (in the next issue), where we'll reveal an unexpected and far-reaching connection between these puzzles and a certain breed of beautiful mathematical monsters already known to *Quantum* readers—the "dragon curves."

The Tower of Hanoi

This puzzle, invented by the French mathematician E. Lucas in 1883, is shown in figure 2. You have to move the cone-shaped tower of disks on the peg A onto peg B using the additional peg C. In so doing you must obey two rules: (1) only one disk can be moved at a time; (2) a disk cannot be placed on a smaller disk. Usually you're asked to find the shortest solution and calculate the number of moves it takes (where a "move" is a single shift of a single disk). The two rules above are often called "Brahma's rules" because of the legend about Brahmins in Benares playing this game with a tower of 64 disks. This story almost invariably accompanies the descriptions of this puzzle in countless books on recreational mathematics (see, for instance, *Mathematical Recreations and Essays* by W. W. Rouse Ball and H. S. M. Coxeter). I'll skip the legend, but I'll repeat the solution, because it's the simplest and clearest example of what should be done with puzzles of this sort. (Of course, it would be better if you try to solve this puzzle on your own first, if you've never done this before. You can use ordinary playing cards with different values in place of disks of varying sizes.)

Denote simply by k a move that carries the kth (in order of size) disk onto the next peg clockwise (from A to B, from B to C, or from

Figure 1

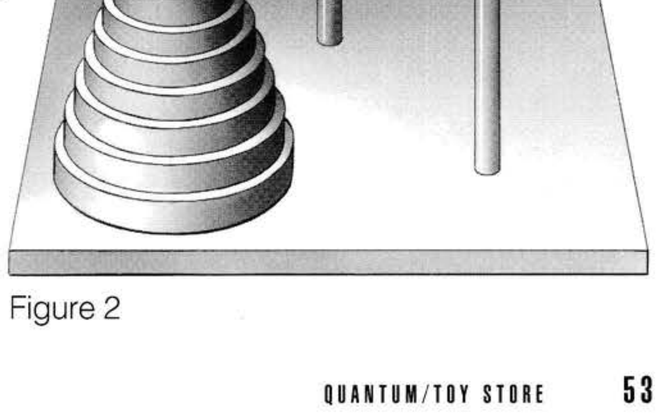


Figure 2

C to A), and by $\overline{k}$ a move of this disk in the opposite direction. Let T_n be the shortest series of moves that takes the tower of n disks to the next peg clockwise. The inverse operation can be obtained by reading T_n from the end and reversing each single move in it (replacing k by $\overline{k}$ and vice versa); it will be denoted by $\overline{T}_n$. We have to find T_n. For small values of n this can be done directly: $T_1 = 1$, $T_2 = \overline{1}2\overline{1}$. After a little thought you'll see that $T_3 = 1\overline{2}131\overline{2}1$. At this point you'll probably guess the general formula. But it's not necessary to make guesses, because it can be derived from the following simple argument.

Start with an n-disk tower on peg A and consider the first time the nth disk is moved. Obviously before this can be done we must remove the top $n - 1$ disks from peg A and collect them on another peg, leaving the third peg empty for the nth disk to be placed there. Since we want the shortest operation, the free peg must be B. So the entire rearrangement begins with moving the $(n - 1)$-disk subtower from A to C (which is done by the process $\overline{T}_{n-1}$); then we transfer the nth disk from A to B (move n) and end by moving the $(n - 1)$-subtower from C to B, on top of disk n (process $\overline{T}_{n-1}$ again). Summarizing, we have the relation

$$T_n = \overline{T}_{n-1}n\overline{T}_{n-1},$$

with $T_1 = 1$. This complies with our formulas for T_1, T_2, T_3 given above. Let's apply it once more:

$$T_4 = \overline{T}_3 4 \overline{T}_3 = \overline{1}2\overline{1}\overline{3}\overline{1}2\overline{1}4\overline{1}2\overline{1}\overline{3}\overline{1}2\overline{1}.$$

We can view this formula, as well as the general formula for T_n, as "grown from left to right" by successively splicing two reversed copies of each previous formula with a move of the corresponding disk inserted at the joint. Alternately, it can be viewed as if filled in from inside (see figure 3) in a process similar to marking a binary ruler by halving each successive space. Either interpretation can be used to calculate the number t_n of moves in T_n.

Problem 1. Use the "splicing" process to show that

$$t_n = 2t_{n-1} + 1$$

and derive that $t_n = 2^n - 1$ from this recursive equation. Use the "binary scale" process to show directly that $t_n = 1 + 2 + 4 + \ldots + 2^{n-1} = 2^n - 1$.

Look again at the binary scales in figure 3. Take the spacing obtained after the nth halving as a unit length. Then the largest scale division will mark the length 2^{n-1} from one end of the segment. The second-largest divisions will mark the lengths 2^{n-2} and $2^{n-1} + 2^{n-2}$—that is, multiples of 2^{n-2} that are not multiples of 2^{n-1}. Then we mark the multiples of 2^{n-3} that are not multiples of 2^{n-2}, and so on. The finest divisions mark odd numbers. In terms of our algorithm for solving the Tower of Hanoi, this means that *if a number k is divisible by 2^{m-1} but not by 2^m* (or, equivalently, its binary representation ends in $m - 1$ zeros), *then at the kth step we move the mth disk.*

Problem 2. How does the direction of this move depend on the size n of the tower?

If you keep count of the moves, you'll always know what disk must be moved next and where. In fact, it suffices to know whether the next move is even or odd, or what disk was moved last. Odd moves (and only odd moves) are made by the smallest disk and always in the same direction (solve problem 2 and you'll see why). Even moves are uniquely determined by Brahma's rules—the smaller disk goes on the bigger one.

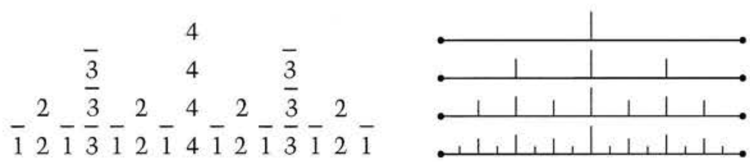


Figure 3

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Publication title: Quantum. (2) Publication No. 008-544. (3) Filing Date: 10/1/95. (4) Issue Frequency: Bi-Monthly. (5) No. of Issues Published Annually: 6. (6) Annual Subscription Price: $39.00. (7) Complete Mailing Address of Known Office of Publication: 175 Fifth Avenue, New York, NY 10010-7858. (8) Complete Mailing Address of Headquarters or General Business Office of Publisher: 175 Fifth Avenue, New York, NY 10010-7858. (9) Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor: Publisher: Bill G. Aldridge, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000, in cooperation with Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010-7858. Editors: Larry D. Kirkpatrick, Mark E. Saul, Albert L. Stasenko, Vladimir Dubrovsky, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. Managing Editor: Timothy Weber, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. (10) Owner: National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201. (11) Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities: None. (12) The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 Months. (13) Publication Name: Quantum. (14) Issue Date for Circulation Data Below: 9/95. (15) Extent and Nature of Circulation: (a.) Total No. Copies (Net Press Run): Average No. Copies Each Issue During Preceding 12 Months, 10,625; Actual No. Copies of Single Issue Published Nearest to Filing Date, 10,800. (b.) Paid and/or Requested Circulation: (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales: Average No. Copies Each Issue During Preceding 12 Months, 1,630; Actual No. Copies of Single Issue Published Nearest to Filing Date, 2,238. (2) Paid or Requested Mail Subscriptions: Average No. Copies Each Issue During Preceding 12 Months, 7,208; Actual No. Copies of Single Issue Published Nearest to Filing Date, 6,828. (c.) Total Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 8,838; Actual No. Copies of Single Issue Published Nearest to Filing Date, 9,066. (d.) Free Distribution by Mail: Average No. Copies Each Issue During Preceding 12 Months, 270; Actual No. Copies of Single Issue Published Nearest to Filing Date, 270. (e.) Free Distribution Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months, 631; Actual No. Copies of Single Issue Published Nearest to Filing Date, 660. (f.) Total Free Distribution: Average No. Copies Each Issue During Preceding 12 Months, 901; Actual No. Copies of Single Issue Published Nearest to Filing Date, 930. (g.) Total Distribution: Average No. Copies Each Issue During Preceding 12 Months, 9,739; Actual No. Copies of Single Issue Published Nearest to Filing Date, 9,996. (h.) Copies Not Distributed: (1) Office Use, Leftovers, Spoiled: Average No. Copies Each Issue During Preceding 12 Months, 284; Actual No. Copies of Single Issue Published Nearest to Filing Date, 804. (2) Return from News Agents: Average No. Copies Each Issue During Preceding 12 Months, 602; Actual No. Copies of Single Issue Published Nearest to Filing Date, 0. (i.) (Sum of 15g, 15h(1), and 15h(2)): Average No. Copies Each Issue During Preceding 12 Months, 10,625; Actual No. Copies of Single Issue Published Nearest to Filing Date, 10,800. Percent Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 90.74; Actual No. Copies of Single Issue Published Nearest to Filing Date, 90.70. (16) This Statement of Ownership will be printed in the January/February 1996 issue of this publication. I certify that all information furnished on this form is true and complete.

Craig Van Dyck

Craig Van Dyck
Senior Vice President

Another simple rule that gives the right order for rearranging the disks was proposed by Misha Fyodorov, a high school student from Chernogolovka, near Moscow.[1] Any move involves two of the three pegs, while the third peg remains "idle." Misha noticed that *the idle peg always moves in the same direction*. For instance, if the first idle peg is B, then the second is C, then A, B, C, ...

Problem 3. Prove that the idle peg always moves in the same direction. For a given n, which is the first idle peg?

Together, these rules enable us to determine the next move, provided we know something about the preceding moves. But what if we've lost count of the moves or forgotten what the last move was? Can we get our bearings directly from the current position of the disks? Problem 4 below provides the answer to this question.

Denote by p_i the label of the peg occupied by the ith disk (so that $p_i = A$, B, or C). Any position of the puzzle is given by the "word" $p_n p_{n-1} \ldots p_1$. (Notice that there are 3^n conceivable "words" of this kind, but moving the tower from one given peg to another in a shortest way, we'll pass through only 2^n of them—the initial position and the results of all the $2^n - 1$ moves.)

Problem 4. Suppose a position $p_n p_{n-1} \ldots p_1$ has emerged after k moves. Find a rule for calculating k from the "code" $p_n p_{n-1} \ldots p_1$, and vice versa. (Hint: it will be helpful to use the binary notation of k.) Show that k has the same parity as the number of letter changes in the sequence of letters $A p_n p_{n-1} \ldots p_1$.

Physically, the Tower of Hanoi can be realized in various shapes. In figure 4 you see a version that was available in Moscow toy stores a few years ago. It differs from the original in design only: numbered square pieces sliding in a frame with three connected slots here play the role of disks of different sizes stacked on three pegs. Another version (fig. 5) is a bit more clever. It was presented by its author, Youshiyuki Kotani of Japan, at the 14th International Puzzle Party (Seattle, 1994—see the Toy Store instalment in the July/August 1995 issue of *Quantum*). Here the disks are replaced by U-shaped elements of different sizes stacked in three notches on a special rack. The order of the sizes is reversed here: the smallest "U" corresponds to the largest disk in the Tower, and so on. Referring to the shape and size of the elements, the author changed the small round "o" in "Hanoi" to a big U and called his toy the "HanUi Puzzle." This design automatically forces you to follow the second Brahma rule, because it's impossible to put a smaller inverted U on a bigger one. Another interesting feature of this version is the small holes in the U-elements. You can insert a thin stick in one of these holes and thus prevent this element from being placed in the notch C. This is the same as if you were not allowed to put a certain disk of the Hanoi Tower on peg C.[2] Of course, this additional restriction changes the solution.

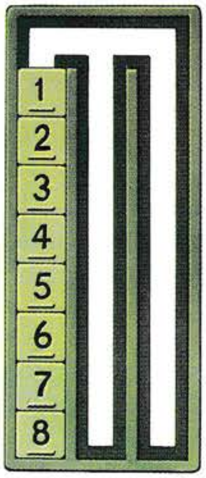


Figure 4

Problem 5. Prove that if the stick is inserted in the kth U-element and the number of elements is n, the shortest solution of the puzzle consists of $2^k \cdot 3^{n-k} - 1$ moves.

Another way to modify the puzzle is to change the initial arrangement of disks. It turns out that even if they are initially put on all three pegs in any order, the problem remains solvable.

Problem 6. Prove that the number of moves needed to collect n disks on a given peg following "Brahma's rules" is no greater than 2^{n-1}, no matter which peg receives the first disk and no matter what order they are placed in initially. The only exception is the case when the largest (nth) disk starts from the peg where the disks have to be stacked in the end and has at least one disk underneath. In this case, no more than $2^n + 2^{n-2} - 1$ moves are needed, and this number cannot be decreased (for $n > 1$).

Panex

In the early 1980s the Japanese Magic Company manufactured a puzzle that resembles the Tower of Hanoi (or rather, its Russian version in figure 4) and can be analyzed in a similar way, though is much more

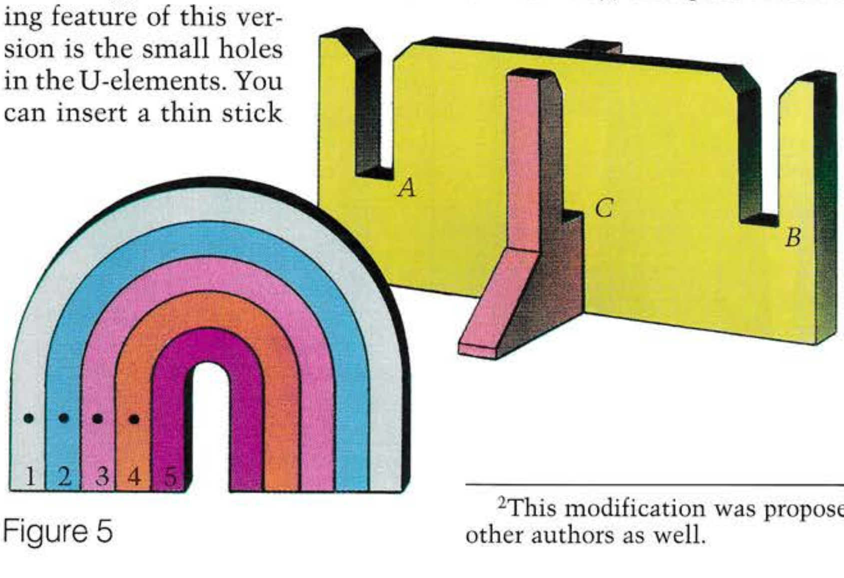


Figure 5

[1]It was described in the article about the Tower of Hanoi published in *Kvant*, the Russian sister magazine of *Quantum* (No. 11, 1991).

[2]This modification was proposed by other authors as well.

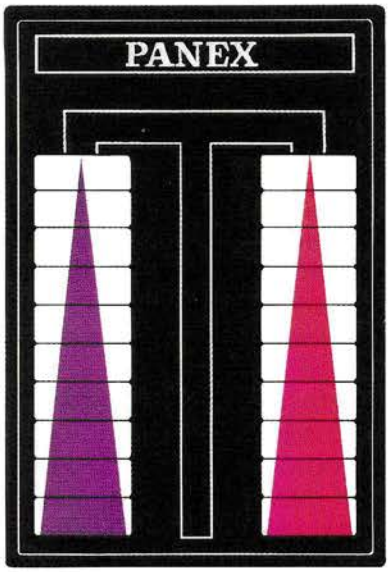


Figure 6

difficult to solve. It bears the mysterious name Panex (I have no idea what this could mean in Japanese). Figure 6 shows what it looks like from the outside, and figure 7 explains its probable hidden construction (compare with figure 4!). Your goal is to swap the left and right columns of pieces (preserving their order in the columns, of course). The central groove is used as a "sorting track." There are no additional rules as to how the pieces can be moved, because the construction itself is rather restrictive: it's impossible to push a piece below its initial level, even if it has no pieces underneath. For instance, either top piece can take only six positions: in the horizontal connecting groove above any of the three columns (we can consider this as three possible positions and call them the "zeroth level") and on its initial level (the first level) in any of the three columns. So these two pieces, unless they are both in the same column, simply block two of the three columns! Such a narrow maneuvering space makes the puzzle look almost unsolvable at first sight—and this is what a really good puzzle should look like. Of course, a solution exists, although it may be that its actual completion requires more time than you're willing to spare. In their book *Puzzles Old And New*, J. Slocum and J. Botermans mention that mathematicians at Bell Laboratories estimated the minimum number N of moves needed to swap the columns in Panex as $27{,}564 \le N \le 31{,}537$. (My own solution is about 1,700 moves longer.) Besides, once you come up with an algorithm and have checked it for five or six levels of the puzzle, you'll most probably say "and so on" to yourself and put it aside. This is a negative feature peculiar to all nesting puzzles.

Whereas the solution for k levels of the Hanoi Tower reduces to the $(k-1)$-level solution almost by itself, similar recursive relations for Panex are much less obvious and more complicated. Not only that—since it makes no difference for the pieces on a certain level what color the pieces above them are, the exchange at this level can't be linked directly to the higher-level exchanges. Recursions show up in certain auxiliary, but indispensable, operations that you'll inevitably have to master in the course of solving the puzzle. Here I'll only introduce these operations and write out recursive relations between them, leaving the rest of the work to interested readers.

Notice first that each piece must leave its column at a certain moment to stay in the other two columns at least for a few moves. Then it may come back, only to leave its "home" column again later. It may even repeat such shuttle trips a number of times. But if you consider the first of such moments for the two pieces in, say, the $(k+1)$st level (counting from level 0), you'll see that for each of them there are only two possible arrangements of the other pieces at this moment. For the left piece, these arrangements, *a* and *b*, are shown in figure 8 (the numbers on the pieces indicate their initial levels). Let's concentrate on arrangement *a*—it leads to a slightly shorter algorithm. It's more or less clear (and will be taken for granted) that the shortest operation L_k that creates this arrangement leaves the right column intact.

Problem 7. Show that for $k \ge 3$ the operation L_k can be represented as $D_k L_{k-2}$, where D_k is the shortest operation that moves the two pieces k and $k-1$ as illustrated in figure 9.

From this problem we get the following equation for the numbers of moves l_k and d_k in L_k and D_k:

$$l_k = k_k + l_{k-2},\ k \ge 3. \qquad (1)$$

It's easy to check directly that

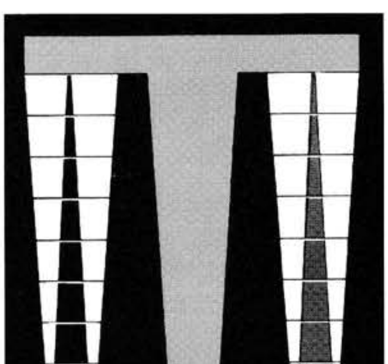


Figure 7

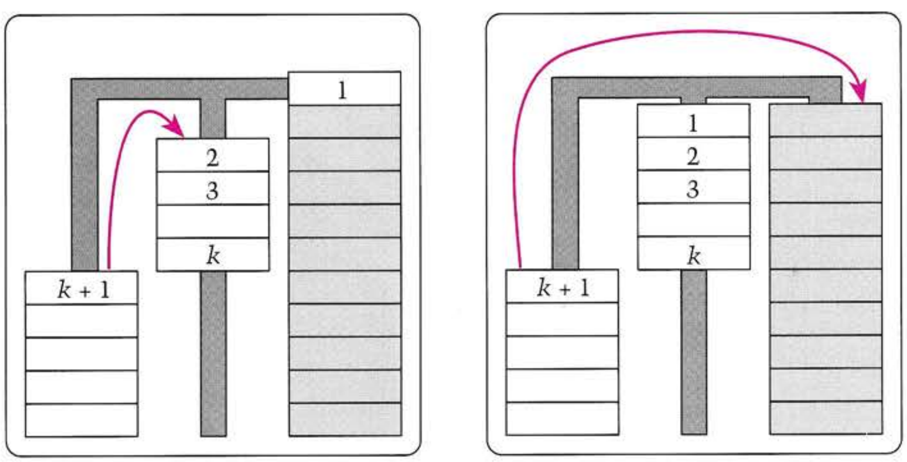
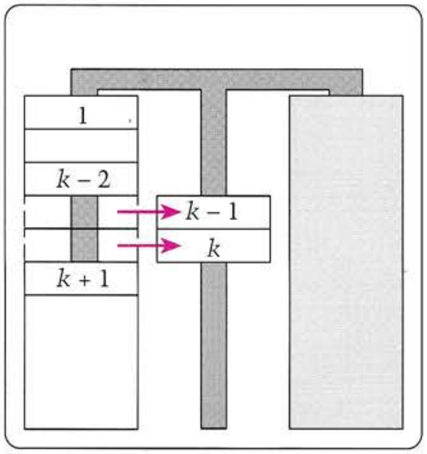


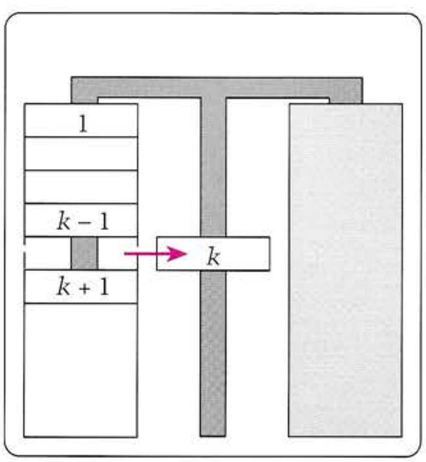
Figure 8



operation D_k

Figure 9

$l_1 = 1$, $l_2 = 2$, $d_2 = 3$ (d_1 is, naturally, undefined). Additional trial-and-error work shows that $d_3 = 8$, so $l_3 = d_3 + l_1 = 9$. Further values of d_k (and l_k) can be found by using the recursive relations between D_k and another auxiliary operation, S_k, which moves a single piece in the kth level from the left to the central column starting from the initial arrangement (see figure 10). It's important to notice that this operation will work even if the kth level in the central column is initially occupied (say, by the corresponding right piece). In this case it will bring the piece from the left kth space to the central $(k - 1)$st space. But it won't work if the latter space is occupied, because this wouldn't leave enough room to clear the way for the transferred piece. The same



operation S_k

Figure 10

remark is true for the operation d_k.

Problem 8. Let s_k be the number of moves in S_k. Show that

$$\begin{gathered} s_k = d_k + s_{k-1}, \\ d_{k+1} = d_k + 2s_{k-1} + 8, \qquad (2) \\ k \geq 4. \end{gathered}$$

The values of s_k for small k can be found directly: obviously, $s_1 = 1$, $s_2 = 3$. For $k = 3$, a slight change in the order of moves causes the value $d_3 + s_2 = 11$ given by our equation to decrease by 1, so that $s_3 = 10$. Similarly, $d_4 = 21$, which is one less than the value given by our formula. Now, using equations (2) we can calculate step by step all the values of s_k and d_k for $4 \leq k \leq 10$, and then, using equation (1), the values of l_k. My calculation gave the following values for $k = 10$: $s_{10} = 6{,}891$, $d_{10} = 4{,}039$, $l_{10} = 4{,}874$.

Now the operations we introduced can be combined into a complete solution of the Panex puzzle in its original formulation.

Problem 9. Write out a complete solution to the Panex puzzle. ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 49

Readers write . . .

And call . . .

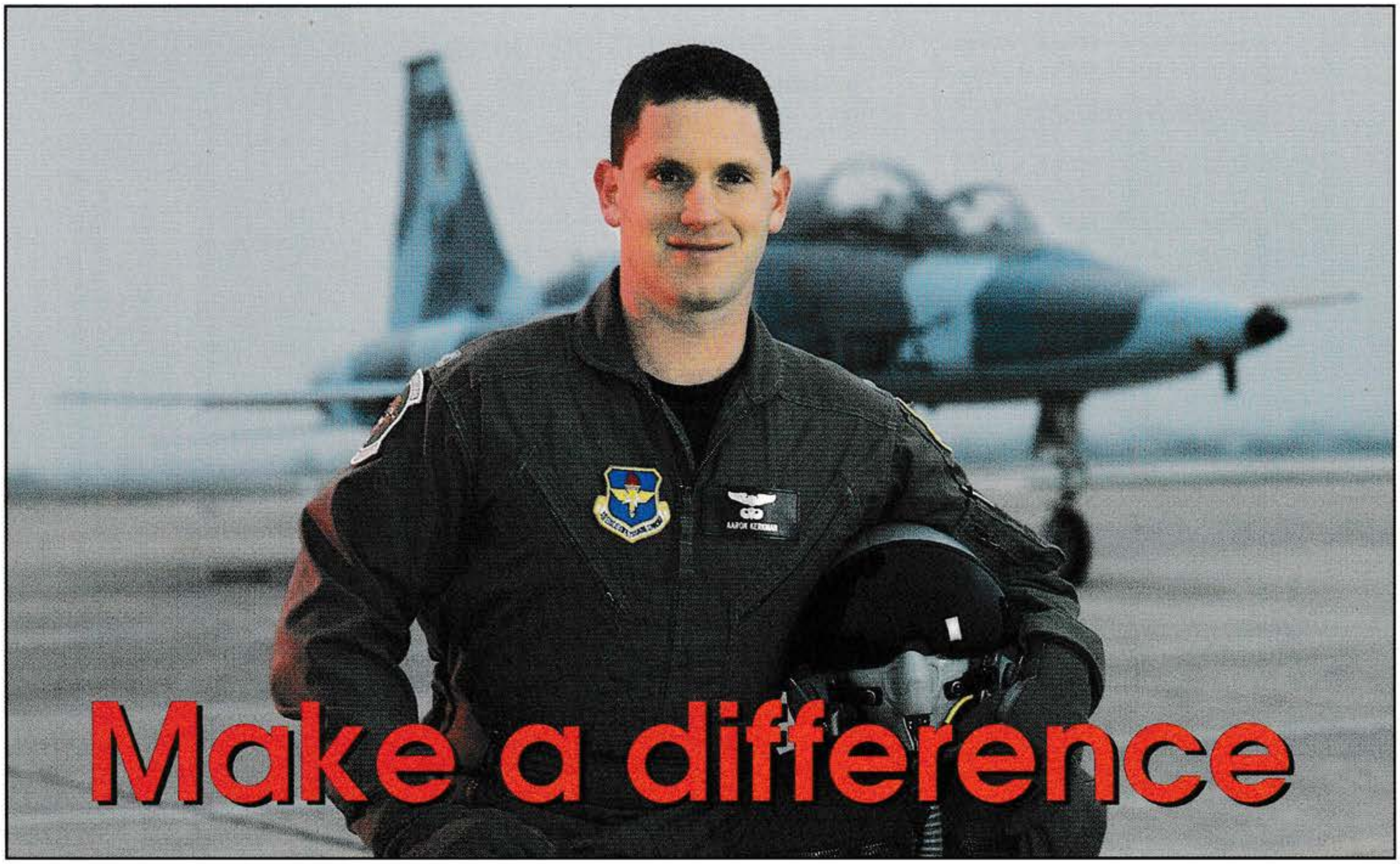
Goran Grimvall, a professor at the Royal Institute of Technology in Stockholm, Sweden, and a specialist in the theory of materials, wrote to us via e-mail:

In the September/October [1995] issue, there is a nice article on elasticity (p. 32) that contains a common error. A red-hot wire (question 14) is easy to tear off because thermally activated motion of dislocations is easier at higher temperatures. The thermal expansion lowers the forces between the atoms typically by only 10%–20%, and that is seen in the elastic parameters. One must distinguish between elastic and plastic deformation. Lead is soft not because its interatomic forces are particularly small, but because room temperature is already "hot" for lead.

Michel Dekking contacted us from the Netherlands, also by e-mail:

It was a pleasure for me to read the paper "Dragon Curves" by Vasilyev and Gutenmacher in the September/October issue of *Quantum*, which originally appeared in 1970 in Russian. Since then a lot of insight into dragon curves has been obtained. I would like to mention one example. In exercise 11(c), a space-filling and quasi-semifolding property is mentioned with the phrase: "This is a difficult theorem proved first by Donald Knuth." After this the authors indicate that representations of complex numbers are needed to prove the result. Actually there is a simple geometrical proof, which is mentioned in the paper "Folds!" by M. Mendes-France, A. v. d. Poorten, and myself (*Mathematical Intelligencer*, No. 4 (1982), 130–38 ff.).

Howard Swift called to inform us that, contrary to the attribution in the July/August 1995 Toy Store, he is not the creator of the puzzle shown in figure 5. That honor goes to Christiaan Vreeling and Enneke Treep of the Netherlands. Swift makes and sells puzzles, and he encountered No-Knot at the 13th International Puzzle Party, held in the summer of 1993 outside Amsterdam (see the report in the July/August 1994 issue of *Quantum*).


Make a difference