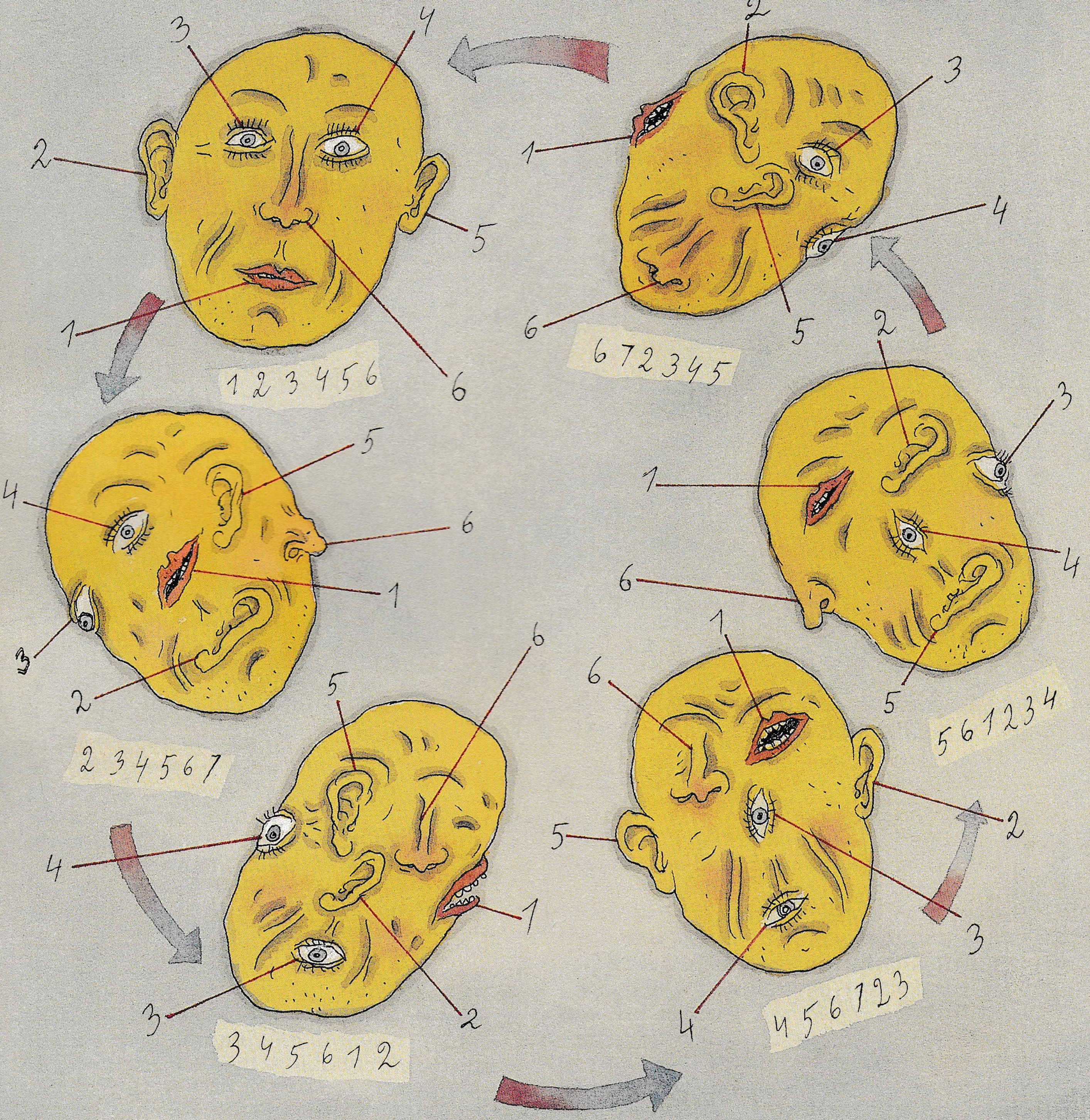
QUANTUM

NOVEMBER/DECEMBER 1991

$5.00



Tishkov

The student magazine of math and science

SPRINGER INTERNATIONAL

GALLERY Q

National Gallery of Art, Washington (Collection of Cornelius Van S. Roosevelt)

Reptiles (1943) by M. C. Escher

MAURITS CORNELIS ESCHER (1898–1972) is indeed the "mathematicians' favorite" and a natural to participate in this issue devoted to group theory.

"Reptiles" is one of a number of works Escher characterized as "an interplay between the stiff, crystallised two-dimensional figures of a regular pattern and the individual freedom of three-dimensional creatures capable of moving about in space without hindrance." He goes on in a curiously sociological vein: "On the one hand, the members of planes of collectivity come to life in space; on the other, the free individuals sink back and lose themselves in the community."

Here's how Escher described "Reptiles":

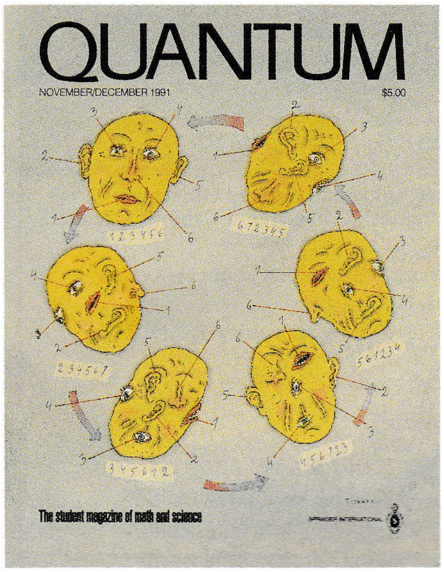
> The life cycle of a little alligator. Amid all kinds of objects, a drawing book lies open, and the drawing on view is a mosaic of reptilian figures in three contrasting shades. Evidently one of them has tired of lying flat and rigid amongst his fellows, so he puts one plastic-looking leg over the edge of the book, wrenches himself free and launches out into real life. He climbs up the back of a book on zoology and works his laborious way up the slippery slope of a setsquare to the highest point of his existence. Then after a quick snort, tired but fulfilled, he goes downhill again, via an ashtray, to the level surface, to that flat drawing paper, and meekly rejoins his erstwhile friends, taking up once more his function as an element of surface-division. (*The Graphic Work of M. C. Escher*, New York: Ballantine Books, 1971, p. 12)

In case you're wondering: the "little book of Job," as Escher puts it, contains Belgian cigarette papers.

Turn to the Kaleidoscope for a fascinating look at Escher through the prism of group theory.

QUANTUM

NOVEMBER/DECEMBER 1991 VOLUME 2, NUMBER 2



Cover art by Leonid Tishkov

The unfortunate gentleman on our cover is having his face rearranged by repeated application of a cyclic permutation of his eyes, ears, nose, and mouth. The fact that these six features retain their shapes, even when appearing in very odd positions, lends a certain fearful symmetry to the resulting series of portraits. Although this particular sort of symmetry may not improve our friend's appearance, various forms of symmetry are often aesthetically pleasing. The study of symmetry has led to one of the most beautiful of mathematical theories: the theory of groups.

The mathematical content of this issue is devoted entirely to different aspects of group theory. You may not have encountered it in your high school curriculum, so we hope *Quantum* can help you get a grip on groups.

FEATURES

DEPARTMENTS

QUANTUM

THE STUDENT MAGAZINE OF MATH AND SCIENCE

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

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York Inc. Volume 2 (6 issues) will be published in 1991–1992. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by the Academy of Sciences of the USSR and the Academy of Pedagogical Sciences of the USSR, as well as original material in English. **Editorial offices:** NSTA, 3140 North Washington Boulevard, 2nd Floor, Arlington, VA 22201. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. Production Editor: Glenn M. Edwards

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (202) 328-5800; (New York) Roger Jänecke (212) 460-1663; and G. Probst, Springer-Verlag, Heidelberger Platz 3, 1000 Berlin 33, FRG, telephone (0) 30-82 07-1, fax 30-82 07-1.

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

Subscription Information: North America

Student rate: $14; Personal rate (nonstudent): $18; Institutional rate: $28. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus, NJ 07096; or call 1-800-SPRINGER (in New York, call [201] 348-4033).

All Countries Outside North America

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

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

1991
U.S. TEAM—
WINNERS OF
GOLD AND
SILVER
MEDALS

XXIII International Physics Olympiad
Helsinki, Finland
July 2–13, 1992

The American Association of Physics Teachers (AAPT) is looking for participants to represent the United States at the 1992 International Physics Olympiad. The goals of the Olympiad are to encourage excellence in physics education and to reward outstanding physics students. The Olympiad is a 10-day international competition among pre-university students from more than 32 nations. At the International Physics Olympiad, the competitors are asked to solve challenging theoretical and experimental physics problems. The 20 members of the U.S. Physics Team are selected through two competitive examinations. Team members are invited to attend a nine-day training camp held in late May or early June during which team members will refine their problem-solving and laboratory skills. The top five members will represent the U.S. Physics Team at the competition in Finland. Team members' expenses are paid by the contributors to the U.S. Physics Team.

Teachers who have students interested in participating in the 1992 Olympiad are encouraged to contact:

Bernard V. Khoury
U.S. Physics Team
AAPT
5112 Berwyn Road
College Park, MD 20740
(301)345-4200

Deadline: January 13, 1992.

The 1991 U.S. Physics Team, winners of a gold and silver medal, with Walter Massey, Director of the National Science Foundation.

Contributors to the 1991 United States Physics Team

Acoustical Society of America
Addison-Wesley Publishing Company
Allyn & Bacon, Inc.
American Association of Physicists in Medicine
American Association of Physics Teachers
American Crystallographic Association
American Institute of Physics
American Physical Society
American Vacuum Society
AT&T
Beckman Instruments
Bellcore
BP Research
COMSAT Laboratories
Ford Motor Company
General Electric Co.
GTE Laboratories, Inc.
Hughes Aircraft Company
IBM
Janis Research Company, Inc.
Lockheed Corporation
McGraw-Hill Book Company
W.W. Norton and Company
Optical Society of America
Phillips Petroleum Company
Prentice-Hall, Inc.
Princeton University Press
Saunders College Publishing
Schlumberger-Doll Research
The Society of Rheology
University of Maryland
Westinghouse Foundation
John Wiley & Sons, Inc.
Worth Publishing, Inc.

Reaching back

Lessons from the Black Hills

I HAVE HEARD OF A NATIVE American who leads hiking expeditions for young people in the Black Hills of South Dakota. Before stepping out into the exhilarating but rough terrain, the hikers must agree to follow three simple rules: (1) *No whining* (just say what you need and we'll try to find a solution); (2) *Help the person behind you* (we're all in this together); (3) *Have fun* (fill your mind with positive thoughts).

Now, I don't suppose any *Quantum* readers are whiners, and I'm sure you have fun reading this beautiful but rocky magazine. In fact, I'm sure you enjoy learning in general. So I'd like to focus on the second rule: Help the person behind you.

No doubt you've heard that the United States faces a major crisis in science education. Maybe you're tired of hearing it. After all, you're doing just fine. You're headed to a good university and possibly a satisfying career in science or technology. But what will your working environment be like in 10 or 20 years? What will society be like then? Will there be anyone outside your field who can understand what you're engaged in?

Not many young people in this country are like *Quantum* readers. Only about 20% of high school graduates in the US study physics and the math required for physics. Since only about 70% of students entering 9th grade eventually graduate, only about 14% of Americans ever learn physics and the related mathematics. Those of you who also study a second year of physics or advanced math are members of an even smaller population.

The sad truth is that far more of your friends could learn physics and mathematics than do so. In fact, almost everyone can learn these subjects at rather high levels, but most don't because they're led early on to believe that they can't handle them. They're made to feel that it takes some special aptitude or talent. A great body of evidence now shows that this is simply not true. There are, among you, individuals with extraordinary talent in math and physics, but those persons are rare indeed—only about one in 10,000 students. The rest of you study hard and take pleasure in learning math and physics. That's what keeps you going.

Much of what I'm saying was stimulated by a recent report of the Carnegie Commission on Science, Technology, and Government. The report cites evidence to discredit four fallacies: (1) "math and science ability is innate, and that many or most young people cannot learn mathematics and science," (2) "women and minorities will not be in the front ranks of technical achievement," (3) "students disadvantaged by poverty, race, or language in urban and rural schools are not needed to support the technical base of this country," and (4) "these subjects are only important for the immediately college bound." The point, of course, is that our nation needs all of its talent, and we cannot afford to waste talent using these fallacies as excuses.

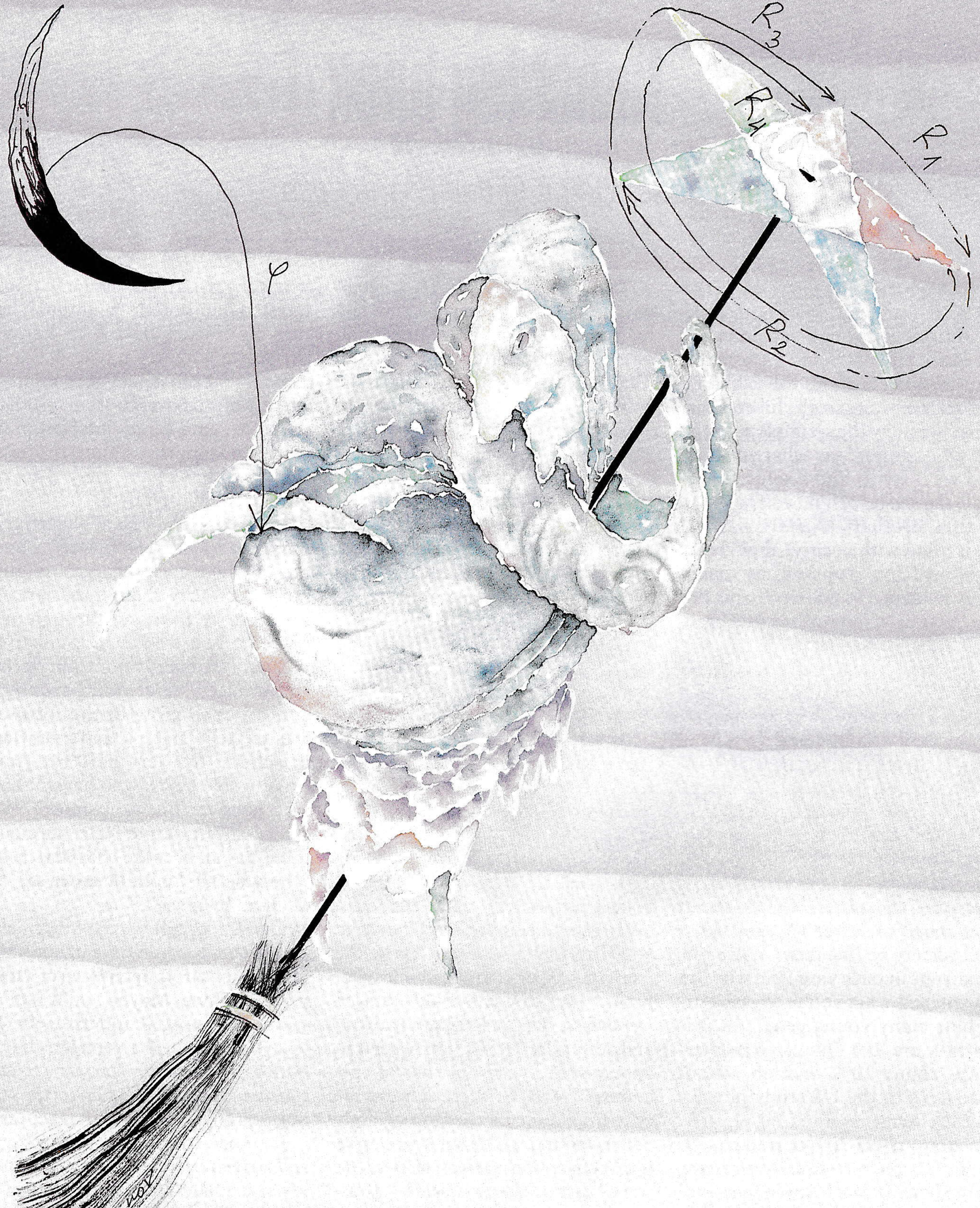
You know a lot of your friends who don't do well in school, and you know that many of them are pretty smart. You also know that, unlike you, they're unmotivated and uninterested. But I think "all human beings by nature desire to know" (Aristotle, slightly updated). Help them if you can. Share your interest and enthusiasm with them so that they, too, gain some satisfaction in learning something they thought they couldn't. Then they'll learn quite well on their own.

Many who don't learn science and math are from disadvantaged groups; most of them are categorized as "minorities." Many sharp minds are allowed to go to waste through our neglect. You who read *Quantum* regularly are mainly from among the "advantaged," and your advantage continues to grow in the form of awards, scholarships, and attention in school. It's very hard for your equally able peers, who haven't had your good fortune, to compete. We can't help but admire their enormous drive. But all of us are obliged to help others, because none of us can say, "I did it all by myself."

One important way you can help is to set aside some time in your life to teach (or tutor). Spend two or three years teaching. Maybe you'll like it and keep doing it indefinitely. If you've ever helped a younger sibling learn to ride a bike or conquer a tough equation, you have an inkling of the satisfaction and enjoyment of teaching.

It's good to set your sights high and strive to achieve your goals. But it's also good to look back once in a while and extend a hand to your friends who started climbing after you or who just need a little help and encouragement.

—Bill G. Aldridge


R_3
R_4
R_1
R_2
φ
S. Ivanov

Marching orders

Atten-SHUN! The topic for today is: Finite Groups

by Alexey Sosinsky

THE NOTION OF A "GROUP," viewed only 30 years ago as the epitome of sophistication, is today one of the mathematical concepts most widely used in physics, chemistry, biochemistry, and mathematics itself. In 1956 James R. Newman, editor of *The World of Mathematics*, described group theory as "the supreme example of the art of mathematical abstraction." Actually, though, groups—especially finite groups—are everyday things, on a par, say, with counting numbers (integers); they're more fundamental and much simpler than lots of things you've been taught in school, like plane geometry or infinite decimal fractions (real numbers).

Some illustrations: groups of actions

A family of actions that can be performed one after the other is said to be a *group* if for each action in this family the inverse action (which "undoes" the given one) also belongs to the family and the result of successively performing two actions from the family is also an action from the family.

As our first illustration, consider the actions performed by a soldier standing at attention on the parade grounds (see figure 1). These four actions constitute a group:

$$R(\blacksquare) = \{S, R, L, F; \circ\}.$$

The result of successively performing the commands R and F ("**R**ight face!"

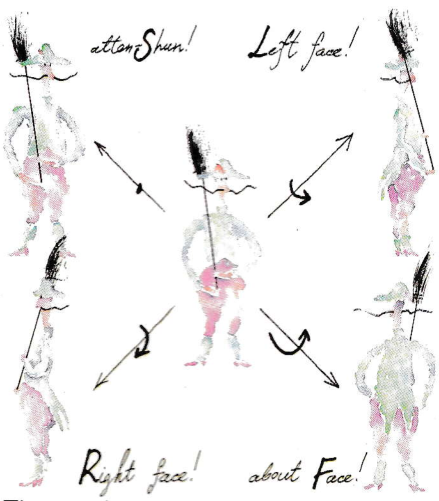


Figure 1

and "about **F**ace!") is the same as doing L ("**L**eft face!"). This can be written in the form of an equation:

$$F \circ R = L$$

Similarly we have

$$R \circ R = L \circ L = F,$$
$$L \circ R = R \circ L = F \circ F = S.$$

Other relations in our group can be read from its *multiplication table*, shown in figure 2. Notice that our group contains the "doing nothing action" S, expressed in military jargon by the command "atten-**S**hun!" and performed by "looking sharp and not moving." Such an action is contained in any group, since it can be obtained by performing an arbitrary action and then its inverse. In our

$\circ$	S	R	L	F
S	S	R	L	F
R	R	F	S	L
L	L	S	F	R
F	F	L	R	S

Figure 2

case the actions R and L are inverse to each other, the action F is inverse to itself.

Now let's consider another group, also having to do with turning—namely, the group of rotations $R(\bigstar)$ of a five-pointed star about its center (see figure 3). Let's denote "doing nothing" (in this case, rotation by 0°)

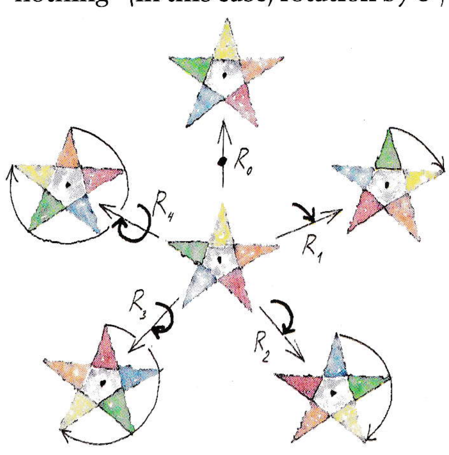


Figure 3

$\circ$	R_0	R_1	R_2	R_3	R_4
R_0	R_0	R_1	R_2	R_3	R_4
R_1	R_1	R_2	R_3	R_4	R_0
R_2	R_2	R_3	R_4	R_0	R_1
R_3	R_3	R_4	R_0	R_1	R_2
R_4	R_4	R_0	R_1	R_2	R_3

Figure 4

by R_0, and the other rotations (by 72°, 144°, 216°, 288°) by R_1, R_2, R_3, R_4, respectively. Here we have

$$R_2 \circ R_1 = R_1 \circ R_2 = R_3,$$
$$R_3 \circ R_3 = R_1,$$
$$R_1 \circ R_4 = R_0.$$

The last relation means R_1 and R_4 are inverse to each other. Other relations may be read from the multiplication table (fig. 4). It's easy to see that

$$\mathrm{R}(\bigstar) = \{R_0, R_1, R_2, R_3, R_4; \circ\}$$

is a group.

As our last illustration, consider the actions related to putting on a sock (fig. 5):

N = "do **N**othing" (leave the sock as is),

S = "**S**witch feet" (take the sock off and put it on the other foot),

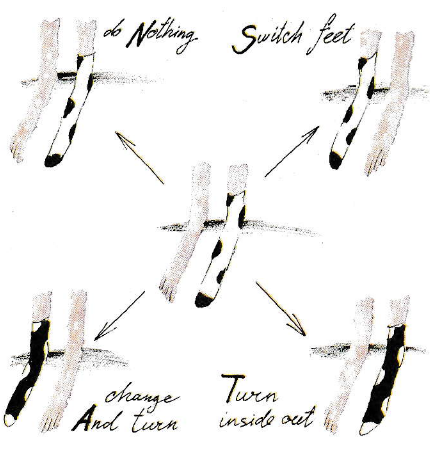


Figure 5

T = "**T**urn inside out" (take the sock off, turn it inside out, and put it on the same foot),

A = "turn inside out **A**nd switch feet" (take the sock off, turn it inside out, and put it on the other foot).

Here the "doing nothing" action is N; each of the actions is inverse to itself:

$$N \circ N = S \circ S = T \circ T = A \circ A = N;$$

and other relations (for example, the obvious $S \circ T = T \circ S = A$) can be read from the multiplication table (fig. 6).

$\circ$	N	S	T	A
N	N	S	T	A
S	S	N	A	T
T	T	A	N	S
A	A	T	S	N

Figure 6

Here again we have a group

$$S = \{N, S, T, A; \circ\},$$

which, like $R(\blacksquare)$, consists of four actions. However, the groups $R(\blacksquare)$ and S are different in principle: their multiplication tables differ not only in the notation used for the actions but in their structure. Thus, the main diagonal of the multiplication table for S contains only one action N (repeated four times), while different actions appear on the main diagonal for $R(\blacksquare)$.

I suspect that the serious reader is starting to get annoyed at all this: "What's going on here? Stupid soldiers doing their drills, ridiculous actions with socks—that's not mathematics, that's not science!"

To prove to such a reader how scientific all of this actually is, let me just give the official, mathematical names of the three groups that we've looked at so far. The first one (about the soldier) is known to mathematicians as the *cyclic group of order 4* (or the *group of residues modulo 4*), and the last (the one about the sock) is the *Klein group*, also known as the *symmetry group of the rectangle*. As to the group based on the star, it's an instance of a *simple finite group*, a type of group about which thousands of research articles have been written, culminating in the discovery of *sporadic groups*, such as the so-called *Big Monster* and *Baby Monster*.

For the more pragmatic reader, who is not impressed by fancy scientific-sounding terms, let me say only that groups are used with great success in such disparate branches of human activity as code theory (for instance, by the CIA and KGB), quantum physics, algebra, . . . and puzzles (see Y. P. Solovyov's article and the Toy Store in this issue).

Symmetry groups of geometric figures

To each geometric figure F we can assign a certain group, called the *symmetry group of the figure F*, denoted by $S(F)$; by definition, $S(F)$ consists of all the motions that take F onto itself. For example, the symmetry group of the square $S(\blacksquare)$ consists of eight actions: four rotations (about the square's center by 0°, 90°, 180°, 270°) and four reflections (in the two diagonals and in the two "midlines" of the square).

The symmetry group of the equilateral triangle $S(\blacktriangle)$ has 6 elements; that of the rectangle $S(\blacksquare)$ has 4. This last group (with its multiplication table) is shown in figure 7.

Take a good look at the multiplication table of the group $S(\blacksquare)$. Doesn't it have a *déjà vu* air about it? Sure enough, it's just like the multiplication table for the "sock group" S except that its actions are denoted by other letters. If we rename the actions constituting S as follows—

$$N \to R_0, S \to R_1, T \to S_1, A \to S_2,$$

—the table for S becomes identical to the table for $S(\blacksquare)$.

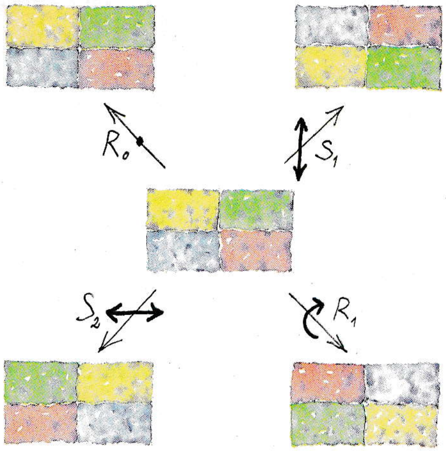


Figure 7

∘	R_0	R_1	S_1	S_2
R_0	R_0	R_1	S_1	S_2
R_1	R_1	R_0	S_2	S_1
S_1	S_1	S_2	R_0	R_1
S_2	S_2	S_1	R_1	R_0

Groups that have identical multiplication tables (after an appropriate renaming of their actions) are called *isomorphic*. We've just shown that the groups S and S(▮) are isomorphic (in honor of the great German mathematician Felix Klein they're often denoted by the letter K), while we previously observed that K is not isomorphic to R(■), the soldier's group.

You've probably guessed by now why we use the notation R(■) for the latter group—it's isomorphic to the *rotation group of the square* (with respect to its center) by the angles $2k\pi/4$, $k = 0, 1, 2, 3$. (Note: the author didn't intend to snub the military by implying that the soldier is a square.) This group R(■) is a particular case (for $n = 4$) of the *rotation group of the regular n-gon* (with respect to its center, by the angles $2k\pi/n$, $k = 0, 1, \ldots, n-1$), also known as the *cyclic group of order n* and usually denoted by $\mathbf{Z}_n$.

In algebra, groups are ordinarily studied "to the point of isomorphism"—that is, algebraists don't distinguish isomorphic groups: they don't care what the groups and their elements are called and how they are denoted; they're only concerned with the structure of the groups' multiplication tables.

Permutation groups and their subgroups

Consider a finite set of distinct objects—say, five. Denote these objects by numbers and the whole set by

$$N_5 = \{1, 2, 3, 4, 5\}.$$

A *permutation* $i \in S_5$ of these objects is any one-to-one map $i: N_5 \to N_5$; or, to put it more simply, a renumbering of these objects. The new number $i(k)$ of the kth object will be denoted by i_k. The permutation i is best displayed by writing it as an array:

$$i = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ i_1 & i_2 & i_3 & i_4 & i_5 \end{pmatrix}.$$

This notation is convenient for finding the product of two permutations $k = i \circ j$ (the result of successively performing the two permutations j and i). For example, if

$$i = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix},$$

$$j = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 4 & 1 \end{pmatrix},$$

then $k(3) = (i \circ j)(3) = i(j(3)) = i(5) = 2$, so that

$$i \circ j = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 1 & 4 \end{pmatrix}.$$

(Notice that when we write $k = i \circ j$, we mean that j is performed before i, and this may be important: in our case $i \circ j \neq j \circ i$, as you can readily check.) It's also very easy to find the permutation inverse to a given one (by reading "backwards"—from the second row up to the first); for example,

$$i^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 2 & 1 & 3 \end{pmatrix}.$$

The set of all permutations S_5 of five objects, as can be verified without much trouble, is a group consisting of $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$ elements, called the *permutation group of degree 5*. The *permutation group of degree n* for any natural number n is defined in a similar way (take n objects instead of five).

Permutation groups are particularly interesting in that they contain numerous *subgroups* (that is, parts that are groups themselves). Permutation groups contain subgroups isomorphic to all the groups we've considered up to now. The interested reader can verify this by studying figure 8, which may also yield other significant observations.

In particular, you may see that certain numerical regularities appear in this figure. For example, if we call the number of elements in a group its *order*, and define the *order of an element g* as the least positive integer k for which $g^k = e$, where e is the identical permutation

$$e = \begin{pmatrix} 1 & 2 & \ldots & n \\ 1 & 2 & \ldots & n \end{pmatrix}$$

and g^k is short for $g \circ g \circ \cdots \circ g$ (k times), then the following theorem can be stated.

Lagrange's Theorem. *The order of any subgroup, as well as the order of any element of a group, divides the order of the group.*

I'll omit the proof (which isn't too difficult).

Relationships between groups: homomorphisms

Groups are studied not only in and of themselves but also as they relate to each other. We say that a map $\gamma: G \to H$ is a *homomorphism* of a group G into the group H if for all elements $g, g' \in G$ we have

The group $\mathbf{Z}_{12} = \left\{ R^{\frac{2\pi}{12}} = z, z^2, z^3, \ldots, z^{11}, z^{12} = e \right\}$

1 element of order 2: $\{z^6, e\} \cong \mathbf{Z}_2$

2 elements of order 3: $\{z^4, z^8, e\} \cong \mathbf{Z}_3$

2 elements of order 4: $\{z^3, z^6, z^9, e\} \cong \mathbf{Z}_4$

2 elements of order 6: $\{z^2, z^4, z^6, z^8, z^{10}, e\} \cong \mathbf{Z}_6$

4 elements of order 12: z, z^5, z^7, z^{11}

The group $S_4 = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ i_1 & i_2 & i_3 & i_4 \end{pmatrix} \text{: a total of } 4! = 24 \text{ permutations} \right\}$

9 elements of order 2: $\begin{cases} (12),(13),(14),(23),(24),\ \boxed{(34),e} \cong \mathbf{Z}_2 \\ (12),(34),(13),(24),\ \boxed{(14),(23),}\ \boxed{e} \cong \mathbf{Z}_2 \end{cases}$

8 elements of order 3: $\begin{cases} (123),(124),(134),\ (234),\ e \cong A_4 \\ (132),(142),(143),\ (243),\ e \cong \mathbf{Z}_3 \end{cases}$

6 elements of order 4: $\begin{cases} \boxed{(2341),(3412),(4123),e} \cong \mathbf{Z}_4 \\ \boxed{(2431),(4312),(3124),e} \cong \mathbf{Z}_4 \end{cases}$

$\{(12),(34),(12),(34),e\} \cong K$ (the Klein group)

$\left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ i_1 & i_2 & i_3 & 4 \end{pmatrix} : i_n \in \{1,2,3\} \right\} \cong S_3 \cong S(\Delta)$

The group $S_5 = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ i_1 & i_2 & i_3 & i_4 & i_5 \end{pmatrix} \text{: } 5! = 120 \text{ elements} \right\}$

There are elements of orders $2, 3, 4, 5, 6$;

example of an element of order 6: $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix}$.

There are subgroups isomorphic to $\mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4, \mathbf{Z}_5, \mathbf{Z}_6$.

There are subgroups isomorphic to S_3, S_4, and K.

There are subgroups isomorphic to A_4 (of order 12).

There is a unique subgroup A_5 (of order 60).

Figure 8

Subgroups of the cyclic group $\mathbf{Z}_{12}$ and of the permutation groups S_4 and S_5. Cyclic subgroups $\mathbf{Z}_k$ are shown in red; blue distinguishes the so-called alternating groups A_4 and A_5. In our description of elements of the groups S_n, the numbers in parentheses denote cycles (that is, permutations that interchange elements in circular order)—for example:

$$(123)(4) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix} \text{(that is, } 1 \to 2 \to 3 \to 1,\ 4 \to 4\text{)},$$

$$(24)(1)(3) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix},\ (13)(24) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}.$$

$$\gamma(g \circ g') = \gamma(g) \circ (g').$$

(Briefly, one can say that a homomorphism is a map preserving the composition operation $\circ$.)

The scatterbrained soldier who ignores the commands "about **F**ace!" and "atten-**S**hun!" and reacts to both commands "**R**ight face!" and "**L**eft face!" by turning around thereby defines the homomorphism

$$\beta: R(\blacksquare) \to \mathbf{Z}_2 = \{S, F; \circ\}$$

according to the rule $\beta(S) = \beta(F) = S$, $\beta(R) = \beta(L) = F$. The totally absent-minded soldier who doesn't react to any command at all determines the *trivial homomorphism* into the trivial (one-element) group

$$\alpha: R(\blacksquare) \to \{S\}.$$

Nontrivial homomorphisms don't always exist. For example, any homomorphism $\alpha: \mathbf{Z}_5 \to \mathbf{Z}_2$ or $\beta: \mathbf{Z}_5 \to \mathbf{Z}_6$ is trivial. (Why?)

Abstract groups and Cayley's theorem

So far we've considered only concrete examples of groups, consisting of specific actions: rotations, symmetries, renumberings, and other transformations. But it's possible to approach the notion of a group from a more formal, abstract position. Then groups are assumed to consist of elements of an arbitrary, unspecified nature (not necessarily "actions"), and the operation in the group is also arbitrary (not necessarily the composition $\circ$ of actions). The corresponding definition is then expressed in *axiomatic form*:

The set G of arbitrary elements, supplied with the *binary operation* $*$ (which assigns to each pair of elements $a, b \in G$ their *product* $a * b = c$, also an element of G) is said to be an *(abstract) group* if

(1) the operation $*$ is *associative*—that is, for all $a, b, c \in G$ we have

$$a * (b * c) = (a * b) * c;$$

(2) G contains a unique *neutral* element $e \in G$, for which

$$a * e = e * a = a$$

for any $a \in G$;

(3) for each $a \in G$ there exists a unique *inverse* element $a^{-1} \in G$ such that

$$a^{-1} * a = a * a^{-1} = e.$$

This general definition yields many new examples of groups (not necessarily finite). Thus, the integers **Z** form a group (for $*$ take the operation +; then the neutral element is 0, and the inverse to $a \in \mathbf{Z}$ is $-a$); the nonzero real numbers $\mathbf{R} - \{0\}$ constitute a group with respect to multiplication; and so on.

The abstract approach, however, yields nothing fundamentally new: it turns out that any abstract group is isomorphic to a certain group of actions. We'll prove this here only for finite groups.

CAYLEY'S THEOREM. *Any finite group G is isomorphic to a certain subgroup of one of the permutation groups S_n.*

PROOF. Let $G = \{e = g_1, g_2, \ldots, g_n\}$. To each element $g_k \in G$ assign the array

$$\begin{pmatrix} 1 & 2 & 3 \ldots & n \\ i_1 & i_2 & i_3 \ldots & i_n \end{pmatrix},$$

where i_1 is the number of the element $g_k * g_1 = g_k * e$ (actually, $i_1 = k$); i_2 is the number of the element $g_k * g_2$; ...; i_n is the number of the element $g_k * g_n$. Then all the i_s are distinct (so our array is indeed a permutation), and the assignment

$$g_k \rightarrow \begin{pmatrix} 1 & 2 \ldots & n \\ i_1 & i_2 \ldots & i_n \end{pmatrix}$$

determines a homomorphism $h: G \rightarrow S_n$ (which follows from associativity); h maps G one-to-one onto a subgroup $h(G) \subset S_n$ (which follows from axioms (2) and (3)). So G and $h(G) \subset S_n$ are isomorphic.

This theorem means we now already know all the finite groups that exist—they are (to the point of isomorphism) the permutation groups and their subgroups. This doesn't mean the topic is closed. We still have a lot to learn about them: what their main properties are, where and how they are applied. But that's another story—or many other stories.

Heart waves

And the flutter that follows when they break down

by A. S. Mikhailov

APPROXIMATELY ONCE A second an excitation signal travels through the heart of a healthy person. It causes the heart muscle to contract, draw blood from the veins into a collecting chamber (the atrium), and then push it into the arteries. The signal is generated by a pacemaker nerve (the sinus node) that fires periodically, after being stimulated by the brain. Oscillations that receive an external supply of energy are called auto-oscillations, so that we may say that the sinus node works in an auto-oscillational mode.

The signal sent by the sinus node can be visualized as a wave decreasing the voltage between the outside and inside of heart cell membranes.

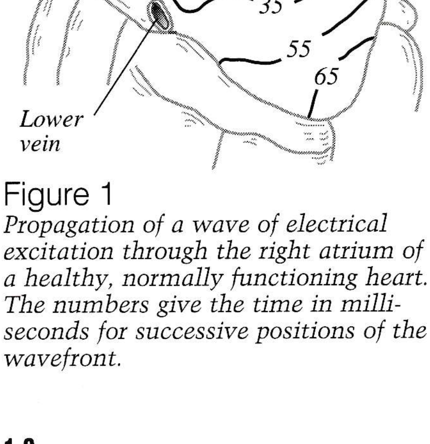


Figure 1
Propagation of a wave of electrical excitation through the right atrium of a healthy, normally functioning heart. The numbers give the time in milliseconds for successive positions of the wavefront.

This wave of electrical excitations also results in minute alterations in the electrical potential of various parts of the body. By monitoring them on an electrocardiogram, doctors can determine whether the heart is functioning properly.

Propagation of the wave of electrical excitations in a healthy heart is shown in figure 1. The wave pattern is drastically different in an unhealthy heart. Instead of infrequent waves sent regularly by the sinus node, waves rotate around certain defects. On the other hand, in the type of heart attack called paroxysmal tachycardia, the normal heartbeat breaks down and the heart contracts rapidly and irregularly. In fact, contractions of the heart as a whole may cease totally, so that only chaotic twitching of its various parts, called fibrillation, is present. This means the heart has stopped working. Paroxysmal tachycardia and fibrillation are the main causes of death from a heart attack.

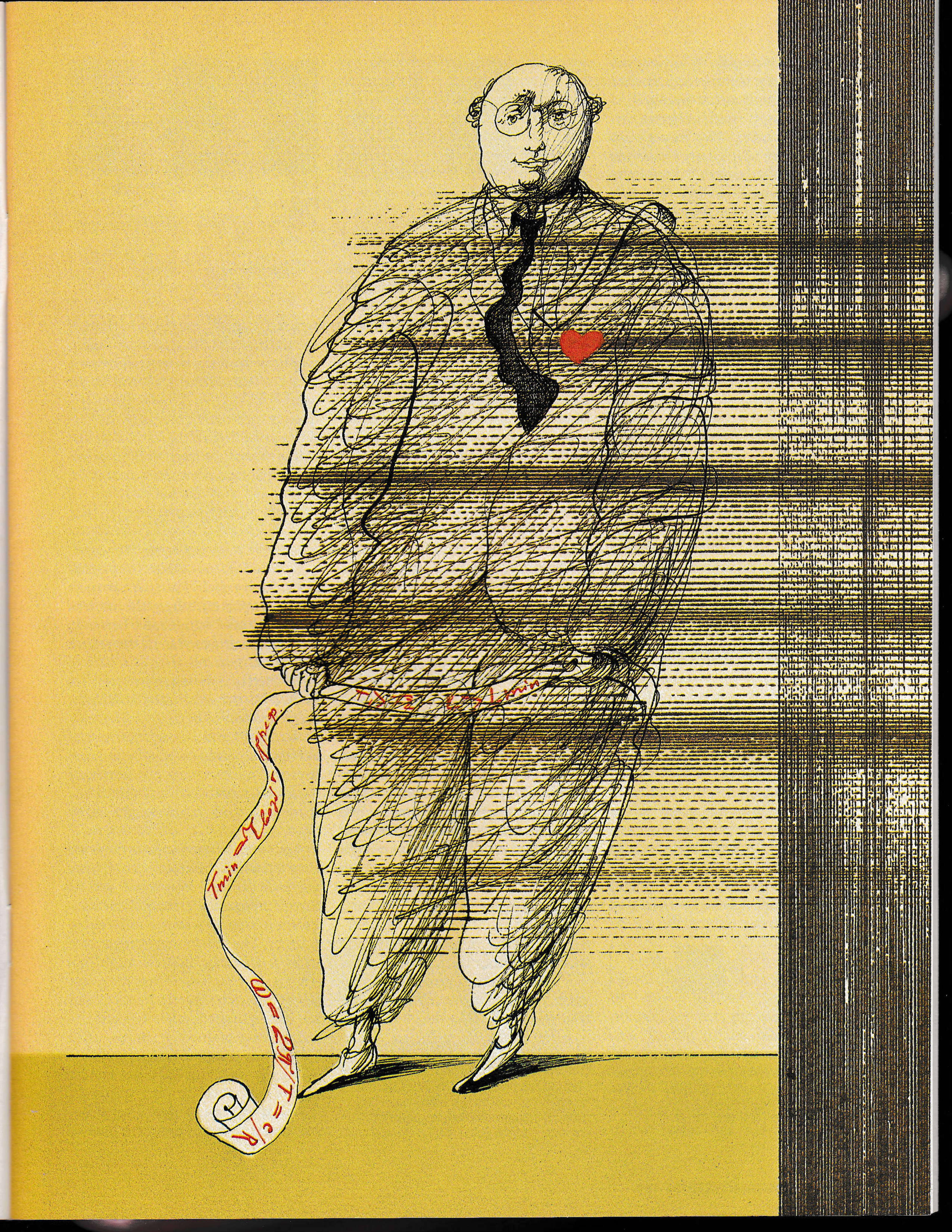
Why do tachycardia and fibrillation occur? What is the mechanism for the breakdown of the normal wave pattern in the heart? Answers to these questions are found in the mathematical theory of waves in excitable media. In this article I'll just try to give a qualitative explanation of the phenomena causing dangerous heart arrhythmia.

Excitation waves

Have you ever seen a detonation fuse burning? The flame, confined at every moment to a small region of the fuse (the combustion zone), runs along without going out, so that we can imagine a solitary wave whose shape gives us the dependence of temperature on time at various points on the fuse. A very thin region—the flame (or wave) front—separates the combustion zone from the unburnt fuse. When the flame front reaches a given point, the temperature suddenly soars; it returns to the initial value when the combustion material is used up. Obviously, when two such waves meet, they extinguish each other.

Waves in living organisms are much like combustion waves. A good example of a biological wave is the excitation impulse traveling along a nerve fiber. This is an electrical impulse that runs along a fiber (less than 0.025 mm in diameter and up to 1.5 m long) at a constant speed (generally several dozen meters per second) without fading or becoming distorted. The nerve fiber is a conductor, but a very poor one. Its electrical resistance per unit length is approximately 10^9 to 10^{10} ohms/cm, which is a hundred million times that of a copper wire with the same diameter. If the impulse had no external supply of energy during its movement along the nerve fiber, it would fade very quickly. There are electromechanical processes in living organisms that ensure the supply of energy for stable propagation of nerve impulses, but they're rather involved and we won't discuss them here.

Art by Yury Vashchenko



Waves that propagate without fading, and whose features remain constant because of a continuous infusion of energy from without, are called autowaves. The way a combustion wave spreads in a medium that can replenish itself would provide a good visual analogy. The simplest example of this process would be a prairie fire. Each time a combustion wave passes through an area, all the grass is burned; but then it slowly grows back, only to be subjected to another prairie fire. So the prairie, as an "excitable medium," is able to regenerate itself—that is, restore its initial state.

Individual muscle fibers in the heart are similar to nerve fibers in their electrical properties, but the heart's muscle tissue consists of densely intertwined fibers that are electrically connected to one another at the points of contact. An excitation impulse can go from one fiber to another, and so excitation waves are able to propagate through the heart in any direction. In this the excitable medium of heart tissue resembles the grassy prairie, whose autowaves of fire can travel in any direction.

Axiomatic model of an excitable medium

In the summer of 1945 the American mathematician Norbert Wiener, who is considered the father of cybernetics, was staying at the summer home of his friend Arturo Rosenblueth, director of the Mexican Cardiology Institute. It had been known for decades that electrical excitations propagate through the heart and that certain heart arrhythmias are linked with the circulation of excitation waves around anatomical obstacles. Wiener became interested in this phenomenon and together with Rosenblueth constructed a mathematical model that gives a simplified picture of the processes occurring in excitable media.

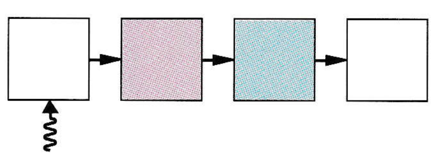


Figure 2
Sequence of changes in an element's state in an excitable medium. White, red, and blue indicate the rest, excited, and refractory states, respectively.

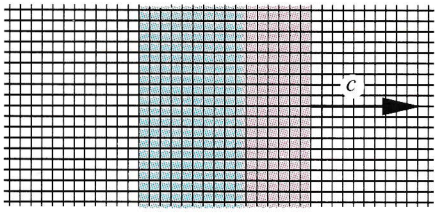


Figure 3
Excitation wave in the Wiener–Rosenblueth model of an excitable medium.

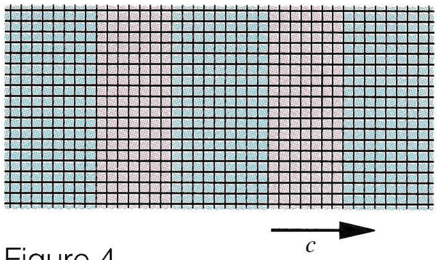


Figure 4
Sequence of closely spaced excitation waves.

In their model Wiener and Rosenblueth assume that the excitable medium is formed by a network of elements, each of which can occupy one of three states: rest, excitation, or refractoriness.[1] Transitions between these states are discontinuous and subject to certain rules.

How do these states arise? In the absence of outside influences, an element remains in the state of rest for an indefinite time. An outside influence can move it from this state to that of excitation. After being in the excited state for period of time τ_{exc}, the element moves into the refractory state, in which it remains unexcitable for a period of time τ_{ref}. After the refractory period has passed, the element recovers the ability to become excited, and it returns to the rest state. (These transitions are presented graphically in figure 2.) Another supposition of the model is that an excitation can move from excited elements to neighboring elements in the rest state, so that excitation waves can propagate through the medium (fig. 3). The propagation velocity c of these waves is constant.

According to the Wiener–Rosenblueth model, excitation waves must obey the following rules.

1. There is a minimum time interval between two successive excitation waves, equal to the sum

$$T_{min} = \tau_{exc} + \tau_{ref},$$

where τ_{exc} and τ_{ref} are the durations of the excited and the refractory states, respectively. If the waves follow each other within the minimum time span—that is, with maximum frequency—all the elements of the medium are either in the excited or the refractory state (fig. 4). The minimum distance between waves is

$$L_{min} = cT_{min} = c(\tau_{exc} + \tau_{ref}),$$

where c is the velocity of the wave.

2. When waves collide, they extinguish each other. This isn't hard to picture: upon colliding, the excited elements are "squeezed" between two approaching areas in the refractory state.

3. Excitation waves aren't reflected off the medium's boundary; they're absorbed by it.

4. "Fast" and "slow" sources of excitation waves coexist in the medium. The fast sources generate waves frequently, within short time intervals, and suppress the slow ones.

Let's verify this. Suppose there are two periodic sources generating waves in the time intervals T_1 and T_2. Figure 5 presents a graph of the evolution in time of waves sent by these sources. The left source sends waves more frequently than the right one (that is, $T_1 < T_2$). Over time the point where the two wavefronts meet and are extinguished shifts toward the right (lower frequency) source, until that source is suppressed completely. So in the long run only the fastest source "survives" in the medium.

5. An excitation wave can circulate along a ring or around a hole in

[1] Literally, the state of resisting control or authority. In physiology, the "refractory phase" is the brief period immediately after the response of a muscle or nerve before it's able to respond again.—*Ed.*

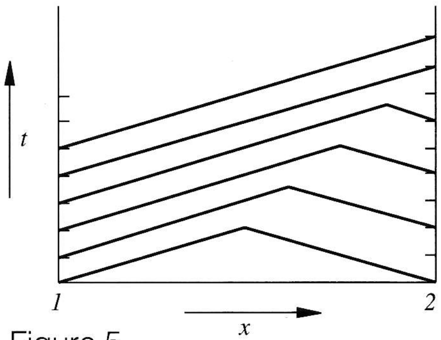


Figure 5
Effect of suppressing slow sources. The development in time of the wavefronts sent by two sources with different periods is shown. The left, high-frequency source suppresses the right, low-frequency one. The time is arrayed upward on the vertical axis; the x-coordinate is arrayed horizontally along the line connecting the centers of the two sources.

the excitable medium for an indefinite time.

This last supposition has important consequences for the possible shape of excitation waves. To see this, let's perform the following imaginary experiment. Take a strip of the excitable medium, which we'll consider two-dimensional (like a sheet of paper) but having all the properties (1–5) listed above, and roll it into a ring. If the length l of the strip is greater than L_{min} given above, an excitation wave can circulate along the ring for indefinite time (see figure 6a). The period of circulation $T = l/c$, where c is again the velocity of the wave.

Now let's enlarge the outer radius of the ring so that it tends to infinity. We get something that looks like an infinite sheet of the excitable medium with a hole in it. What happens to the excitation wave in the ring when we stretch the ring like this? It will continue to rotate, but the wavefront (the imaginary line separating excited elements from those still at rest) will bend.

In fact, the wave front can't be a radial line rotating around the hole at an angular velocity ω; if that were the case, the velocity of the wavefront at a distance r from the origin would be equal to $c = \omega r$ and would tend to infinity as r increases. But a requirement stated above calls for the velocity of excitation waves to be constant. So distant portions of the wavefront will start to lag behind those closer to the origin, so that the wavefront will take the form of a spiral.

Thus, a spiral wave can rotate around a hole in an excitable medium for an indefinite time without fading. The period of rotation is the time interval needed for an excitation impulse to travel around the perimeter l of the hole—that is, $T = l/c$. If the hole is circular with radius R, then $T = 2\pi R/c$, and the rotation frequency $\omega = 2\pi/T = c/R$.

These are the main points of the Wiener–Rosenblueth model. The question is, how can we use the model to explain the outbreak of heart arrhythmia?

Spiral waves and heart arrhythmia

The walls of the atrium are thin enough to allow us to look at heart tissue as a two-dimensional excitable medium. Because of the blood shortage caused by a heart attack, small areas of heart tissue may die, losing the ability to conduct excitation waves. These areas of dead tissue are, in effect, holes in the excitable medium, and according to the arguments given above they will cause the circulation of excitation waves around them. What will the final outcome of such circulation be?

The answer depends on the ratio of the circulation period to the frequency at which the sinus node sends waves. If the rotation frequency of the spiral wave is greater than the impulse frequency of the sinus node, the spiral wave turns out to be a faster source and suppresses the activity of the sinus node. And that's generally what happens, because the radii of the holes (the damaged regions of the heart) are small, and the circulation frequency of the wave is inversely proportional to the radius. Then the wave rotating around the hole takes over the entire atrium, and the normal functioning of the heart breaks down. Instead, the heart begins to "flutter"—that is, arrhythmia sets in.

Excitation waves can also be generated by discontinuities in the fronts of other excitation waves. To see this, let's suppose that the medium contains a region D of elements with a longer refractory period. Suppose two successive excitation waves travel through the medium within a short span of time (fig. 7a). After the first wave passes through region D, elements located in D will remain in the refractory state for a long time. As a result, the excitation from the second wave can't penetrate the region—the wave front tears (fig. 7b). After a while the refractory state in region D ends, and its elements regain the ability to become excited. If at this point the second wave has not yet passed through region D, an excitation from the front of this wave will begin to propagate inside that region (fig. 7c). Then the wave will begin to curl around the tear, and after a while a spiral wave is generated (fig. 7d).

Heart tissue is not homogeneous, especially with regard to the duration of the refractory state. This nonhomogeneity increases if the normal "feeding" of the cells is disrupted by a heart attack. When the heart is functioning normally, the time between two successive excitation waves is usually 1 s, and the duration of the refractory state is approxi-

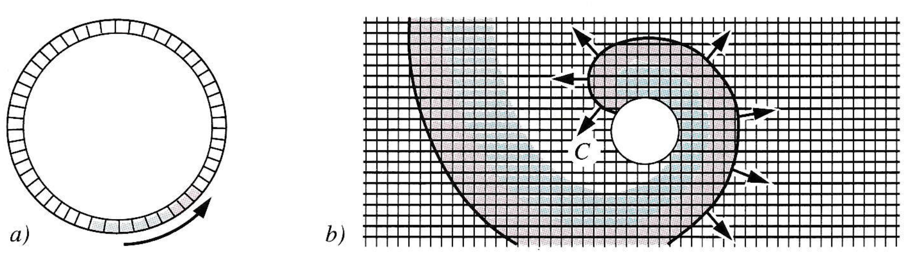


Figure 6
Circulation of the excitation wave: (a) a solitary wave rotating in a ring; (b) a spiral wave rotating around a hole in the excitable medium.

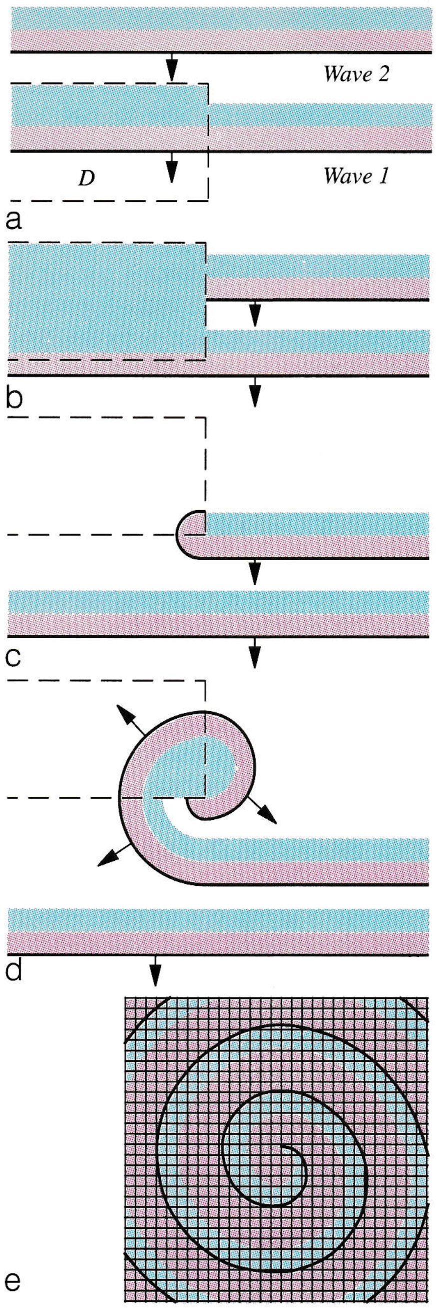


Figure 7
Generating a spiral wave (D is a region with long refractory periods): (a) the initial positions of the two waves; (b) the second wave doesn't enter region D, discontinuities in the front appear; (c) the excitation wave begins to propagate in region D; (d) the spiral wave appears; (e) the spiral wave after a sufficiently long period of time.

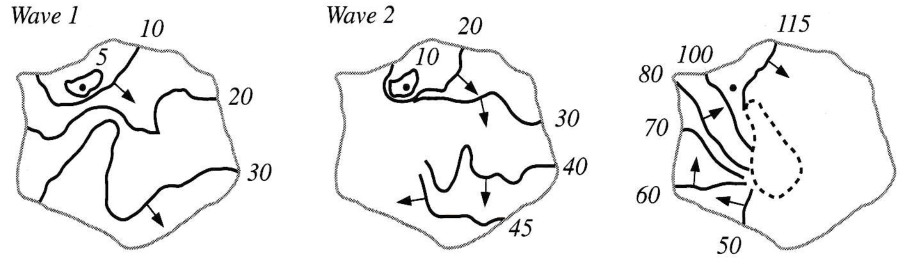


Figure 8
Generating a spiral wave in a strip of heart tissue from a rabbit. The numbers indicate the time in milliseconds. Wave 2 is broken into regions with a long refractory period. The spiral wave appears. Because of the small size of the strip, only a small portion of the spiral actually forms.

mately 0.1 s; so a small scattering in the values of the refractory time, which is natural in living tissue, doesn't cause tearing at the fronts of excitation waves. If the heart tissue is damaged, however, the sinus node can generate an additional wave that follows the previous one by a very short time interval. If this interval is short enough—less than the sum of the maximum duration of the refractory state and the duration of the excited state—the front of such a wave will be torn at the boundary of the region with the longer refractory state, and a spiral wave will be generated. Figure 8 shows the generation of a spiral wave by a tear in the wavefront in an experiment with a tissue sample from a rabbit's heart.

Understanding the dynamics of spiral waves is important for the study of serious heart illnesses. Let's make this point clear by the following examples. We know that in nonhomogeneous tissue the lifetime of a spiral wave, which is the alleged cause of tachycardia, is finite. Because of the excessively rapid action of the heart, or the nonhomogeneous structure of the tissue, the wave travels randomly and its shape changes. Sooner or later the wave arrives at the boundary of the tissue and disappears. So we have a happy ending: when the rotating spiral wave disappears, the tachycardia stops.

But another outcome is also possible. Secondary tears can appear at the front of the excitation wave, in its outlying region, when the wave passes through a region with a long refractory period, so that new spiral waves can appear. If the number of waves generated per second is greater than that of the waves absorbed at the boundary, the number of spiral waves will grow in time until the entire medium is filled with small scraps of spiral waves. This chaotic situation is observed when fibrillation occurs—that is, contractions of the heart as a whole cease and only various parts of it flutter chaotically.

The important point is that the onset of fibrillation requires that the heart's tissue exceed some "critical mass." Until the tissue sample reaches a certain size, self-supporting fibrillation can't be generated. For example, the critical mass is approximately 20 mg for the atrium tissue of a rabbit.

Within the framework of the theory given above, this phenomenon isn't hard to explain. Assuming that the refractory states are randomly distributed in the medium, we see that the probability of generating a new spiral wave by a tear at the front of a wave that already exists is approximately the same at any point in the medium. The total number n_+ of new spiral waves generated in the medium per unit of time is proportional to the area occupied by the medium, which we'll consider two-dimensional, and the number of existing spiral waves N. So we can write the equation

$$n_+ = \alpha L^2 N,$$

where L is the linear size of the medium and α is a constant coefficient. The number of spiral waves that disappear at the boundary per second is given by

$$n_- = \beta N L$$

—that is, we assume it to be proportional to the length of the boundary (β is a constant coefficient). In the chaotic mode observed during fibrillation, the rate at which spiral waves are generated is greater than that of their disappearance—that is, $n_+ > n_-$. According to the arguments given above, this requires that the linear size of the medium be greater than the critical value $L_c = \beta/\alpha$. For smaller samples, the generation of spiral waves will be canceled by the boundary.

An understanding of the mechanisms leading to heart arrhythmia is, of course, essential for finding effective methods of curing or preventing it. The examples given above make it clear that certain simple arguments derived from the physics of vibrations may provide a clue.

In this short article I've confined myself to the simplest and most "graphic" explanations. The Wiener–Rosenblueth model played its role in creating a theory of autowave phenomena. In the 1960s a group of scientists studied the model closely in one of I. M. Gelfand's seminars[2] and made some important generalizations, so that new and more detailed models are now available for waves in excitable media such as heart tissue. These more sophisticated models make use of the theory of partial differential equations and can provide more precise answers to the questions discussed in this article. But the Wiener–Rosenblueth model is still useful when we need to get a qualitative picture of a phenomenon by the simplest possible means.

[2]See "A Talk with Professor I. M. Gelfand" (Jan./Feb. 1991).—*Ed.*

Challenges in physics and math

Math

M36

Fifth follows four: right angles in a square. Point L divides the diagonal AC of a square $ABCD$ in the ratio 3:1 (fig. 1), K is the midpoint of side AB. Prove that angle KLD is a right angle. (Y. Bogaturov)

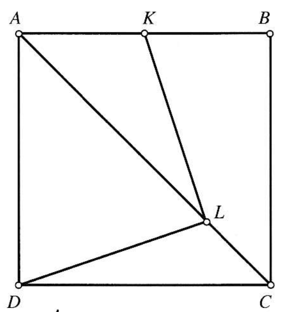


Figure 1

M37

Annihilating array. The numbers 1, 2, ..., 100 are arranged in a 10 x 10 square table in their natural order (1 in the top left corner, 100 in the bottom right corner). The signs of 50 of these numbers are changed in such a way that exactly half of the numbers in each line and each column get the minus sign. Prove that the sum of all the numbers in the table after this change is zero. (S. Ageyev)

M38

Fifth follows four: parallel segments in a pentagon. Each of four sides of a convex pentagon is parallel to one of its diagonals (having no end points in common with the side—see figure 2). Prove that the same holds for the fifth side, too. (E. Turkevich)

M39

Baker's dozen cuts. Is it possible to draw 13 straight lines across a chessboard such that, after cutting it along the lines, each piece will contain at most one center of a square? (The lines must not pass through the centers.) (A. Pechkovsky)

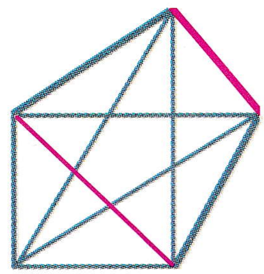


Figure 2

M40

Fifth follows four: numbers in a sequence. In the sequence 1, 9, 9, 1, 0, 9, ... each number, beginning with the fifth one, is the last digit of the sum of the preceding four. Will we ever meet in this sequence the succession of numbers (a) 1, 2, 3, 4; (b) 1, 9, 9, 1 a second time; (c) 2, 1, 9, 9? (G. Gurevich)

Physics

P36

What will the speed be? A constant force F starts to act on a body that is moving with a constant velocity v. After an interval of time τ the body's speed becomes $v/2$. During the next interval of the same duration the speed is again halved. Figure out what the body's speed will be three time intervals after the force begins to act. (S. Krotov)

P37

Child on a hill. A child climbing slowly up a snowy hill is pulling a sled with a rope such that the rope is always parallel to the hill's surface. The hilltop is at a height h and distance l from the foot of the hill. Find the work the child does in lifting the sled to the hilltop. (The sled's mass is m, the coefficient of friction is μ.) (A. Buzdin)

P38

Alpine helicopter. Helicopter pilots in the mountains prefer to fly early in the day, right after dawn. Fine morning weather (which might seem to be the reason) doesn't matter in this case. What's the real reason? (G. Myakishev)

P39

What's the resistance? The shape shown in figure 3 is made of wire of constant cross section. The side of the bigger square is a, and a one-meter length of the wire has a resistance ρ. Determine the resistance between points A and B. (A. Zilberman)

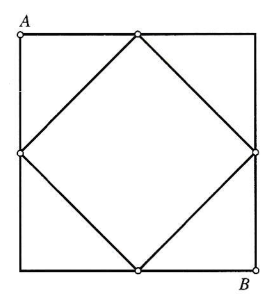


Figure 3

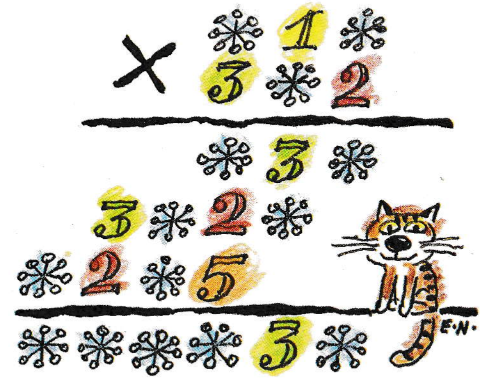
P40

Semitransparent mirrors. Measurements show that a semitransparent mirror lets through about 1/5 of the incident light; the rest is reflected. If two identical mirrors of this kind are set perpendicular to the incident light beam, we might expect the pair to let through 1/25 of the light. But actually about 1/10 of the incident light passes through. What's the reason? (V. Golubev)

ANSWERS, HINTS & SOLUTIONS ON PAGE 58

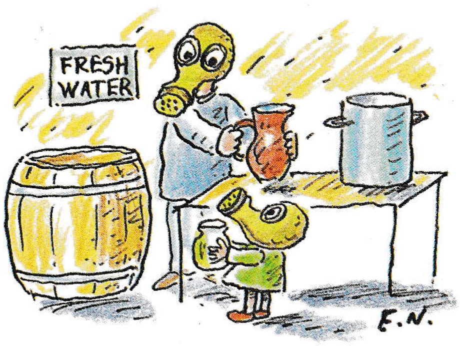
BRAINTEASERS

Just for the fun of it!



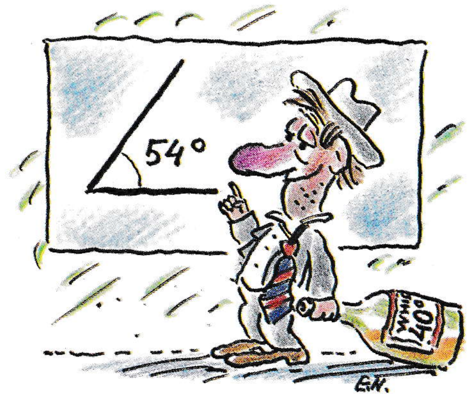
B36

Restore the sum. The figure at right shows a sum in which some of the figures are rubbed out and replaced by asterisks. Restore the absent figures. (L. Yakovleva)



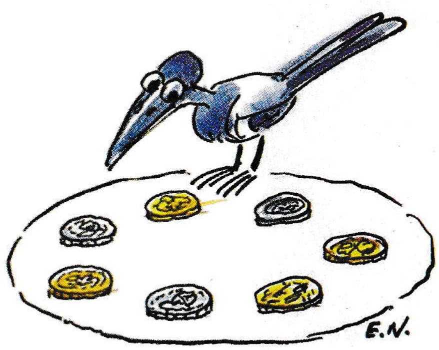
B37

Jug, pot, and barrel. You have a four-liter jug, a six-liter pot (like in the drawing at left), and a big barrel of water. Can you measure exactly one liter of water? (G. Kushnirenko)



B38

Possibility of trisection. One of the three famous ancient geometrical problems is the problem of trisecting an angle with ruler and compass. It was proved long ago that this problem, along with the other two, is in general unsolvable. But there are some exceptions to this rule. Trisect an angle of 54° using only a compass (that is, construct points that lie on the rays dividing the angle into three congruent parts). (A. Shvetsov)



B39

Turning seven by fives. Seven coins lie along a circle. Can you turn them all upside down if you're only allowed to turn over any five coins in succession at a time? Can you do it by turning over a succession of four coins at a time? (A. Shvetsov).

B40

Traveling light. Dashing sea captains wouldn't fill their holds completely when they transported cotton from Australia to England. It would have been to their advantage to take as much as they could, but they didn't. Why not? (L. Mochalov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Art by Edward Nazarov

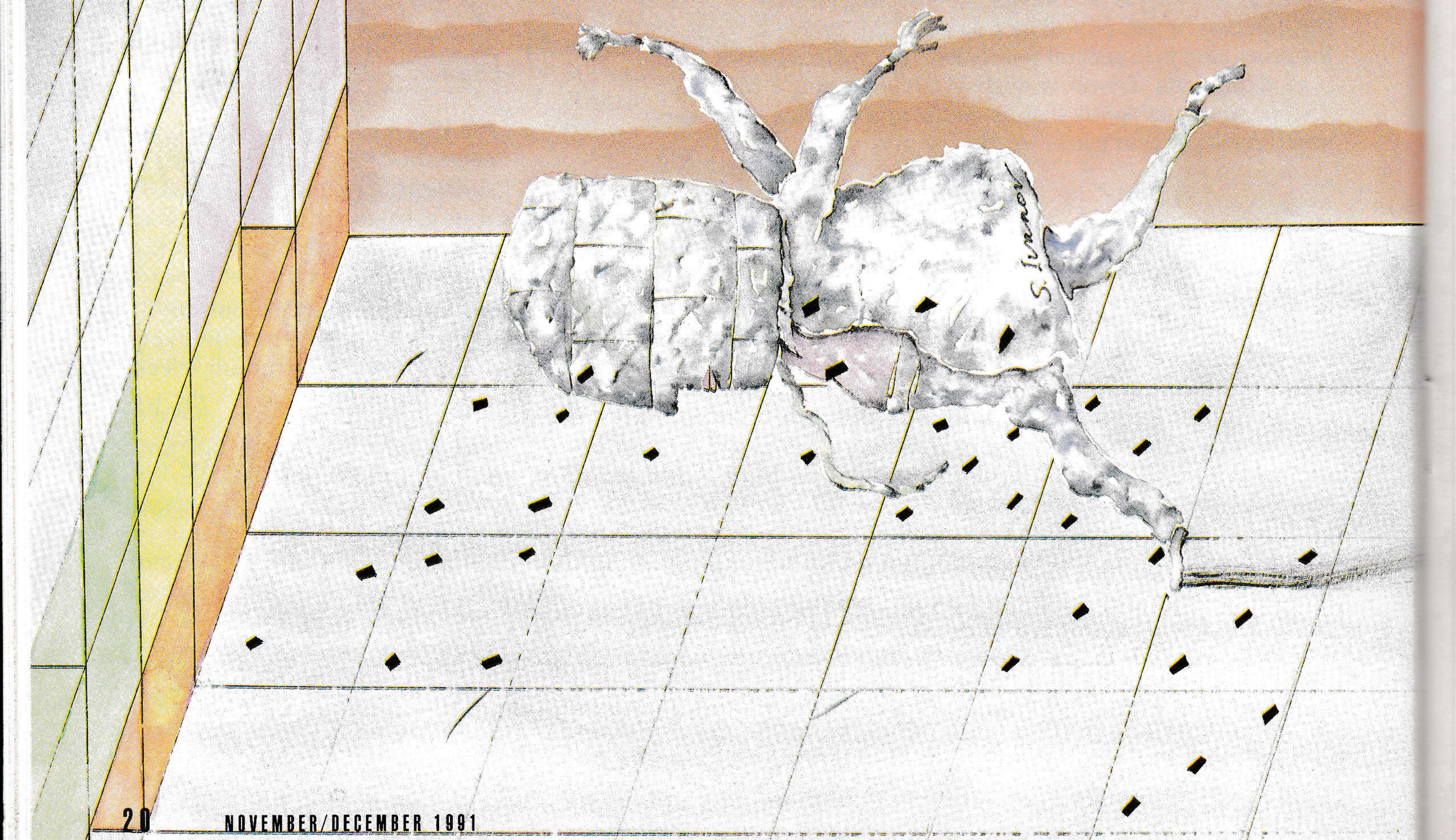
Getting it together with "polyominoes"

It all started with the humble domino . . .

by Dmitry V. Fomin

YOU'VE PROBABLY PLAYED, or at least heard of, the exciting computer game Tetris™, which was invented in the Soviet Union and is now popular around the world. In this game the player tries to manipulate plane figures, made up of four squares, that "fall from the sky" so as to make as many solid rows of squares as possible. You can see all the seven shapes in figure 1; by analogy with dominoes, they're called tetrominoes. The rules of the game allow the figures to be shifted and rotated 90 degrees: that's why there are just seven of them (all possible unbroken four-square figures with sides touching are equivalent to these seven, excluding translations and rotations).

Another puzzler, pentominoes, was popular long before Tetris ap-



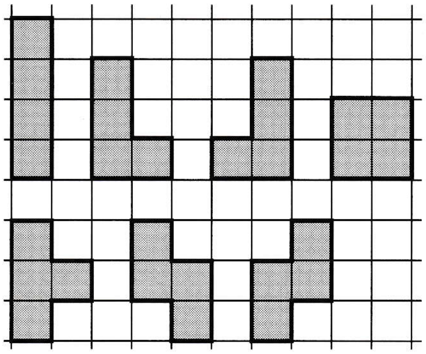


Figure 1

peared. Its building blocks are a set of wooden or plastic five-square tiles, which can be not only moved but turned over. The set comprises twelve pieces—every possible shape occurs once. You can use them to tile, for instance, a checkerboard with the central 4 x 4 square cut out. (Try it yourself!) This problem was published way back in 1907 by the famous puzzle inventor Henry E. Dudeney, but pentominoes and other polyominoes owe their fame to professor Solomon W. Golomb of the University of Southern California. The detailed story of this type of puzzle, math problems connected with it, and how they're solved can be found in his excellent book *Polyominoes* (New York, Charles Scribner's Sons).

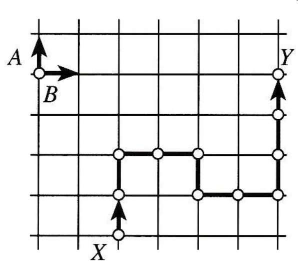
As for this article, we'll deal with just one (though probably the most numerous) class of polyomino problems, asking whether a given figure on the checkered plane can be tiled by polyominoes of some given type. In order to give a *positive* answer to such a problem, it's sufficient simply to produce the appropriate tiling (which is usually found by some cleverly arranged search, often a very laborious one). But using the brute force of a search method to prove that the required tiling *does not exist* can prove fruitless because of the astronomical number of possibilities to be checked. Instead, proofs of nonexistence of the required tiling often employ a special coloring of the plane (see Golomb's book). I'll demonstrate an approach to tiling problems that is based on group theory. It was suggested to me by the young Leningrad mathematician Oleg Izhboldin in 1987. All the information you need about groups can be found in Alexey Sosinsky's article on page 6.

From polyominoes to groups

Consider an infinite grid of unit squares on the plane. A figure consisting of a finite number of squares is called a polyomino if any two of its squares can be connected by a chain of squares such that every subsequent square in the chain adjoins the preceding one along a whole side. For simplicity, we'll confine ourselves to polyominoes bounded by one closed non-self-intersecting polygonal curve—that is, polyominoes without holes.

Let XY be a directed polygonal curve, originating at node X and ending at node Y, formed by the lines of our grid (unit segments). Let's move along it from its starting point to the end point. Each segment between adjoining nodes of the grid will proceed in one of four directions, so that we can designate our curve as a sequence of directions of movement.

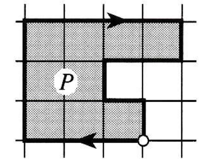
Now let's take an arbitrary group G and two elements A and B in it. We'll label an element directed upward A and one directed to the right B; we'll label the inverse elements—those directed downward and to the left—A^{-1} and B^{-1}, respectively. We'll consider this entire curve to be an element of the group $t(XY)$, equal to the product of these four elements taken in the order in which the directions are encountered when we move along the curve (fig. 2).



$t(XY) = AABBA^{-1}BBAAA = A^2B^2A^{-1}B^2A^3$

Figure 2

The same product can be written for a polygon bounding some polyomino P if we denote the starting point (grid node O) and choose the direction of travel (say, clockwise). We'll denote it by $t_O(P)$ (fig. 3).



$t_O(P) = B^{-3}A^3B^4A^{-1}B^{-2}A^{-1}BA^{-1}$

Figure 3

Try to work through the exercises that follow—you'll find them useful later in the article.

Exercise 1. Prove that $t(YX) = t(XY)^{-1}$, where YX is the same polygonal curve XY but traced in the opposite direction, from Y to X.

Exercise 2. Let O and O_1 be two nodes on the boundary of polyomino P. Prove that

$$t_{O_1}(P) = t(O_1O)t_O(P)t(OO_1).$$

It follows from these exercises that if the product $t_O(P)$ is the unit (or neutral) element e of group G, then the equality $t_O(P) = e$ still holds even if we reverse our direction of travel on the boundary or switch our starting point (from O to O_1), since in this case

$$\begin{aligned} t_{O_1}(P) &= t(O_1O)t(OO_1) \\ &= t(O_1O)t(O_1O)^{-1} = e. \end{aligned}$$

So we can simply write $t(P) = e$ without indicating either the origin or the direction.

Now everything is ready, and we can state and prove the

Basic Theorem. *If a polyomino P can be cut into polyominoes P_1, P_2, ..., P_k such that, for elements A and B of some group G, $t(P_1) = t(P_2) = ... = t(P_k) = e$, then $t(P) = e$ as well.*

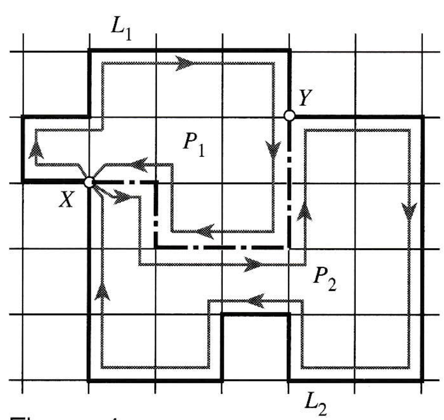


Figure 4

Art by Sergey Ivanov

Let's begin by proving the theorem for a polyomino P composed of just two pieces, P_1 and P_2, that border on each other along the polygonal curve XY. Starting from X, move along the common section of the boundaries of P_1 and P, complete a circuit around P_1, and then make a full circuit around P_2 (see figure 4), writing down the product of the elements of G traced by this path. We pass along curve YX forward and backward, so the corresponding factors in the product cancel out, and two successive circuits (around P_1 and P_2) amount to one circuit around P:

$$\begin{aligned} t_X(P_1)t_X(P_2) &= t(L_1)t(YX)t(XY)t(L_2) \\ &= t(L_1)t(L_2) \\ &= t_X(P), \end{aligned}$$

where L_1 is the common portion of the boundaries of P_1 and P from X to Y, and L_2 is the common portion of the boundaries of P_2 and P from Y to X. So $t_X(P) = t(P_1)t(P_2) = e$.

The thorough proof for the case of an arbitrary number k of pieces into which a polyomino P is cut proceeds analogously, using induction over k. We just need to notice that, out of all the polyominoes $P_1, \ldots, P_k$ that constitute P, one will always be found that, combined with the rest, form the same kind of pair as P_1 and P_2 in the preceding argument.

Exercise 3. Prove that if $AB = BA$, then $t(P) = e$ for any polyomino P.

It's clear from the last exercise that it's pointless to use commutative groups in polyomino tiling problems: such groups would not give any information. It's also clear that we can limit ourselves to groups whose elements are all products of some elements A and B and their inverses (any number of them in any order). Such groups are said to be *groups generated by two elements*, and these two elements are called the *generators* of the group. Here are two examples (which also appeared in Sosinsky's article).

The first is the group S_3 of permutations of the numbers 1, 2, 3. It is generated by the permutations

$$A = \begin{pmatrix} 1\,2\,3 \\ 2\,1\,3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1\,2\,3 \\ 1\,3\,2 \end{pmatrix}.$$

Exercise 4. Verify that the four permutations of S_3 distinct from A and B coincide with $A^2 = e$, AB, BA, and ABA.

The second example is the so-called dihedral group D_n ($n = 2, 3, \ldots$)—that is, the symmetry group of a regular n-gon (for $n = 2$, the "n-gon" is considered a line segment). It consists of all the isometries of the plane that map this n-gon onto itself: these are n rotations about the n-gon's center through the angles 0, $2\pi/n$, $4\pi/n$, ..., $2(n-1)\pi/n$, and n line reflections in the midperpendiculars to the sides (fig. 5). We can choose as the generators of the group the rotation R through angle $2\pi/n$ and any line reflection r.

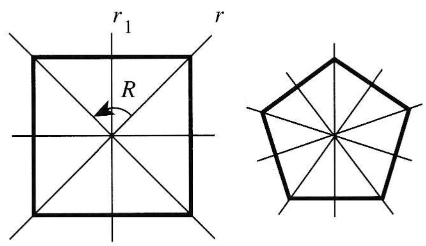


Figure 5

Exercise 5. Check that all rotations in group D_n can be written in the form $R, R^2, R^3, \ldots, R^n = e$, and all the line reflections in the form $r, rR, rR^2, \ldots, rR^{n-1}$. Later, though, another pair of generators, r and $r_1 = rR$, will be used. (They're actually the generators, since $R = rr_1$.) All dihedral groups except D_2 are noncommutative: $rr_1 = R \neq R^{-1} = r_1r$.

Now we can return to the problem of polyomino tiling. Usually a certain set of standard polyominoes is given (for example, the twelve pentominoes) and various figures are to be built from them. The standard polyominoes may be shifted, rotated, and overturned at will. Let's reformulate the Basic Theorem so it can be applied to this problem more conveniently.

But first, one more definition. We'll call a noncommutative group with two generators A and B a *null group* for polyomino P (with respect to A and B) if, for any position of P on the grid, $t(P) = e$.

The Basic Theorem directly results in the following

TILING CONDITION. *If G is a null group for each of the polyominoes $P_1, \ldots, P_k$, then G is also a null group for any polyomino that can be tiled by polyominoes of these types.*

Examples

1. *Monomino.* For a single unit square P, regardless of its position on the grid, $t(P) = ABA^{-1}B^{-1}$ (fig. 6a). The equality $t(P) = e$ would mean that $AB = BA$—that is, that G is commutative. But we've decided not to use any commutative groups: they can't help us. So there is no null group for a monomino. And this is quite natural, since obviously any polyomino can be tiled by monominoes.

2. *Domino.* Dominoes can be placed on the grid in two different ways, and the corresponding products are

$$t_1 = AB^2A^{-1}B^{-2}$$

and

$$t_2 = A^2BA^{-2}B^{-1}$$

(fig. 6b). The group S_3, with the abovementioned generators

$$A = \begin{pmatrix} 1\,2\,3 \\ 2\,1\,3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1\,2\,3 \\ 1\,3\,2 \end{pmatrix},$$

is a null group for dominoes. Indeed, since $A^2 = B^2 = e$, we have

$$t_1 = AeA^{-1}e = AA^{-1} = e,$$
$$t_2 = BB^{-1} = e.$$

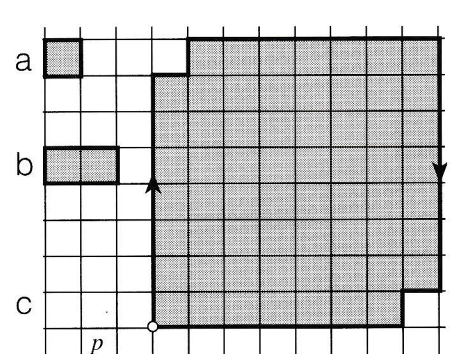


Figure 6

Now we can solve the well-known problem of whether it's possible to tile a chessboard with dominoes when two opposite corner squares—say, a8 and h1—are lopped off (fig. 6c). Trace out the boundary of such a chessboard, starting from one of the corners that's intact (as in figure 6c). We get the product

$$\begin{aligned} t &= A^7BAB^7A^{-7}B^{-1}A^{-1}B^{-7} \\ &= (AB)^4 \\ &= AB \neq e, \end{aligned}$$

since

$$AB = \begin{pmatrix} 1\,2\,3 \\ 3\,1\,2 \end{pmatrix}.$$

Consequently, the required tiling is impossible. Here we might as well use the symmetry group D_3 of a regular triangle (the generators would be the reflections in two of the triangle's heights). You may find it easier to deal with this second group. But S_3 and D_3 are, in fact, two "incarnations" of one and the same abstract group (that is, they're *isomorphic*); you'll see this right away when you number the triangle's vertices and follow the permutations of the numbers under the symmetries of the triangle.

3. *P x q–omino.* Let P be a rectangle measuring $p \times q$ (fig. 6d). Consider the group G of permutations generated by two cyclic permutations of length p (p-cycles), which have exactly one element in common (fig. 7):

$$A = \begin{pmatrix} 1 & 2 & \dots & p-1 & p & p+1 & \dots & 2p-1 \\ 2 & 3 & \dots & p & 1 & p+1 & \dots & 2p-1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 1 & 2 & \dots & p-1 & p & p+1 & \dots & 2p-1 \\ 1 & 2 & \dots & p-1 & p+1 & p+2 & \dots & p \end{pmatrix}.$$

The two products $t_1 = A^pB^qA^{-p}B^{-q}$ and $t_2 = B^pA^qB^{-p}A^{-q}$ correspond to the two possible positions of P on the plane. Since $A^p = B^p = e$, both products are equal to e, so G is a null group for the rectangular polyomino P. The other null group for P is the group G_1 generated by two q-cycles A_1 and B_1 that have a single common element.

Using this construction, let's prove that *a rectangle Q measuring $m \times n$ can be tiled by rectangles measuring $p \times q$ only if each of the numbers p and q is a factor of at least one of the numbers m and n.*

In this case the tiling condition has the form $t(Q) = A^nB^mA^{-n}B^{-m} = e$ or $A^nB^m = B^mA^n$. But for our cyclic permutations A and B this equality is possible only if $A^n = e$ and $B^m = e$ (watch the common element of the cycles during permutations A^nB^m and B^mA^n). But $A^n = e$ only if p divides n, and $B^m = e$ only if p divides m. Replacing cycles A and B by the similar q-cycles A_1 and B_1, we find that either n or m is divisible by q.

It's clear that the required tiling is possible only if both of the numbers m and n has the form $xp + yq$, where x and y are some nonnegative integers, since the side of the larger rectangle must be composed of the sides of the smaller ones. Bringing together all these conditions, we find that *either (a) one of the two numbers m and n must be divisible by p and the other one by q; or (b) one of these numbers is divisible by both p and q and the other one has the form $xp + yq$ ($x, y \geq 0$).*

Exercise 6. Prove that the last condition is not only necessary but also sufficient for tiling a $m \times n$–omino with $p \times q$–ominoes.

Exercise 7. Prove that an $m \times n$ rectangle can be tiled with $p \times p$ and $q \times q$ squares if and only if either (a) both m and n are divisible by p, (b) both are divisible by q, or (c) one is divisible by both p and q and the other is of the form $xp + yq$, where x and y are nonnegative integers. (Hint: the group of permutations generated by a p- or q-cycle having exactly one common element is a null group for both $p \times p$ and $q \times q$ squares.)

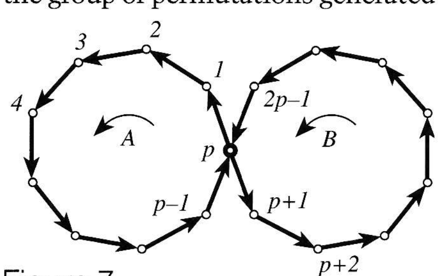


Figure 7

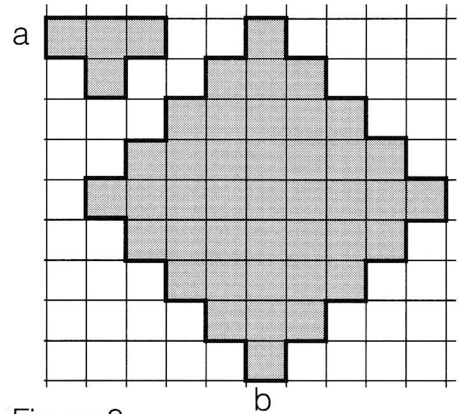


Figure 8

In these exercises it's possible to obtain necessary and sufficient tiling conditions. But I must admit that this is a rare case. Our theorem yields only a necessary condition and can be used, basically, when one must prove that some tiling does not exist.

A few problems with null groups

1. Prove that the symmetry group D_4 of a square with generators r and r_1 (fig. 5) is a null group for the T-tetromino (fig. 8a).

2. Is it true that the group D_4 (from problem A) is a null group for any $4k$-omino ($k = 1, 2, 3, \dots$) with respect to generators r and r_1?

3. Prove that (a) for $n = 2, 3, 4$, (b) for any $n \geq 2$, there exists a common null group for all n-ominoes.

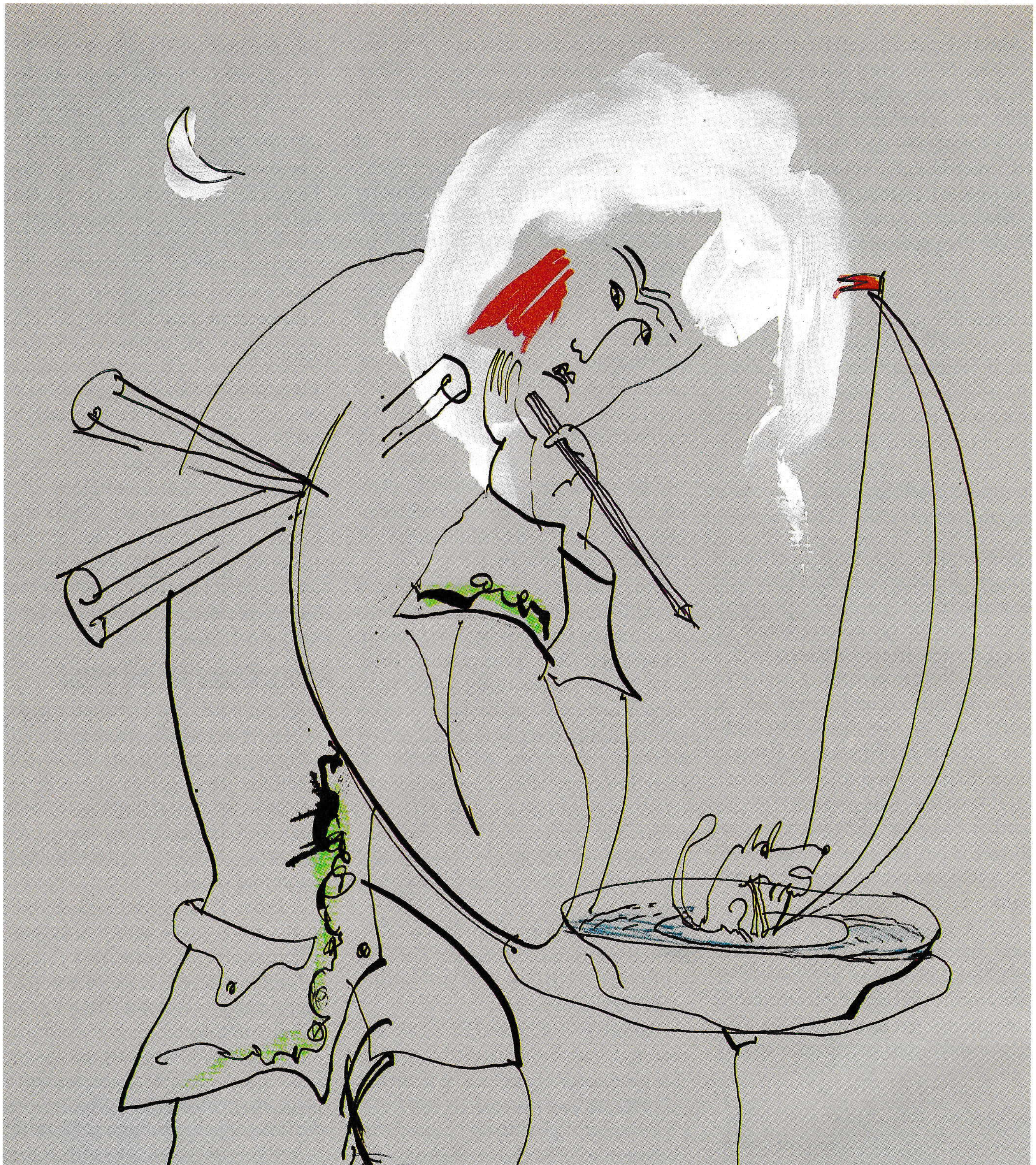
4. Prove that D_{18} is a null group for the polyomino in figure 8b.

5. Prove the

Theorem on Chessboard Coloring. *Color the squares of polyomino P black and white like a chessboard, and denote the absolute value of the difference between the number of black and white squares by $c(P)$. If $c(P) > 0$, then the group $D_{2c(P)}$ is a null group for P; whereas for any $n > 2c(P)$, the group D_n is not a null group for P.*

Now *you* try thinking up other examples of null groups and use them to solve some interesting problems of polyomino tiling.

HINTS ON PAGE 61



The wolf, the baron, and Isaac Newton

About a cartoon, a novel by Cyrano de Bergerac, and laser nuclear fusion

by V. A. Fabrikant

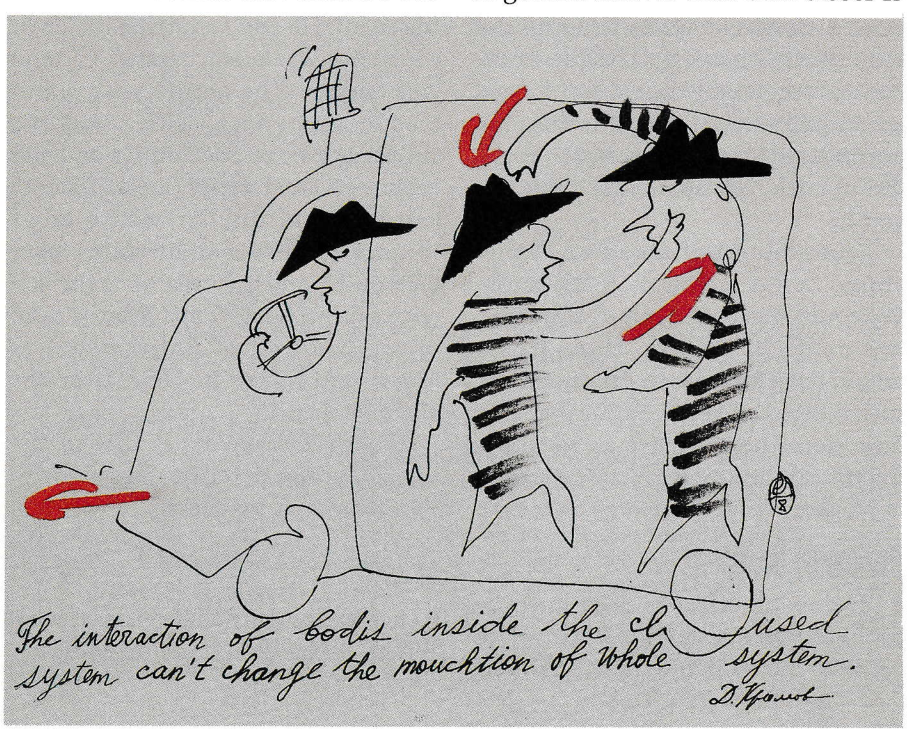
IN THIS ARTICLE YOU'LL FIND a discussion of a physical phenomenon that was shown in a popular Soviet cartoon. A wolf in a sail boat is trying to catch up to a hare traveling by steamship. To increase the speed of his sailboat, the wolf is blowing into the sail. At first glance, it may seem that this situation is vaguely similar to the one described by Baron Münchhausen: when he was sinking in a marsh he managed to pull himself out by his own hair. There is, however, a significant distinction between the two cases. The wolf hasn't broken that fundamental law of mechanics, Newton's third law, while the baron has indeed broken it, which is, of course, impossible.

Sir Isaac Newton, in his *Philosophiae Naturalis Principia Mathematica*[1] ("*Principia*" for short), formulated the third law this way: "To every action there is a reaction; that is, the actions of two bodies on one another are equal and opposite in direction." Newton explained: "If something is pushing something else, or pulling it, then the former is itself pushed or pulled by the latter. If someone pushed one's finger against a stone, the finger is also pushed by the stone." It follows from Newton's third law that no interaction inside a closed system (from which nothing escapes) can influence the movement of the system as a whole. In particular, the interaction of the parts of the Baron's body (hands and hair) could not have caused a change in his rate of sinking, let alone pull him out of the marsh.

To slow down his immersion, the baron would have had to begin undressing and throwing his clothes down into the marsh—that is, he needed to open up the system. His heavy jackboots would have been especially useful. Finally, the baron could have stood on the saddle and jumped from it to solid ground. But in doing so he would have shoved his horse down into the muck and hastened the poor animal's doom.

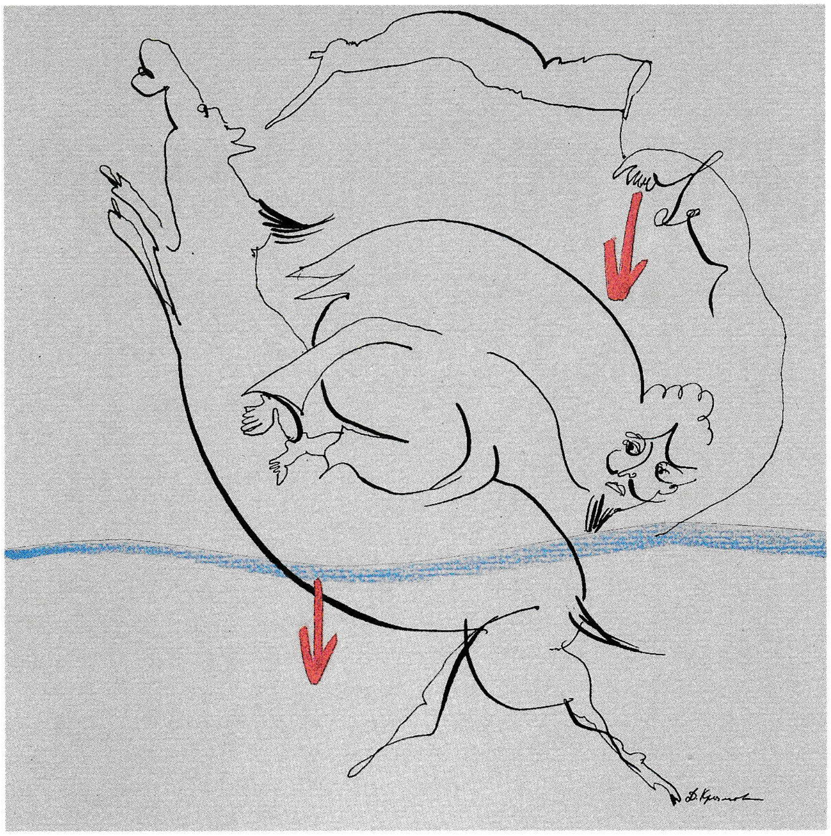
But let's get back to our cartoon. We'll suppose that the sail is made of a kind of material that absorbs the stream of air produced by the wolf. Then the sailboat and the wolf would constitute a closed system, and without wind the boat couldn't budge no matter how hard the wolf blew. The wolf's attempts to speed the boat up would be equally fruitless (if the boat is already moving at the same speed as the wind): the pressure of the air blown by the wolf would be balanced by the action of his claws on the gunwale of the boat. The fact is, the wolf experiences a recoil when he blows forward—he's forced back according to the law of action and reaction.

This is similar to a gunshot: the bullet or shell flies out in one direction and the gun begins to move in the opposite direction. Every hunter or gunner knows this. Still closer is



Art by Dmitry Krymov

[1] "Mathematical principles of natural philosophy" (Latin).



the analogy of a missile, whose exhaust shoots out one end while it takes off in the other direction.

In the cartoon, though, the boat actually has a sail of ordinary material, which naturally deflects the stream of air rather than absorbing it. So the air blown by the wolf comes back after being deflected by the sail and leaves the "sailboat-and-wolf system." In other words, the system ceases to be closed as soon as the wolf starts blowing. The phenomenon of recoil increases the sailboat's speed. (All jet planes exploit this phenomenon: their engines send out gas jets in back to propel the aircraft forward.)

So it turns out the wolf was better versed in mechanics than the baron. We could suggest an improvement in his mode of sailing, though. He would have found it more efficient to turn around and blow in the opposite direction: the sail isn't an ideal deflector, so the stream of air sent back is weakened.

Cyrano de Bergerac

The recoil effect is used in space rockets nowadays. But here we've come to the French writer Cyrano de Bergerac (1619–1655). He is a very curious figure in world literature. Cyrano descended from a impoverished noble family and spent his entire life in need. At that time only poets serving rich aristocrats prospered. Cyrano couldn't serve anybody because he was extremely independent and hot-tempered. Cyrano wrote leaflets of poems against the all-powerful prime minister, cardinal Masarini. He tried to strike it rich at cards, but was unsuccessful. Cyrano was endowed by nature with a grotesque, huge nose, which was the subject of jokes and taunts and the cause of a great many duels. But this gambler and duelist was a great thinker, a follower of the philosopher Gassendi,[2] and the creator of the science fiction novel *The Other World, or the States and Empires of the Moon* (published in 1657 after the author's death).

What's interesting is that in the novel Cyrano describes his flight to the Moon . . . by means of rockets! There's an old engraving of Cyrano flying to the Moon in a basket with rockets attached. The rockets send out fiery jets that drive the basket upward. So Cyrano foresaw the use of rockets in space flight more than 300 years ago. It's also curious that Cyrano asserts in this book, written some 30 years before Newton's *Principia*, that the Moon's gravitational attraction becomes greater than the Earth's at a point nearer the Moon than the Earth because the Moon's mass is less than the Earth's. He even calculates the ratio of these distances (but, alas, got it wrong: 3 instead of about 9).

Cyrano ironically describes the Moon as Eden. He allegedly meets the prophet Elijah there and asks how he had gotten to the Moon. The prophet then describes his method of travel, which contradicts Newton's third law. He had built an iron chariot (it was believed that thunder is the rumbling of Elijah's chariot), sat in it, and began to throw a magnetized iron ball upward. The ball attracted the chariot, and each throw pulled the chariot up until it reached the Moon. The prophet failed to take into account the recoil experienced by the chariot when he throws the ball, and so he became a predecessor of the "truth-loving" Baron Münchhausen.

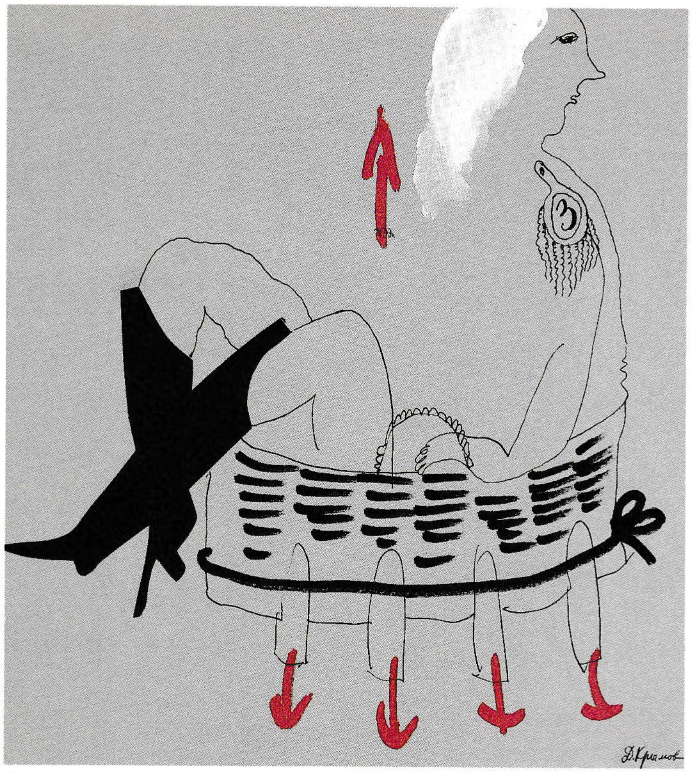
In the second half of the nineteenth century, the French dramatist Edmond Rostand wrote the play *Cyrano de Bergerac*, which continues to enjoy great success on stage and screen.[3] The author called his play a "heroic comedy," which describes it perfectly—it's funny and profound at the same time. Rostand ascribes to Cyrano a method of flying to the Moon "invented" by the prophet Elijah, but doesn't mention rockets. Inexplicably, he also changes the chariot to a piece of sheet metal.

Not so long ago a lunar crater was named Cyrano, which stands to reason.

---

[2]Pierre Gassendi (1592–1655), French philosopher and physicist, who studied mechanics, acoustics, optics, and heat.

[3]The movie "Roxanne," starring Steve Martin, was based on this play.—*Ed.*



A few words about Jules Verne

From Earth to Moon in 97 Hours, 20 Minutes is a science fiction novel written by Jules Verne in 1865. Verne describes a flight to the Moon in a shell fired by a giant cannon. We have to admit that this is a much less rational way to fly than the rocket-powered basket Cyrano dreamed up.

And this is 200 years later! True, in another novel, *Around the Moon,* one of the three travelers, Michel Ardan, takes some rockets with him, but only to soften the impact while landing. In both novels, though, Verne gives a more precise coordinate for the point at which the Earth's gravitational pull equals that of the Moon: 47/52 of the distance from the Earth to the Moon. No wonder he got it right—the relation of masses of the Earth and Moon was already known, as was Newton's law of gravity. Michel Ardan's fellow travelers—the president of the cannon club Barbicain and Captain Nicol—used algebra to perform the necessary calculations.

By the way, the state of weightlessness, according to Jules Verne, can be observed in a spaceship only at the point of zero gravity—that is, where the Earth's gravity is balanced by that of the Moon. Verne was wrong. For the state of weightlessness to appear it's not necessary that the gravitational force acting on the spaceship be zero. Leaving aside the experience of astronauts, who are weightless as soon as the engines stop firing, we know that pole-vaulters, for instance, experience weightlessness (if we ignore air resistance).

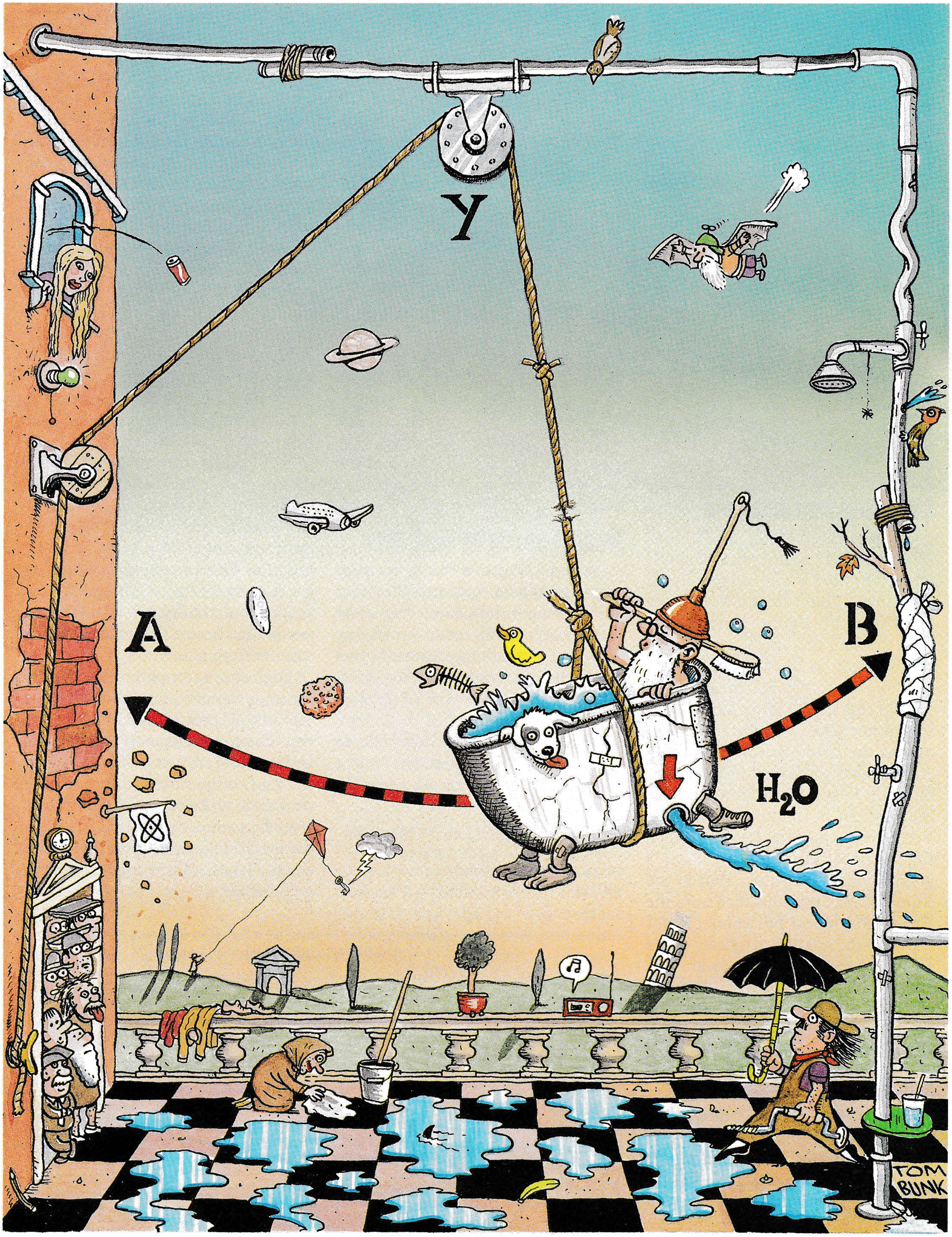
Nuclear fusion and the recoil effect

Now let's return to our century and even take a step into the next. We're going to talk about the latest technology, which is far from being perfected. I'm talking about a form of controlled thermonuclear fusion in which inertia is used to keep a plasma together.[4]

This type of fusion dates back to 1962, when the Soviet physicists N. G. Basov and O. N. Krokhin suggested that lasers be used to trigger thermonuclear fusion. The beams of many high-powered lasers are focused on a small target from all directions. The intensity of the laser beams is changed over time according to a certain formula. At first the beams cause quick evaporation of the target's surface layer. This results in a violent compression of the inner part of the target (by a factor of hundreds or thousands) because of the recoil of evaporated molecules. The compression is needed to bring the nuclei close enough together for a thermonuclear reaction. Premature heating of the inside of the target would inhibit compression, so heating occurs after compression. For the compression to be efficient the laser beams must "illuminate" the target from all sides uniformly, which is no mean feat. Only then will the "winds" arising from evaporation of the surface material take on the structure needed for extreme compression. Otherwise parts of the target that are poorly illuminated will bulge. The "Dolphin" laser device creates 216 beams that strike a target smaller than a pea.

High-speed computers and the latest achievements in nonlinear optics are used to provide automated control of the complex equipment used in laser fusion. Laser fusion is not without competitors—there are now devices that use powerful electron or ion beams. But they are all based on compression from evaporation of the surface layer and the accompanying recoil effect.

[4]Plasma is highly ionized gas, often called a "fourth state of matter" because it doesn't easily fit into the categories of solid, liquid, or gas.—*Ed.*



Y
A
B
H_2O
TOM BUNK

PHYSICS CONTEST

The leaky pendulum

"Would you like to swing on a star,
Carry moonbeams home in a jar . . .?"—*Johnny Burke*[1]

by Arthur Eisenkraft and Larry D. Kirkpatrick

THE MOTION OF SIMPLE pendulums has played an interesting role in physics and technology. Galileo is reported to have made an important discovery about the motion of a pendulum while watching the swinging of a chandelier during a church service. He discovered that the period of a simple pendulum is (almost) independent of the amplitude of its swing. This led to the use of pendulums to measure time intervals and the development of accurate clocks.

Any mass hanging from a pivot is a pendulum. An orangutan swinging by one arm from a branch is one example. Your leg pivoting from your hip as you walk is another. Both of these examples are fairly complicated because the objects aren't rigid structures and the mass is distributed in such a complicated fashion. As physics students, you should practice looking at complicated situations and finding ways to simplify them. What do you imagine would be the simplest pendulum? Certainly not the chandelier that so intrigued Galileo.

The simplest pendulum would probably be a compact mass attached to a long string. Physicists call this "the simple pendulum." If we let a simple pendulum swing through small angles, we find that its period T (the time to complete one cycle) is given by $T = 2\pi\sqrt{L/g}$, where L is the length from the pivot to the center of mass of the pendulum bob and g is the acceleration due to gravity. This is similar to the solution of many other problems in physics where objects repeat their motion over and over again. Notice that the period doesn't depend on the mass of the bob. Does this surprise you? It should! And we recommend that you plan on devoting some quiet time to wondering about the insignificance of the mass of a pendulum.

In this month's contest problem, we'll study the period of a simple pendulum that is slowly losing mass—the so-called leaky pendulum. The pendulum bob is a cubical container of negligible mass with an edge length of $2a$. It is initially filled with a fluid of mass M_0. The cube is tied to a very light string to form a simple pendulum of length L_0. The fluid flows through a small hole in the bottom of the cube at a constant rate r. At any time t the level of the fluid in the container is l and the length L of the pendulum is measured relative to the instantaneous center of mass.

Part A: Find the period as a function of time for small angular displacements.

Part B: Sketch a graph of the period as a function of time, being sure to label the endpoints of the graph.

Part C: How do your answers change if the mass of the container is also M_0 (the same as the initial mass of the fluid) and the center of mass of the container is located at its geometrical center?

Please send your solutions to *Quantum*, 3041 North Washington Boulevard, Arlington, VA 22201. The best solutions will be acknowledged in *Quantum* and their creators will receive free subscriptions for one year.

How the ball bounces

In discussions with some of our readers, we discovered that many of you are successfully solving these contest problems. We then discovered that we haven't been getting responses because the best problem solvers aren't necessarily the best letter writers! Well, that's fine. In order to challenge all of our readers, we'll design our contest problems to cover a range of skills. In other words, watch out for future problems.

The best solutions to the Contest problem in the March/April issue were submitted by Joseph Burke (Massena, NY), Kiran Kedlaya (Silver Spring, MD), Sam Prytulak (Vancouver, BC), and Noam Shomron (Budd Lake, NY).

When the racquetball hits the wall, it undergoes a totally elastic, frictionless collision. This means that the vertical component of its velocity remains unchanged and its horizontal component is reversed in direction. Since this is just like viewing the motion in a vertical mirror,

CONTINUED ON PAGE 31

[1]From "Swinging on a Star," music by Jimmy Van Heusen, which was the Academy Award–winning song in 1944 (sung by Bing Crosby in *Going My Way*).

Art by Tomas Bunk

MATH INVESTIGATIONS

Differing differences

And subsetted sets

by George Berzsenyi

SOME YEARS AGO BASIL Rennie, the editor of *JCMN* (*James Cook Mathematical Notes*), sent me a collection of problems he had compiled as the Creator of Problems for the 1981 Australian Mathematics Olympiad. Among them I found the following problem, devised by C. J. Smyth of James Cook University of North Queensland, whom I also met during my first visit to Australia: "How large a subset of the integers 1, 2, ..., 1981 can be chosen so that no two in the subset differ by either 2 or 3?" In my search for problems to be proposed in the AIME (American Invitational Mathematics Examination), I managed to update Smyth's challenge and posed the following problem as Problem 13 in the 1989 AIME: "Let S be a subset of {1, 2, 3, ..., 1989} such that no two members of S differ by 4 or 7. What is the largest number of elements S can have?" I recommend that you solve these two problems before attempting the more general problem: "Let k, m, and n be positive integers, $S \subset \{1,2,\ldots,n\}$. What is the largest number of elements S can have if no two members of it differ by m or k?"

I'll sketch the solution of yet another special case, $m = 6$, $k = 11$, and $n = 1991$ (what else?!) below—thereby, perhaps, providing a mode of attack on the general situation.

Partition $T = \{1, 2, ..., 17\}$ into the subsets

$$\{1, 7\}, \{2, 13\}, \{3, 9\}, \{4, 15\}, \{5, 11\}, \{6, 17\}, \{8, 14\}, \{10, 16\}, \{12\},$$

and note that each of these can contribute at most one element to S. Moreover, in view of the chain of implications,

$$\begin{aligned} 12 \in S &\Rightarrow 1 \notin S \Rightarrow 7 \in S \Rightarrow 13 \notin S \\ &\Rightarrow 2 \in S \Rightarrow 8 \notin S \Rightarrow 14 \in S \\ &\Rightarrow 3 \notin S \Rightarrow 9 \in S \Rightarrow 15 \notin S \\ &\Rightarrow 4 \in S \Rightarrow 10 \notin S \Rightarrow 16 \in S \\ &\Rightarrow 5 \notin S \Rightarrow 11 \in S \Rightarrow 17 \notin S \\ &\Rightarrow 6 \in S \Rightarrow 12 \notin S, \end{aligned}$$

if $T \subset S$, then T can have at most 8 elements. Indeed, one can choose

$$T = \{1, 3, 5, 8, 10, 13, 15, 17\},$$

and then one can show that T retains its key property (of not having elements differing by 6 or 11) if we add to its elements multiples of 17. Since $1989 = 117 \cdot 17$, it follows that the largest subset of {1, 2, ..., 1989} with the desired properties has $117 \cdot 8 = 936$ elements. A bit more effort will yield 937 as the answer to our problem (that is, when n = 1991).

The successful reader may also wish to generalize even further and consider the folowing problem: "Let n, k, and $m_1, m_2, ..., m_k$ be positive integers, $S \subset \{1,2,\ldots,n\}$. What is the largest number of elements S can have if no two of its members differ by m_i for $1 \leq i \leq k$?"

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

Shapes and sizes

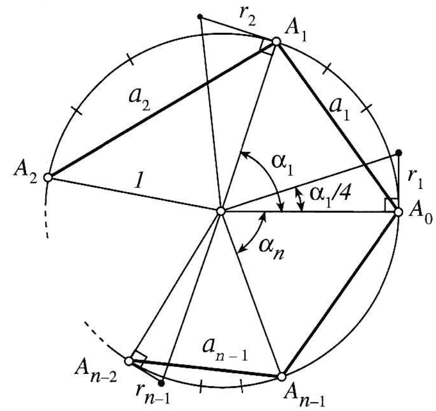
Two readers, Brian Platt from Utah and Bodo Lass from Germany, submitted wonderfully successful investigations of the problems posed in the November/December 1990 issue. Both showed that for $n > 2$, there are infinitely many polygons with n sides of integer length that can be inscribed in a circle of integer radius. Moreover, one can specify positive integers $n_1, n_2, ..., n_k$ such that $n_1 + n_2 + ... + n_k = n - 1$, and for each i, $1 \leq i \leq k$, n_i sides of the polygon are of equal length.

The figure on the facing page provides a scheme for constructing a polygon with n rational sides inscribed in a unit circle; upon multiplying through with the least common multiple of the denominators of the sides, one can obtain from it the desired polygon. The lengths r_1, r_2, ..., r_{n-1} are arbitrary positive rational numbers such that

$$\tan^{-1} r_1 + \tan^{-1} r_2 + \cdots + \tan^{-1} r_{n-1} < \frac{\pi}{2},$$

and for each i, $1 \le i \le n-1$, $\alpha_i = 4 \tan^{-1} r_i$.

I'll leave it to you to show that for $i = 1, 2, \ldots, n-1$, $a_i = 4r_i/(1 + r_i^2)$, $\sin(\alpha_i/2) = 2r_i/(1 + r_i^2)$, and $\cos(\alpha_i/2) = (1 - r_i^2)/(1 + r_i^2)$; that is, all these numbers are rational. Finally, one must also verify that



$$\begin{aligned} a_n &= 2\sin(\alpha_n/2) \\ &= 2\sin\frac{1}{2}\left[2\pi - (\alpha_1 + \alpha_2 + \cdots + \alpha_{n-1})\right] \\ &= 2\sin\left[(\alpha_1/2) + (\alpha_2/2) + \cdots + (\alpha_{n-1}/2)\right] \end{aligned}$$

is a polynomial in $\sin(\alpha_1/2)$, $\sin(\alpha_2/2)$, ..., $\sin(\alpha_{n-1}/2)$, $\cos(\alpha_1/2)$, $\cos(\alpha_2/2)$, ..., $\cos(\alpha_{n-1}/2)$, and so is also rational.

Platt and Lass also noted that one could replace the circle in the problem with a semicircle. Congratulations to them for their excellent work, for which they will receive an interesting book prize.

"LEAKY PENDULUM"
CONTINUED FROM PAGE 29

we can ignore the wall until the last step.

For the vertical component of the motion, we have

$$y = y_0 + v_{0y}t + \tfrac{1}{2}a_y t^2,$$

with the usual definitions for the symbols and $v_{0y} = v \sin\theta$. For our problem, we have

$$0 = H + v_{0y}t - \tfrac{1}{2}gt^2.$$

We can solve this quadratic equation to find the time for the ball to reach the floor:

$$t = \frac{v_{0y} \pm \sqrt{v_{0y}^2 + 2gH}}{g}.$$

Substituting the given values, we get

$$t = 1.08 \text{ s}, -0.342 \text{ s}.$$

Since we are interested in positive times, we choose the first solution. (You should ask yourself the meaning of the other solution.)

During this time the ball travels the horizontal distance

$$R = v_x t = v \cos\theta = 5.17 \text{ m}.$$

Since the ball must travel 2.25 m to reach the wall, it will land 2.92 m from the front wall.

KALEIDO

Ornamenta

Sa ie

Harmony with alg
—Alexander Pushkin,

by Vladimir D

THIS DEPARTMENT, AS ITS name implies, usually contains a motley collection of facts strung on the thread of some math or physics idea. The approach taken in this edition of the Kaleidoscope will be less "kaleidoscopic," yet its topic has a lot to do with kaleidoscopes and with groups, so it fits perfectly right here—the Kaleidoscope pages in this "group issue" of *Quantum*.

The beautiful plates that grace these pages are the work of the famous Dutch artist—a favorite of mathematicians everywhere—Maurits Cornelis Escher, whose imagination was greatly stimulated by kaleidoscope patterns.

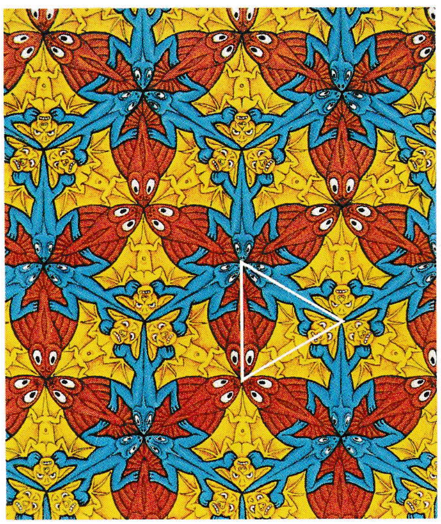


Figure 1

Figure 1 presents a picture very similar to what a real kaleidoscope might show (you must imagine all the patterns here to be extended infinitely). It consists of repeated triangular patterns, and we can regard any one of them (say, the one in the box I've drawn) as if created by pieces of colored glass in a kaleidoscope tube framed by three mirrors, making a regular trihedral prism. When you look into the prism, you see the main triangular pattern—the motif—together with all its multiple reflections. Because the triangle is equilateral, its images tile the whole plane without gaps or overlaps. In much the same way all of Escher's pattern is generated by reflections of any of its triangular fragments, called *unit*

Figure 2

cells, or just *cells*, through its sides. We can change the pattern in the cell—like rearranging the glass pieces by turning the kaleidoscope. Then the entire picture will also change, but its abstract structure—the order and disposition of cell patterns, determined by laws of reflection—naturally remains the same. But when the mirrors are moved a little, separate images of the motif shift, overlap with each other, and the whole picture diffuses and gets blurred. There are only a few positions of the mirrors that by multiple reflection produce a discernible pattern of recurrent motifs like in figure 1. Another pattern of this sort is seen in figure 2: here Escher made use of a "kaleidoscope" with four mirrors forming a square cell.

Exercise 1. Find two other essentially different kaleidoscopes.

The next steps on the road to generalization are the periodic patterns in which the unit cell is reproduced by transformations other than line reflections: *translations*, *rotations*, and *glide-reflections* (which are re-

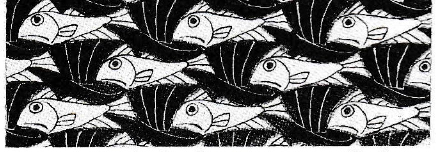


Figure 3

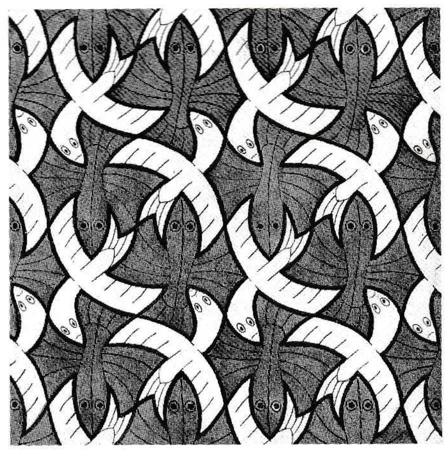


Figure 4

flections combined with translations along the mirror axes). Inspecting figures 1 and 2, you'll find that every cell can be mapped onto many others just by translation or, say, rotation (see exercise 2 below). But here these types of isometries (distance-preserving transformations) arise as a net effect of a series of reflections. On the other hand, in figures 3 through 7 there are no reflections at all. Each of these patterns, however, is composed of images of one cell undergoing rotation and translation. In figure 7, for example, the triangular cell is printed more brightly. Applying all the different kinds of 1/6 turns about its left vertex and 1/3 turns about its upper vertex, we'll tile the plane with its images and get the entire pattern (disregarding the colors). So this is a sort

tal groups

alieri:
I tested
algebra.
kin, Mozart and Salieri

ir Dubrovsky

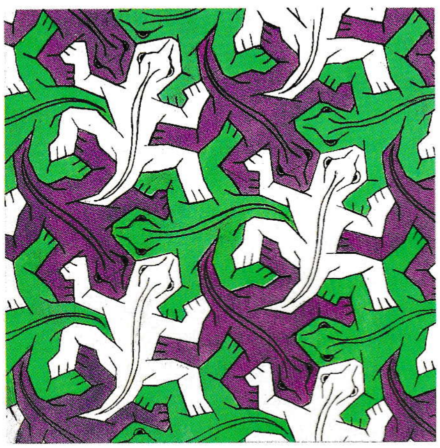


Figure 5

of "rotational kaleidoscope." This figure also shows that a unit cell isn't defined uniquely; we could use, say, the silhouette of a lizard.

Exercise 2. Choose some unit cell in each of figures 1 through 7 and describe the transformations that take this cell into the others. Verify that these transformations can always be reduced to one of the three types of isometries listed above (line reflections are a particular case of glide-reflections). Verify that the transformations you've found map the entire pattern onto itself—that is, they're the symmetries of the patterns.

We can see at a glance that each of our (or rather, Escher's) patterns has its own set of symmetries, differing from those of the others. For example, the three patterns in figures 1, 5, and 7 can be fitted onto themselves by 120° turns, but only the first has line reflections, and only the last has half-turns. It's the set of symmetries that determines the abstract structure of a pattern—that constitutes its framework. And such structures are quite numerous, as one can judge by the designs on mosques, for example. Intricate periodic patterns are especially frequent in the art of Muslim countries because the Koran forbids depictions of people and animals. My goal here is to give, if not a complete, then at least a partial description of these structures. Why is it so important to classify them? The reason is more scientific than artistic: atoms in a crystal form a pattern much like those you see in the illustrations here, though not as elaborate. So in a certain sense, the classification of crystalline forms amounts to the classification of regular periodic systems of points. (You can find out more about crystals and regular systems in the article by R. Galiulin in the January/February 1991 issue.)

Figure 6

And here groups enter the scene. You may have already learned from Alexey Sosinsky's article "Marching Orders" (see page 6)—which is very helpful for what follows—that all the symmetries of any figure constitute its symmetry group. We're interested in the symmetry groups of infinite periodic patterns. Such groups must satisfy two natural conditions:

P (*periodicity*). The group contains two translations, t_1 and t_2, in different directions.

D (*discreteness*). The group contains the *shortest translation* t_0, which is the nonidentical translation by the vector whose length does not exceed the length of the vector of any other nonidentical translation of the group.

Maybe the second condition needs a little clarification. If it doesn't hold, and you can find an arbitrarily short translation in the group, then however small your cell is its images will overlap, so you'll be unable to distinguish them.

Exercise 3. Find the translations t_1, t_2, t_0 for the patterns in figures 1 through 7.

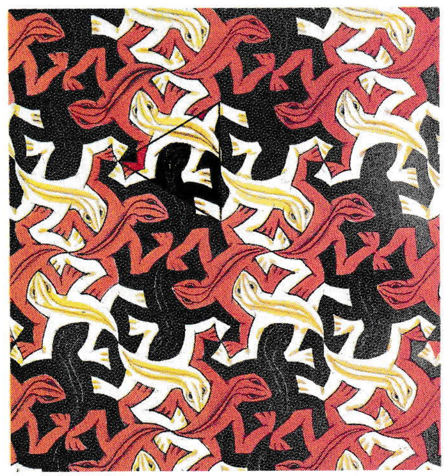


Figure 7

The groups of isometries satisfying conditions **P** and **D** are called *crystallographic groups*. They've been actively investigated in the last century. The prominent French mathematician C. Jordan found 16 plane crystallographic groups (in 1869); L. Sohncke in Germany discovered the missing one but missed three others (1874). The first full list of all 17 groups was given exactly 100 years ago by the outstanding Russian crystallographer and geometer E. S. Fyodorov. A year earlier, in 1890, Fyodorov published a classic work in which he listed all 230 spatial crystallographic groups. The patterns presented here have seven of the 17 groups as their symmetry groups. (To get the right groups in figures 6 and 7 disregard the coloring.) But five of the seven (figures 3 through 7) constitute the complete list of those of the 17 groups that consist only of translations and rotations. (I'll sketch the proof below.) These five groups form the core, in a sense, of all the 17 groups, since any of the other 12 can be obtained by adding to one of the five groups a proper (glide-)reflection and the compositions of this reflection with all the translations and rotations of the original group. For instance, the symmetry group of the pattern in figure 2 is obtained in this way by adding to the group of figure 4 a reflection in the line joining any two neighboring centers of symmetry.

I imagine many of you would like to draw your own recurrent patterns. To construct such a pattern, it suffices to choose a unit cell and generators (transformations) that will reproduce it. Then draw whatever you like inside the cell and tile the plane with the images of your motif, applying the chosen transformations. You can try it yourself, but it's far from easy. The guidelines for your work are given in the exercises at the end of this article.

So, let's try to sort the crystallographic groups consisting of translations and rotations. It will be convenient, and quite legitimate, to think of them as the symmetry groups of some periodic patterns, although patterns aren't involved in the general definition of crystallographic groups. Also, we'll have to use some of the simpler elements of transformational geometry.

We'll begin with the groups comprising only translations. You'll recall that the inverse of a translation by some vector is the translation by the negative of this vector, and the composition of two translations is the translation by the sum of the vectors. It follows that, together with translations by vectors **a** and **b**, our group must contain translations by all vectors of the form $k\mathbf{a} + l\mathbf{b}$, where k and l are arbitrary integers. If we represent these vectors by arrows radiating from some origin O, then their tips will make a grid like the one in figure 8. Let's show that *vectors* **a** *and* **b** *can always be chosen so that the vector of every translation of our group is of the form* $k\mathbf{a} + l\mathbf{b}$ *with some integers* k *and* l.

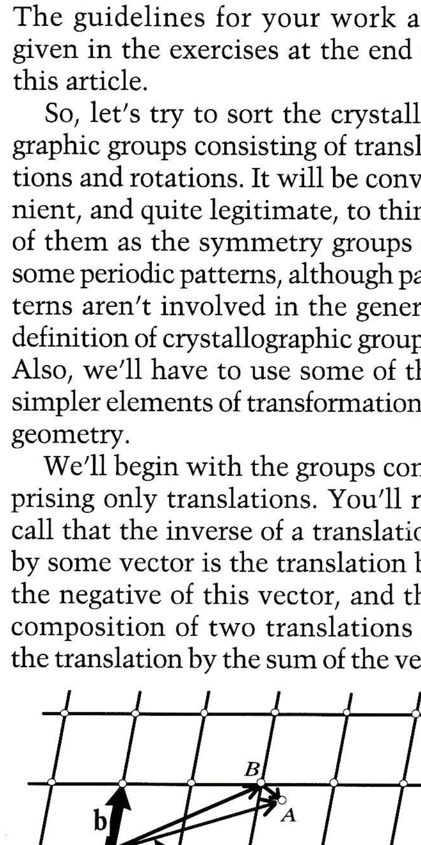


Figure 8

Figure 8 illustrates the proof when the group contains not just one (as condition **D** requires) but two shortest translations in different directions. Let **a** and **b** be their vectors. The corresponding grid, shown in the figure, consists of diamonds. If the group contains a translation by vector **OA**, whose tip A doesn't fall on a node of the grid, then point A must lie inside one of the diamonds, and the distance from A to at least one of the vertices B of this diamond must be less than the length of its side. Translations by **OB**, **OA**, and, therefore, **BA** = **OA** – **OB** belong to our group. This leads to a contradiction, because the length of **BA** is less than that of the shortest vector **a**.

In the general case, we can take for **a** the shortest vector and for **b** the vector whose projection onto the axis perpendicular to **a** has the smallest nonzero length among all the vectors of translations in our group. (I'll omit the details.)

The next step is to find out what the set V of the vectors of all the shortest translations looks like. Notice that the negative of any vector **a** from V also belongs to V (it has the same length as **a** and corresponds to the inverse translation). Also, the angle between any two different vectors from V is not less than 60° (otherwise the group would contain a translation—the one by the difference of these vectors—that is shorter than the shortest possible). It follows immediately that there are only four possible cases for set V, shown in figure 9.

Now we're able to say what values the angles of rotations belonging to the crystallographic group can take.

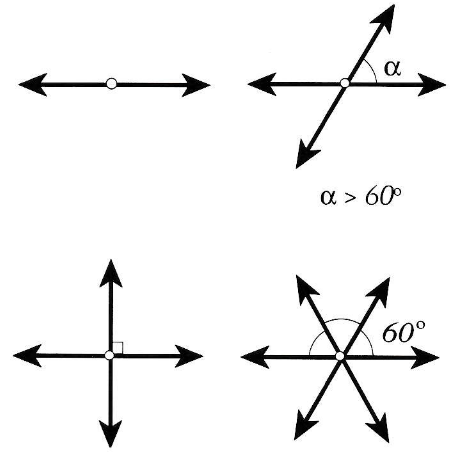


Figure 9

Consider one of these rotations R. It maps the corresponding pattern onto itself. The translation by any vector **v** of set V also takes the pattern onto itself. Clearly the pattern rotated by R is mapped onto itself by vector **v** rotated by R—that is, $R(\mathbf{v})$. But the rotated pattern coincides with the initial one! So the symmetry group of the pattern contains $R(\mathbf{v})$ whenever it contains **v**. But the length of $R(\mathbf{v})$ equals that of **v**. Therefore, *set* V *stays invariant under any rotation from the group.* (When a vector is rotated, only the angle of rotation matters, not the center.) From this we can infer, for example, that a crystallographic group never contains a rotation by 360°/5 = 72° or, say, by 15°, since neither of these rotations keeps any of the sets V in figure 9 invariant.

So what are the possible angles? Taking into account that if there is a rotation by angle α in a group, there are also its square, cube, . . . (that is, rotations by 2α, 3α, ...), we come up with the following sets of rotation angles:

(1) 0°; (2) 0°, 180°; (3) 0°, 120°, 240°; (4) 0°, 90°, 180°, 270°; (5) 0°, 60°, 120°, 180°, 240°, 300°.

Sets (1) and (2) go with all four cases of V in figure 9; (3) and (5) go with the bottom right V; (4) goes with the bottom left V. Sets (1) through (5) correspond to the symmetry groups of patterns in figures 3 through 7, respectively. The standard notation for these groups is **p1**, **p2**, **p3**, **p4**, **p6** (the number corresponds to the maximal order of rotation centers for the respective groups; for example, the center of rotations by $360°k/n$, $k = 0, 1, ..., n - 1$, has the nth order).

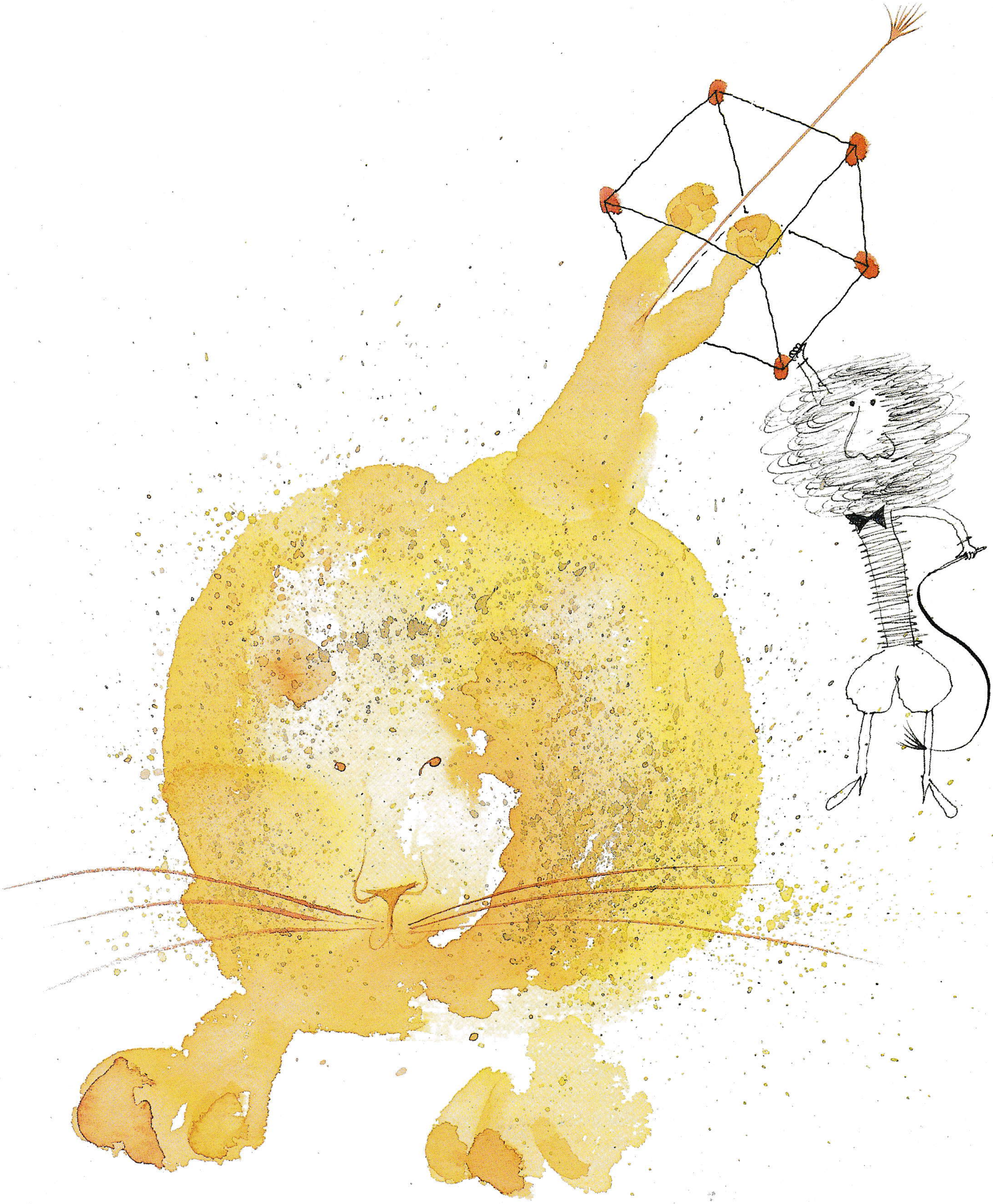
To be fair, we must also prove that a set of rotation angles determines the group uniquely (or rather, to the point of isomorphism, to use the term explained in Sosinsky's article). The main step of the proof is to show that every rotation from the group can be represented as a composition of the rotation by the same angle about some *fixed* center and one of the translations from the group. So the entire group is completely determined by the basic translations (by **a** and **b**), the maximal order of rotation centers, and any of the centers of this order.

Now we can extend the five groups by reflections, as was explained above. You should keep in mind, though, that in order to retain the set of translations and rotations, the added reflection must map the grid of rotation centers onto itself, so the image of any center must coincide with another center of the same order, and so on. The extensions of groups **p1** and **p2** are described in exercise 4 below (**p1** has three extensions and **p2** has four; one of the latter is the "kaleidoscope" group of figure 2). Groups **p3** and **p4** have two extensions each; **p6**, only one. One of the extensions of each of these three groups is again "kaleidoscopic": for **p3** it's the group in figure 1; for **p4** and **p6** they are generated by reflections in the three sides of the right isosceles triangle and the right triangle with an angle of 60°, respectively (and this is the answer to exercise 1). So we have $17 - 5 - 7 - 3 = 2$ groups left; they're described in exercise 5.

Exercise 4. Draw periodic patterns based on a rectangular cell $ABCD$ for the following sets of generating transformations (we'll use the notation t_{XY}, r_{XY}, and gr_{XY} to signify translation by XY, reflection in line XY, and glide-reflection $t_{XY} \cdot r_{XY}$, respectively; K, L, M, N are the midpoints of sides AB, CD, AD, BC of the rectangle): (1) t_{AB}, r_{AB}, r_{CD}; (2) t_{AC}, r_{AB}; (3) gr_{AK}, gr_{DL}; (4) t_{AB}, r_{AB}, gr_{AD}; (5) gr_{KL}, gr_{MN}; (6) r_{AB}, r_{AD}, gr_{DC}.

Exercise 5. Find all the transformations generated by a line reflection and a rotation (a) by 90°, (b) by 120°. Choose a suitable cell and draw the corresponding patterns.

Once you've done the exercises, you'll have discovered all the plane crystallographic groups. To complete the proof of Fyodorov's theorem, you just have to show that all the possible extensions of groups **p1**, **p2**, **p3**, **p4**, and **p6** have indeed been exhausted. ◙



Patterns of predictability

Symmetry, anisotropy, and Ohm's law

by S. N. Lykov and D. A. Parshin

EVERYBODY HAS SOME intuition about symmetry—so many things in nature and art are arranged symmetrically: the petals of a flower, patterns in a piece of jewelry, atoms in a lattice, and so on. The very notion of beauty is connected with symmetry. Symmetry is inherent not only in art and architecture but in mathematics and physics as well. In fact, it plays an exceedingly useful role in the last two areas mentioned. Sometimes considerations of symmetry are all it takes to solve extremely difficult physics problems. We'll try to convince you of this by way of a few examples from solid-state physics.

Imagine we face the task of determining the electrical conductivity[1] of a crystal (whose properties are as yet unknown). Let's cut a rectangular sample from this crystal and insert it into a circuit as shown in figure 1. Having measured the current across the sample and the voltage in it, we'll use Ohm's law to find the conductivity σ:

$$\sigma = \frac{IL}{VS}, \qquad (1)$$

where I is current, L is length, V is electrical potential difference (voltage), and S is cross-sectional area. It doesn't depend on the sample size, so it would seem unambiguously to characterize the sample's ability to conduct current. But why say "it would seem"? Can a crystal really have several values for its conductivity?

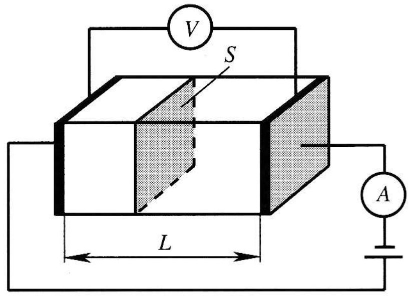


Figure 1

Yes, it can, as it turns out. And there's nothing surprising about that. Imagine, for instance, that the substance was layered: conducting layers alternate with nonconducting ones—that is, with dielectric layers. Then the measurements will depend on how the sample is cut—along or across the layers. It's clear that the current can't flow across the layers but only along them. So measurements in two directions (fig. 2) will give two answers: the "transverse" conductivity $\sigma_\perp$ (fig. 2a) is equal to zero, and the "longitudinal" conductivity $\sigma_\parallel$ is not. (An example of such a layered medium occurring in nature is the graphite crystal. Its longitudinal conductivity is 250 times that of its transverse conductivity.)

But even without layers, the conductivity of a medium can depend on the direction of the current. Many crystals possess this property. Here's why.

Figure 3 shows the arrangement of atoms in a crystal. (Natural crystals are three-dimensional, of course; but our main points can be clearly illustrated in flat drawings.) Compare the two directions shown with red and blue lines: it's obvious that they're not equivalent. Why? We clearly see that along the "red" lines the atoms are much farther apart than they are along the "blue" lines. Choose any other direction and you'll see that it's

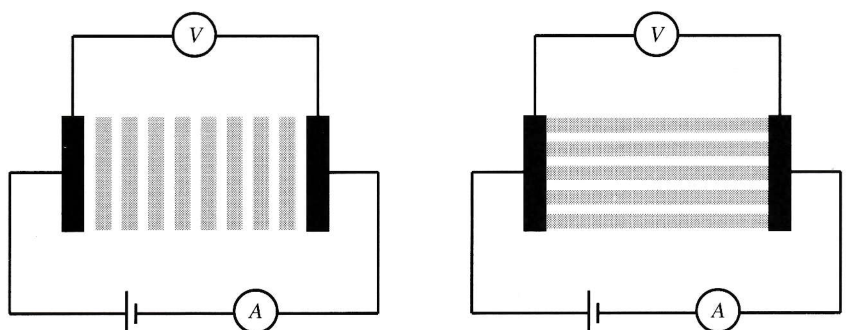


Figure 2 a b

[1]Electrical conductivity is the inverse of the resistivity—that is, $\sigma = 1/\rho$. The resistance R of a rod of the material is given by $R = \rho L/S$.

Art by Pavel Chernusky

not equivalent to the preceding one. It's natural to expect that the electrons somehow "feel" this difference—it's easier for them to move among atoms in some directions than in others. No wonder crystal conductivity can depend on the direction of the current.

Media whose properties are not equal in different directions are called anisotropic. As we now see, the cause of crystal anisotropy is the difference in directions that arises out of the regular arrangement of the atoms.

Another important aspect of anisotropy is the current's ability to flow in a direction that doesn't coincide with that of the electric field. Let's examine this more closely. From formula (1) it's clear that the current density $j = I/S$ is proportional to the electric field $E = V/L$ applied to the sample:

$$j = \sigma E. \qquad (2)$$

But the current and field in the medium are characterized not only by magnitude but by direction as well. The field vector $\vec{E}$ gives the magnitude and direction of the force $\vec{F} = e\vec{E}$ acting upon each charge e. Under the influence of this force the electrons move with an average velocity $\vec{v}$. The current density vector $\vec{j}$ (we'll call it the current vector for short) is parallel to this average velocity v of the electrons.[2] Ohm's law, as it appears in equation (2), links only the magnitude of the current j and the field applied from outside, but it says nothing about the mutual orientation of vectors $\vec{j}$ and $\vec{E}$.

"But that's obvious," you say. "The field $\vec{E}$ and force $\vec{F}$ are parallel. The average velocity of the electrons $\vec{v}$ is parallel to the effective force $\vec{F}$. And the current $\vec{j}$ is parallel to $\vec{v}$. So vectors $\vec{j}$ and $\vec{E}$ are also parallel, and we can write

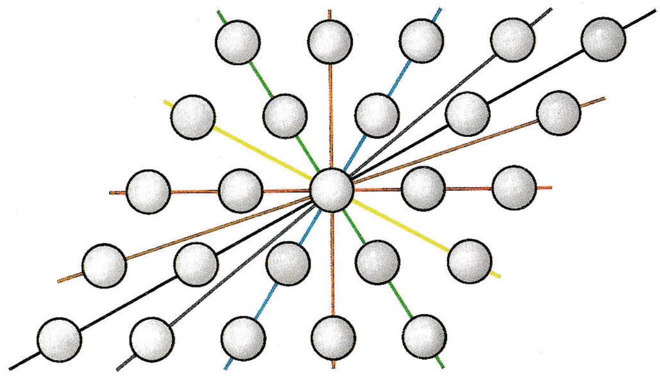


Figure 3

$$\vec{j} = \sigma\vec{E}. \qquad (3)$$

instead of (2). It's very simple!"

But is this reasoning correct? Yes, if we're talking about an isotropic medium—that is, one in which all directions are equivalent. When all directions are equivalent, the electrons don't care where they go, and as soon as we apply an external field they have no choice: vector $\vec{E}$ gives the only direction that is different. The current $\vec{j}$ will flow in that direction.

Whenever we talk about crystals, we can't lose sight of anisotropy! In any crystal there's a rich store of unequivalent directions (recall figure 3). In this case the direction $\vec{E}$ is but one of them, which stands out from the point of view of electron movement. So we can't reason the same way we did in the case of isotropy. The weak link in our reasoning is where the parallelism of vectors $\vec{v}$ and $\vec{F}$ is mentioned. In an anisotropic medium they may turn out to be nonparallel, and then vectors $\vec{j}$ and $\vec{E}$ won't be parallel either! The situation can be imagined rather vividly as follows.

An external force $\vec{F} = e\vec{E}$ drives the electrons "through the structure" of atoms, where there are "easy" and "difficult" directions of travel. Colliding with the atoms, the electron often changes its direction, and it's possible to speak of an electron moving along a straight line only on average. Passing through many interatomic distances the electron manages to move more in an "easy" direction than in a "difficult" one, and not necessarily along $\vec{F}$. Everything depends on the interactions of the electrons with the atoms and on the orientation of the field $\vec{E}$ with regard to the crystal directions that stand out the most.

The thoughtful reader is sure to ask a sly question: in our experiments (fig. 1) the field is always applied along the sample, and the current undoubtedly also flows only along the sample. How, then, could vectors $\vec{j}$ and $\vec{E}$ not be parallel?

We admit that we've concealed an important fact from you, one that eliminates this misunderstanding. We haven't mentioned that (because of anisotropy again) the field $\vec{E}$, which induces the current $\vec{j}$ in the crystal, isn't necessarily the same as the electrical field due to the voltage per unit length V/L applied along the sample. It doesn't agree in magnitude or direction. Figure 4 depicts the case of an anisotropic medium, and you can see the result of the chain of events occurring after the voltage is supplied to the sample. At this moment the field applied along the sample—vector $\vec{E}_\parallel$ in figure 4—begins to act upon the electrons. The current that arises, as was mentioned

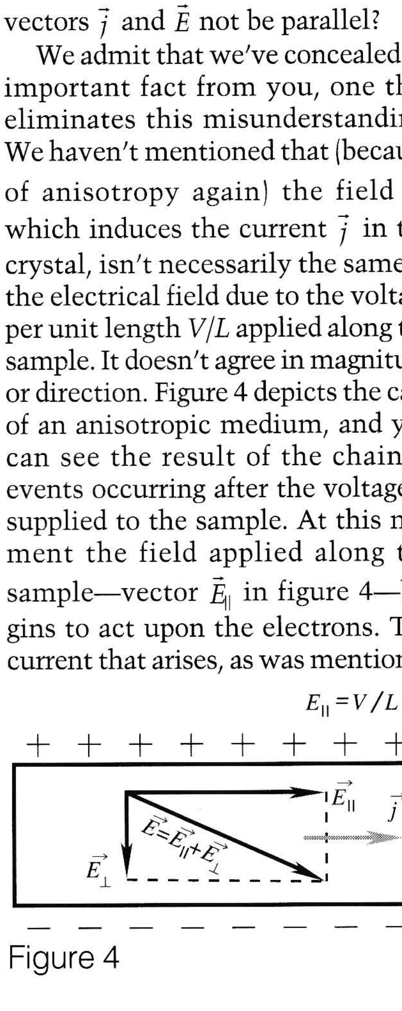


Figure 4

[2]It's easy to show that $\vec{j} = en\vec{v}$, where n is the number of electrons per unit volume moving at a velocity $\vec{v}$. The electron charge e is negative, so $\vec{j}$ and $\vec{v}$ are oriented opposite each other. (Vectors are normally designated typographically by bold face; arrows were used in this article for emphasis.—*Ed.*)

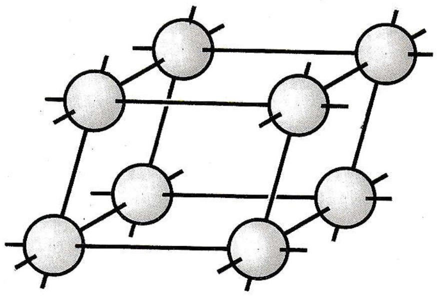


Figure 5

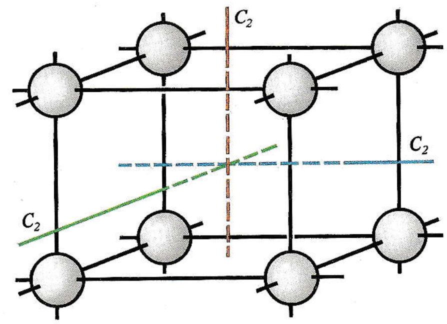


Figure 6

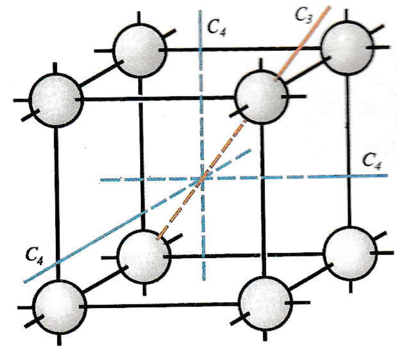


Figure 7

earlier, can be oriented differently because of anisotropy—for instance, to the right and up somewhat (in which case the electrons move left and down a bit). Transverse displacement of the electrons will cause the faces of the sample to become charged (the upper face positively, the bottom negatively), and the transverse field $\vec{E}_\perp$ will appear in the sample in addition to the applied field. But at the same time $\vec{E}_\perp$ prevents the transverse movement of electrons downward, and so it will increase only until this movement stops. All this occurs in the blink of an eye (10^{-13} s) and after that the picture doesn't change: the current $\vec{j}$ continues to flow along the sample, but the resulting field $\vec{E} = \vec{E}_\parallel + \vec{E}_\perp$ is directed at an angle to it. This is easy to verify—there should be a difference of potentials between the upper and lower faces, caused by the additional component $\vec{E}_\perp$ of the field vector $\vec{E}$. This is indeed what is detected experimentally.

So our first acquaintance with Ohm's law for anisotropic conductors shows that a crystal's conductivity varies, generally speaking, in different directions, and that the current $\vec{j}$ can flow at an angle to the electrical field $\vec{E}$.

"We've got a hard nut to crack!" some readers will think. "Do we really have to perform an infinite number of measurements for each imaginable direction?"

No, the situation is simpler than that. It'll become clear after we've put Ohm's law in concise mathematical form. And it will be quite simple for crystals with sufficient symmetry. That's what we'll talk about now.

Figure 3 shows a sample of a crystal structure in which all directions are different. Let's use different examples now—three-dimensional ones. If one isn't a talented artist, it's difficult to describe the picture that a "sightseer" would encounter inside a crystal, walking about and admiring the perspective of the geometrically arranged atoms. To make it easier, only small fragments of three different crystal structures are shown in figures 5–7.

In figure 5 the atoms are arranged at the corners of an ordinary oblique-angled parallelepiped. This crystal (as well as its two-dimensional analogue in figure 3) doesn't have any equivalent directions. But the following examples are more interesting. In figure 6 atoms are at the corners of a rectangular parallelepiped and lines, denoted by C_2, are drawn through its center perpendicular to the faces. In figure 7 we see a cube instead of a parallelepiped and analogous lines are marked C_4; also, a spatial diagonal of the cube (C_3) has been drawn. What's so remarkable about these lines?

The lines C_2, C_3, C_4 are the axes of symmetry of the crystal structures shown in figures 6 and 7. If the entire crystal is rotated about any of these axes, it's possible to obtain an arrangement of atoms that is indistinguishable from the original arrangement. One must only maintain a certain angle of rotation: 180° for axis C_2, 120° for axis C_3, 90° for axis C_4; that is, $2\pi/n$ for axis C_n. Integral multiples of these angles of rotation will also work, since they serve as repetitions of the rotations mentioned above. By the way, C_n is the commonly accepted way of designating an axis of symmetry; the number n is called its "order."

From the point of view of symmetry we can refine the notion of directional equivalency: any directions that "pass into one another" during rotation about the axes of symmetry are equivalent.

The more high-order axes the crystal possesses, the more equivalent directions it has and the stronger symmetry competes with anisotropy. You'll soon see how this competition affects conductivity, but first we advise you to find on your own some examples of equivalent and unequivalent directions in the crystal structures shown in figures 6 and 7. In particular, it's important to note that in a cubic crystal the axes C_4 are themselves equivalent directions (they pass into one another during rotations about C_3), while the analogous axes C_2 in figure 6 are not equivalent.

Well, how are we to apply considerations of symmetry to the question of conductivity in practice? Here are a few examples. Let's start with a crystal with three mutually perpendicular (unequivalent) axes of symmetry C_2 (as in figure 6). For convenience we'll denote them 1, 2, 3 (see figure 8). The question arises: where will the current $\vec{j}$ flow if the field $\vec{E}$ is directed along one of the second-order axes—say, axis 1? The answer is unexpectedly simple: the current will flow along the same axis of symmetry, parallel to the field! We'll prove it by disproving the contrary.

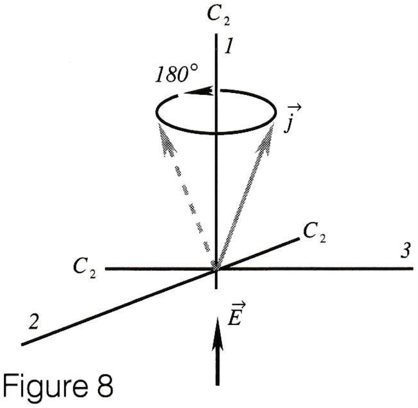


Figure 8

Suppose the current $\vec{j}$ flows at an angle to axis of symmetry 1. Rotate the crystal together with the applied field $\vec{E}$ 180° about the axis. It's pretty clear that everything inside the crystal—the "structure" of atoms and the flow of electrons penetrating it—will rotate too. And as the vector $\vec{j}$ is supposed to be not parallel to the axis of rotation, its new orientation (the broken line in figure 8) will differ from the initial one. Yet the microscopic pattern defining the direction of the current $\vec{j}$ looks the same as before: the new arrangement of the atoms cannot be distinguished from the initial arrangement, and the field $\vec{E}$ was and remains parallel to direction 1. This means that the orientation of current $\vec{j}$ must be the same as it was before.

So we've arrived at a contradiction, and the only way to avoid it is to acknowledge that vector $\vec{j}$ is oriented in the same direction as vector $\vec{E}$—that is, along the axis 1 (fig. 9).

It's worth noting that this conclusion isn't sensitive to the details of the crystal's macroscopic structure—we didn't have to explain how exactly the atoms are arranged and which directions are the easiest for electron movement. We also didn't use any other axes of symmetry. Even the order of the axis of symmetry along which field $\vec{E}$ is directed is of no importance—the reasoning will hold for any angle of rotation (not a multiple of 360°) allowed by the symmetry.

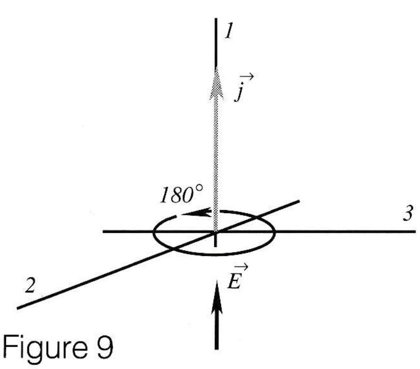


Figure 9

So if the crystal has an axis of symmetry and a field $\vec{E}$ is directed along it, then the current $\vec{j}$ flows along the field. Under these conditions the relation between $\vec{j}$ and $\vec{E}$ is given by equation (3), where the proportionality coefficient σ is the conductivity along this axis of symmetry. Taking a measurement on a sample cut along the crystal's axis of symmetry (fig. 1), we can find σ according to equation (1). Unlike the situation described in figure 4, the field will certainly not have components transverse to the current; therefore $E = V/L$, and equation (1) follows immediately from (3).

Consequently, three measurements on samples oriented along the axes of symmetry 1, 2, 3 give us three values of the conductivity: σ_1, σ_2, σ_3. Generally speaking, we can say that they'll turn out to be different, since the directions 1, 2, 3 are not equivalent:

$$\sigma_1 \neq \sigma_2 \neq \sigma_3. \qquad (4)$$

And what is the picture if field $\vec{E}$ in our crystal is not directed along the axis of symmetry? Where will the current $\vec{j}$ flow? To answer this question we'll find the projection of the vector $\vec{E}$ on the axes 1, 2, 3—that is, imagine the field as the sum $\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3$ (fig. 10). Now we'll see the contribution of each of these items to the current.

Field $\vec{E}_1$ is directed along the axis of symmetry 1 and so gives rise to a current flowing in the same direction:

$$\vec{j}_1 = \sigma_1 \vec{E}_1.$$

Similarly, the two other components of the field give rise to currents moving along the axes of symmetry 2 and 3:

$$\vec{j}_2 = \sigma_2 \vec{E}_2, \quad \vec{j}_3 = \sigma_3 \vec{E}_3.$$

These expressions determine the resulting current $\vec{j} = \vec{j}_1 + \vec{j}_2 + \vec{j}_3$:

$$\vec{j} = \sigma_1 \vec{E}_1 + \sigma_2 \vec{E}_2 + \sigma_3 \vec{E}_3. \qquad (5)$$

Because of inequality (4), the ratios of the projections of the current and field ($j_1/E_1 = \sigma_1$, and so on) in the directions 1, 2, 3 are not equal. This means that the vectors $\vec{j}$ and $\vec{E}$ are not parallel (fig. 11). This fact (not at all new to us) is directly connected with the difference in conductivity in different directions.

Here's what's so interesting. Formula (5) tells us that for any orientation of field $\vec{E}$ the magnitude and direction of current $\vec{j}$ are completely determined by the conductivity values along our crystal's axes of symmetry. We don't need innumerable measurements; all we need is to find σ_1, σ_2, σ_3. True, we first need to find the directions of the axes of symmetry themselves, but that's a different story.

Another example is a cubic crystal (fig. 7). But this is a special case of the preceding, isn't it? And the most interesting one to boot! First of all, cubic symmetry is the highest possible symmetry for crystals. Second, it occurs in many semiconductors that are of great practical use—Ge, Si, and others. Third, it will be the best example of the influence of symmetry on crystal conductivity.

So there are three mutually perpendicular axes of symmetry C_4. As before, we'll designate the conductiv-

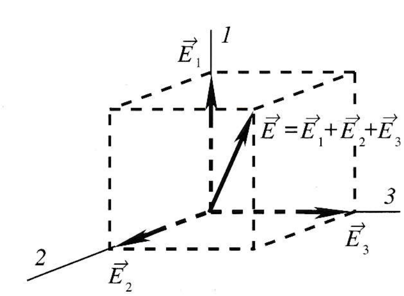


Figure 10

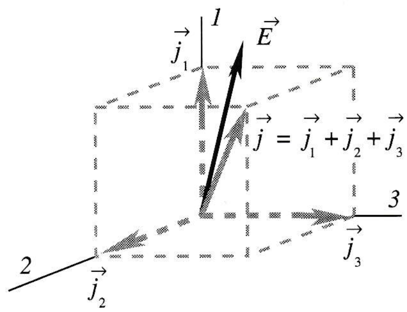


Figure 11

ity along them as σ_1, σ_2, σ_3. Then the reasoning we've used before will again lead us to equation (5), which is valid for any orientation of vector $\vec{E}$. Now let's recall that all three directions C_4 in a cubic crystal are equivalent. From this it follows, as you may have guessed, that

$$\sigma_1 = \sigma_2 = \sigma_3 = \sigma. \qquad (6)$$

Yes, symmetry guarantees that the conductivity in these three directions is one and the same. But only in these directions? Taking equation (6) into account, we can turn equation (5) into equation (3), known to us as Ohm's law for an isotropic medium:

$$\begin{aligned} \vec{j} &= \sigma\vec{E}_1 + \sigma\vec{E}_2 + \sigma\vec{E}_3 \\ &= \sigma(\vec{E}_1 + \vec{E}_2 + \vec{E}_3) \\ &= \sigma\vec{E} \end{aligned}$$

—that is, (1) regardless of the direction of the applied field $\vec{E}$ the current $\vec{j}$ is always parallel to field $\vec{E}$, and (2) its magnitude is determined only by one value of the conductivity σ. The conductivity of a cubic crystal is equal in all directions!

This result is unexpected and certainly beautiful, isn't it? But not all directions in a cubic crystal are equivalent—compare, for example, lines C_4 and C_3 in figure 7. Both are axes of symmetry, and the current is certain to be parallel to field $\vec{E}$ if it's applied along any of them. But why should the magnitude of the current be identical for both orientations of $\vec{E}$? The electrons "feel" the unequivalence of the directions, right? And this would have to be reflected in different values for the corresponding conductivities. But no! In whatever orientation the sample is cut off, measurements must give only one value of σ. This is what we've just proved, using only considerations of symmetry.

So in the given case symmetry has the upper hand over anisotropy: with respect to electrical conductivity, a cubic crystal is similar to an isotropic medium in which all directions are equivalent, as we've already mentioned.

It's interesting to compare the microscopic structures of a crystal and an isotropic medium. An example of the latter is an amorphous body, characterized by a random arrangement of atoms. Such an arrangement, you'll say, isn't symmetrical at all—any rotation (not a multiple of 360°) will give different patterns. But all the patterns will be identically random and in this sense can't be distinguished from the original pattern. Disorder equalizes (averages out) the properties of the medium in all directions. It makes all directions equivalent. Paradoxical as it may be, directional symmetry is higher in an amorphous body than in a cubic crystal! Any direction is an axis of symmetry. In light of this, Ohm's law—expressed in equation (3)—would seem to be a truism for isotropic media.

Yes, directional symmetry is a good thing. But what if there is no such symmetry at all (for example, the crystal in figure 5)? We're forced to turn to Ohm's law in its most general form. Let's try to define it.

We'll arbitrarily take three mutually perpendicular directions x, y, z (the coordinate axes). Let's break the field $\vec{E}$ down into components E_x, E_y, E_z. We'll also search for the current $\vec{j}$ as three components j_x, j_y, y_z.

What does j_x consist of? The coordinate axes aren't the same thing as the axes of symmetry! The three currents resulting from the three components of the field don't necessarily flow along the axes x, y, z. This means every current will have x-, y-, and z-components—that is, not only the x-component of the field but also the y- and z-components of vector $\vec{E}$ contribute proportionally to j_x. We'll designate the corresponding coefficients of proportionality as σ_{xx}, σ_{xy}, σ_{xz}.

And what do j_y and j_z consist of? The answer is the same, almost word for word—only the coefficients are different.

Thus:

$$\begin{cases} j_x = \sigma_{xx}E_x + \sigma_{xy}E_y + \sigma_{xz}E_z, \\ j_y = \sigma_{yx}E_x + \sigma_{yy}E_y + \sigma_{yz}E_z, \\ j_z = \sigma_{zx}E_x + \sigma_{zy}E_y + \sigma_{zz}E_z. \end{cases} \qquad (7)$$

These three equations represent Ohm's law for a medium with an arbitrary degree of anisotropy. It states: the magnitude of the current $\vec{j}$ is proportional to the magnitude of the field $\vec{E}$, but their directions may be different. Equations (7) establish a linear relation between vectors $\vec{j}$ and $\vec{E}$. We see that in the given system of coordinates this relation is written as nine "coefficients of conductivity" σ_{ik} (where $i, k = x, y, z$).

All the examples considered earlier are particular cases of (7). So if it's possible to choose axes x, y, z along axes of symmetry 1, 2, 3, we must do this. Then all the coefficients, except $\sigma_{xx} = \sigma_1$, $\sigma_{yy} = \sigma_2$, $\sigma_{zz} = \sigma_3$, will be equal to zero (no problem measuring *these*!), and equations (7) will reduce to equation (5). Also, if two of the axes are equivalent—say, 1 and 2—then $\sigma_1 = \sigma_2$ (one measurement less!). Finally, if all three axes are equivalent (cubic symmetry), we refer to equation (3) (and one measurement gives us all the information about the crystal's conductivity!). So it pays to know symmetry.

And so ends our tale of conductivity and how it's influenced by symmetry. Of course, this by no means exhausts the problems that can be solved through "considerations of symmetry." There are many other physical phenomena in which symmetry plays a crucial role. But we'll have to return to this some other time. ◘

The short, turbulent life of Évariste Galois

A revolutionary in politics and mathematics

by Y. P. Solovyov

"IN THE SPRING OF 1832 Paris was boiling and ready for a revolutionary explosion, although for three months without reprieve cholera had stultified the minds and cast a gloomy calm over the agitated feelings of the populace. The great city was like a loaded cannon that lacked but a single spark to set it off." These lines are from *The Outcasts* by Victor Hugo, who was a witness to the shattering events that followed. On Tuesday, June 5, a great rebellion broke out in Paris. Only a few people were likely to pay attention to a short paragraph in the Parisian newspapers on those alarming days. On the morning of May 30, the reports stated, Évariste Galois, a young man of 20 years, famous for his political speeches and a graduate of the Collège Royal de Louis-le-Grand, was killed in a duel. On Saturday, June 2, he was buried in a common grave at Montparnasse cemetery. Today there is no trace of this burial place.

Sixty pages of an incomplete mathematical manuscript were left after his death. They fell into the hands of Évariste's friend Auguste Chevalier, but he could find no one who would agree to publish the paper. It wasn't until 1846 that the manuscript was printed. The theory expounded in the paper has influenced not only mathematics but all the natural sciences for 145 years now . . .

This is the only existing portrait of Évariste Galois drawn from life. He was 15 or 16 years old at the time.

Life and death

"Nitens lux, horrenda procella, tenebris aeternis involuta."[1]

Évariste Galois was born on October 26, 1811, in the ancient town of Bourg-la-Reine, about ten kilometers from Paris. His father, Nicholas-Gabriel, was the head of an educational institution for young men. In 1815 he was elected mayor of the town and held that post until his death. During the first twelve years of his life, Évariste was brought up and educated by his mother. The boy studied Greek and Latin and spent much of his time reading the works of Plutarch and Livy.

In October 1823 Évariste entered the Collège Royal de Louis-le-Grand (now the Lycée Louis-le-Grand), a well-known educational institution where Molière, Victor Hugo, Robespierre, and Delacroix were students. Galois received a scholarship and boarded at the Collège. For the first three years he was regarded as one of the best students; he studied languages, literature, and history with pleasure. In October 1825 Galois began attending the senior course at the school—the rhetoric class; but he began to show signs of exhaustion, so at the director's recommendation he repeated the course in January 1827. Again he became one of the best students without great effort, and he received an award for his Greek translations as well as honors in four other subjects. These years were a turning point for Évariste: he had discovered and entered the world of mathematics.

You see, before taking the rhetoric class all the students of the Collège were taught one and the same curriculum, including mostly humanities and only the rudimentary basics of the exact sciences. But students interested in the exact sciences could attend a preparatory mathematics class during the last two years of their

[1] "Dazzling lights, a terrible storm surrounded by eternal pitch-darkness" (Latin). These were the words with which Évariste Galois ended one of the letters he wrote on the eve of his duel.

studies. Those who wanted to devote themselves to mathematics then had to attend a year-long basic mathematics class and spend another year studying a mathematical specialty. As a student repeating a year of studies, Galois took the opportunity to enter the preparatory mathematics class. His extraordinary mathematical abilities were revealed almost at once: without any difficulty he comprehended the rather complicated book *The Principles of Geometry* by Adrien Marie Legendre and began studying the works of Joseph Louis Lagrange: *The Solution of Numerical Equations*, *The Theory of Analytic Functions*, and *Lectures on Function Theory*.

In the autumn of 1827 Évariste came back to the rhetoric class and continued his studies in the preparatory mathematics class. The school routine was a burden to him; he was fully absorbed in mathematics. One of Galois's teachers said, "A great passion for mathematics possesses him; I think it would be better for him to study only this science, if his parents agree: as a student in the rhetoric class he is wasting his time, annoys his teachers, and incurs anger and punishment."

At this time Évariste became acquainted with the works of Gauss and Abel and felt himself capable of doing even more. He was only a student in the preparatory class, but without any help he was preparing for the entrance exams of the Polytechnical School—the best institution of higher learning in France at the time. Évariste believed that all the powers and energy of his mind would be put to full use at this school.

His attempt to enter the Polytechnical School failed. The failure greatly distressed Galois, and according to Paul Dupuy, a historian of mathematics, it was "the first of the injustices that eventually poisoned Évariste's life." Évariste had to return to the Collège he had grown sick and tired of; and, skipping the basic course, he took the specialized mathematics class. At that time the course was taught by Louis-Paul Richard, a remarkable teacher who loved his science with a passion. In addition to Galois, the famous astronomer Leverrier and the outstanding mathematician Hermite studied under him at various times.

Richard was very attentive to his young student Galois, whom he considered the most gifted of all. His remarks about him were laconic: "Galois deals only with the highest domains of mathematics," "He is much more talented than his fellow students." Under Richard's supervision Évariste wrote his first scientific work, "The Proof of a Theorem about Periodic Continued Fractions," which was published in *Les annales de mathematiques pures et appliquées* in March 1829. At that time, under the influence of Joseph Lagrange, Galois began to seriously study one of the most difficult mathematical problems of his time: solving algebraic equations in radicals.

This problem has a long history. It was the Babylonians who discovered the method of solving a second-degree equation $ax^2 + bx + c = 0$. In modern notation, its roots can be expressed by the formula

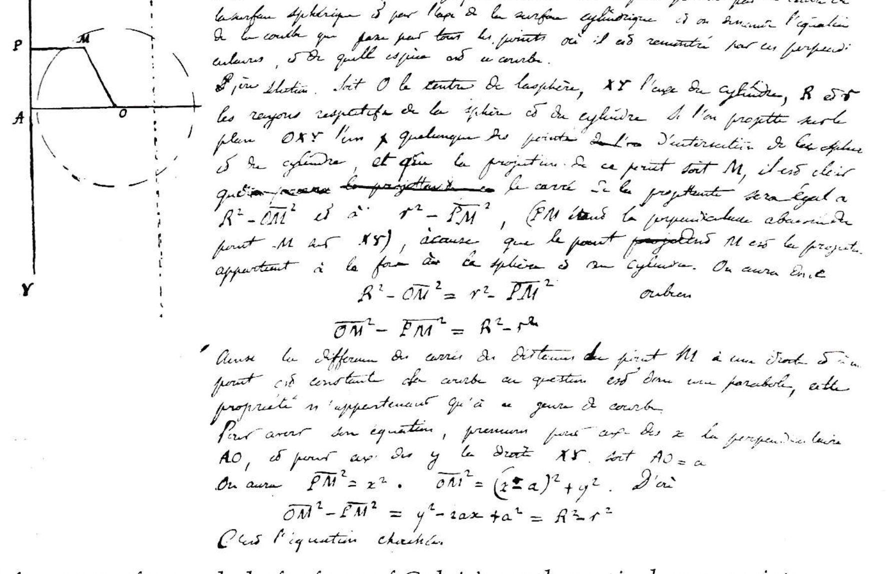
$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

which entails four arithmetic operations on coefficients and also a square radical. At the beginning of the 16th century Scipione del Ferro and Niccolo Fontana (known as Tartaglia) obtained a formula for the roots of a third-degree equation that included four arithmetic operations as well as square and cubic radicals. Some time later Lodovico Ferrari discovered a formula for the roots of a fourth-degree equation that involves radicals of not more than the fourth degree. It was only natural to expect that the roots of an algebraic equation of degree n

$$a_0 x^n + a_1 x^{n-1} + \dots + a_n = 0 \quad (a \neq 0)$$

must be expressed in terms of radicals of at most the nth degree. But despite tremendous efforts by the most prominent mathematicians over the course of three centuries, nobody succeeded in obtaining such a formula even for fifth-degree equations. By the end of the 18th century mathematicians began to suspect that formulas in radicals for equations of degree $n \geq 5$ simply didn't exist, which is why no one ever found any.

An important step in investigating algebraic equations was made by Lagrange, who discovered that the



A fragment of a rough draft of one of Galois's mathematical manuscripts.

A portrait of Galois painted by his brother Alfred from memory and published in 1848 in Magasin pittoresque. *"This portrait," says a note in the magazine, "reproduces as exactly as possible the expression of Galois's face. It was drawn by Alfred Galois, who has been creating a veritable cult in his unfortunate brother's memory for 16 years now."*

solution of equations in radicals was closely connected with permutations of their roots. Lagrange's idea, which he called "the true philosophy of solving equations," was substantially developed by another genius of mathematics, the Norwegian Niels Henric Abel. In 1824, at the age of 20, Abel proved that there were no formulas for solving algebraic equations of degree $n \geq 5$ in general form by means of radicals.

After Abel's theorem appeared, another problem emerged: to find the necessary and sufficient condition for the coefficients $a_0, a_1, ..., a_n$ of any equation that would allow it to be solved in radicals.

Évariste completely solved this extremely difficult problem in the years 1829–31. Already in the spring of 1829 he had obtained his first results in the theory of equations. Galois sent them to the Academy of Sciences. One of the most distinguished French mathematicians, Augustin Cauchy, was supposed to examine Galois's work, but he misplaced it!

After finishing his specialized mathematics class, Galois again tried to enter the Polytechnical School, and again he failed. What happened at the exams nobody knows for sure. Recalling that incident later, Galois described the "insane laughter of the examiners" that punctuated his performance. His examiners were Binet and Lefebure de Fourcy. We don't know how they graded Galois; in any case, he didn't get into the Polytechnical School.

While Évariste was preparing for his entrance exams, a great misfortune befell him. On July 2, 1829, his father, hounded by the local curé and Jesuits, committed suicide. Évariste spent those difficult days at home with his mother and younger brother Alfred.

Following Richard's advice, Évariste decided to enter the Preparatory School, a pitiful vestige of the formerly glorious Teachers School (École Normale), established during the French revolution. In 1822 the Bourbons closed the Teachers School; it was reopened in 1826 and renamed the Preparatory School as the continuation of the Collège Louis-le-Grand. The three-year course of study at the Preparatory School was meant to train teachers and civil servants. On February 20, 1830, Évariste Galois became one of its students.

The first year at the Preparatory School turned out to be the most successful in Galois's life. It was there that he got acquainted with Auguste Chevalier, and they soon became fast friends. Galois studied mathematics with enthusiasm. He wrote three works and submitted them to the Academy for a competition.

Another blow struck him quite unexpectedly. Galois's manuscript fell into the hands of the secretary of the Academy Fourier, who dropped dead soon after. Galois's work was not found among Fourier's papers—like the first one it had disappeared, and with it all hope of winning the first prize in mathematics. True, Évariste had kept copies of the works he had sent and managed to publish them in the *Bulletin of Mathematical Sciences* in May and June, but that was a poor consolation. He couldn't believe his recurrent misfortunes were mere accidents. He concluded that they were the result of a faulty organization of society, one that doomed talent to endless hardship while mediocrity prospered. With all the ardor of youth, Évariste entered into the struggle for the political reconstruction of society.

In July 1830 the dark clouds that had been gathering over the Bourbon regime burst in a revolutionary storm—King Charles X was deposed. Galois fully sympathized with the republicans, taking an active part in the work of revolutionary circles and joining the People's Friends Society. But the expectations of the republicans failed to materialize: a protégé of big business, "the bourgeois king" Louis-Philippe, came to power. The unrest in Paris wasn't over yet.

In the autumn of 1830 Évariste published a sharp attack on the double-dealing of Gineot, director of the Preparatory School, during the July events. He was expelled on December 9 as a result. His hopes of a career in mathematics were dashed. Évariste joined the artillery of the National Guard, which consisted mostly of members of the People's Friends Society. This was the army of the revolution, and the Louis-Philippe government moved quickly to dissolve it. Évariste was left without any means of subsistence—only private lessons allowed him to make ends meet.

At that time his mind was preoccupied with revolutionary ideas—mathematics took second place. Still he managed to find the time and energy to send the manuscript that had been lost the previous year to the Academy. It was to be reviewed by two academicians, Lacroix and Poisson. After dragging things out for a long time they finally informed Galois that they couldn't give the manuscript a positive assessment.

In June 1831 Galois was brought to trial for "provoking an attempt on the life of the French king by a public statement at a meeting." The jury found him not guilty, but he fell under the surveillance of the secret police.

On July 14, 1831, Galois took part in a demonstration. Its participants were protesting the prohibition of street demonstrations by the Louis-Philippe government.

Évariste was in the front lines, wearing his National Guard uniform and carrying a rifle. As soon as the demonstrators appeared in the center of the city, on the Pont Neuf, they were surrounded by the police. Galois was arrested along with the other demonstrators and ended up in the Sainte Pélagie prison in Paris. This time he was sentenced to nine months' imprisonment for illegally wearing a military uniform and carrying a weapon.

In the Sainte Pélagie prison Galois turned 20, and it was there that he finished writing his fundamental mathematical treatise. On March 16, 1832, Galois was transferred to the prison hospital—it was suspected that he had cholera. In the hospital he met a young woman who was to play a fateful role in his life. She was Stéphanie-Felicie Dumotel, the daughter of the prison doctor.

On April 29, 1832, Galois's term was up. There is reason to believe he stayed at the hospital a while longer. No traces of his life during May 1832 have survived. Nothing is known of his further relations with Stéphanie Dumotel. All we know is that Stéphanie was the cause of an argument between Galois and two of her friends, and that it resulted in a duel. Évariste wrote in one of his last letters: "I am dying. I'm the victim of a wicked coquette and two fools who are devoted to her." The conditions of the duel are unclear; we don't even know for certain who his antagonist was.

On the morning of May 30 a passerby found him badly wounded after a pistol duel near the bank of Glassier pond in the Parisian suburb of Gentilly. The wounded man was taken to the Cochin Hospital, where he died at 10 A.M. the next day with his younger brother at his bedside.

Before he died he wrote a letter to his friend Auguste Chevalier, asking him to show his manuscript to the German mathematicians Jacobi and Gauss. But it wasn't published for fourteen years, and even then it went practically unnoticed. Galois's ideas weren't fully appreciated until the 1870s, after the book *Algebraic Equations and the Theory of Substitutions* by Camille Jordan appeared.

Immortality

"In the theory of equations I examined the cases in which equations can be solved in radicals; it gave me the opportunity of making the theory more profound and of describing all possible transformations of an equation that are admissible even when it can't be solved in radicals."

This is a quotation from Galois's manuscript "A Treatise on Conditions for Solving Equations in Radicals." The ideas in the treatise weren't understood by his contemporaries, and they're still considered difficult. Nevertheless, the formulation of Galois's theorem isn't complicated. But it's necessary to understand some new concepts—primarily, the notion of a permutation group. (By the way, the term "group" was introduced by Galois, but the notion itself had first appeared in the works of Abel and Ruffini.) All the basic definitions can be found in "Marching Orders" by Alexey Sosinsky in this issue of *Quantum*, and you should look through that article before reading further. You'll find various examples there, but the definitions that are most important for us will be illustrated here once more by way of the group S_3 of permutations of three objects. There are only six such permutations:

$$p_0 = \begin{pmatrix}123\\123\end{pmatrix},\ p_1 = \begin{pmatrix}123\\312\end{pmatrix},\ p_2 = \begin{pmatrix}123\\231\end{pmatrix},$$

$$p_3 = \begin{pmatrix}123\\132\end{pmatrix},\ p_4 = \begin{pmatrix}123\\213\end{pmatrix},\ p_5 = \begin{pmatrix}123\\321\end{pmatrix}.$$

And here is the multiplication table for this group:

	p_0	p_1	p_2	p_3	p_4	p_5
p_0	p_0	p_1	p_2	p_3	p_4	p_5
p_1	p_1	p_2	p_0	p_4	p_5	p_3
p_2	p_2	p_0	p_1	p_5	p_3	p_4
p_3	p_3	p_5	p_4	p_0	p_2	p_1
p_4	p_4	p_3	p_5	p_1	p_0	p_2
p_5	p_5	p_4	p_3	p_2	p_1	p_0

Table 1

It's clear now, for example, that if we first perform p_2 and then p_3, we'll get p_5 as the result; performing these permutations in reverse order, we'll get p_4. So the commutative law isn't valid in multiplying permutations: we don't always have $a \cdot b = b \cdot a$. If this equation holds for all pairs of elements a, b of a certain group G, then G is called a *commutative* or *Abelian* group. You can verify on your own that of all the groups S_n, $n > 1$, only the group S_2 commutative.

A portion of the table, at the intersection of the rows and columns for p_0, p_1, and p_2, is displayed in blue. All the blue letters are p_0, p_1, or p_2 as well. This means that these three permutations themselves form a group, or, to be more exact, a *subgroup* of the group S_3. We denote it by $\mathbf{Z}_3$. (In Sosinsky's article the same notation was used for the rotation group of an equilateral triangle. But if you relate the permutation p_i to the rotation by the angle $120° \cdot i$, where $i = 0, 1, 2$, then you'll see at once that these two seemingly different groups have the same structure—that is, they're isomorphic.) Besides $\mathbf{Z}_3$, there are four more subgroups in the group S_3: one subgroup consists of only one element p_0, and three other groups each have two elements: $\{p_0, p_3\}$, $\{p_0, p_4\}$, and $\{p_0, p_5\}$. All these subgroups are commutative. You can check for yourself that there are no other subgroups in S_3. If H is a subgroup of G, we write $H \leq G$.

Let G be a group and a, b be elements of G. The expression $[a, b] = aba^{-1}b^{-1}$ is called the *commutator* of elements a and b: it can be used as a correcting term for a and b to change places:

$$ab = [a, b]ba.$$

If $ab = ba$, then $[a, b] = e$. It's clear that the more commutators there are in group G that are not e, the greater the deviation of group G from a commutative one. Let's call the subgroup G', consisting of all possible products of the form

$$[a_1, b_1] \cdot [a_2, b_2] \cdot \ldots \cdot [a_k, b_k],$$

where $a_1, \ldots, a_k, b_1, \ldots, b_k$ come from G, the *derived subgroup* of group G. It's clear that if group G is commutative, then G' consists only of the identical permutation e. Here's a simple exercise for you: verify that the commutators of the group S are given by the following table:

$[p_i,p_j]$	p_0	p_1	p_2	p_3	p_4	p_5
p_0	p_0	p_0	p_0	p_0	p_0	p_0
p_1	p_0	p_0	p_0	p_2	p_2	p_2
p_2	p_0	p_0	p_0	p_1	p_1	p_1
p_3	p_0	p_1	p_2	p_0	p_1	p_2
p_4	p_0	p_1	p_2	p_2	p_0	p_1
p_5	p_0	p_1	p_2	p_1	p_2	p_0

Table 2

For G' we might also consider the derived group $(G')' = G''$, called the *second derived group* of group G. Continuing this process, we'll get the *kth derived group* $G^{(k)} = (G^{(k-1)})'$. It's evident that $G^{(k)} \le G^{(k-1)}$. This gives rise to a nested chain of subgroups:

$$\ldots \le G^{(k)} \le G^{(k-1)} \le \ldots \le G'' \le G' \le G.$$

If this chain breaks at the subgroup consisting only of the identity permutation—that is, if $G^{(m)} = e$ for a certain number m, then the group G is called *solvable*.

It's obvious that *any commutative group is solvable*. In particular, the group S_2 is solvable. Let's show that the group S_3 is solvable too. Table 2 shows that all commutators from S_3 are in $\mathbf{Z}_3$, so $S_3' = \mathbf{Z}_3$. We can see from table 1 that the group $\mathbf{Z}_3$ is commutative, so $S_3'' = \mathbf{Z}_3' = e$.

But many groups are not solvable. For example, *the groups S_n are solvable only when $n = 2, 3, 4$.* (This far from simple statement first appeared Galois's treatise.)

Now we're ready to explain the main point of Galois's theory. Let

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n = 0$$

be an arbitrary equation of the nth degree, where $a_0, a_1, \ldots, a_n$ are given. Already at the end of 18th century Karl Friedrich Gauss proved that for any $a_0, a_1, \ldots, a_n$ the given equation has n complex roots $\alpha_1, \ldots, \alpha_n$. We'd like to find out whether there are formulas that express the roots $\alpha_1, \ldots, \alpha_n$ in terms of the coefficients $a_0, a_1, \ldots, a_n$ by means of the four arithmetic operations and radicals. For simplicity let's suppose that $a_0, a_1, \ldots, a_n$ are rational numbers and that all the roots $\alpha_1, \ldots, \alpha_n$ are different. We'll assign to the system of roots $\alpha_1, \ldots, \alpha_n$ the set $Q(f)$ consisting of all numbers of the form

$$\frac{P(\alpha_1, \alpha_2, \ldots, \alpha_n)}{R(\alpha_1, \alpha_2, \ldots, \alpha_n)},$$

where P and R are polynomials in n variables with rational coefficients. Let's look at transformations of $Q(f)$ that take the sum of the numbers into a sum and the product into a product and leave the rational coefficients unchanged. If β is a root of our equation—that is,

$$a_0\beta^n + a_1\beta^{n-1} + \ldots + a_n = 0$$

—and φ is such a transformation, then

$$\begin{aligned}\varphi(a_0\beta^n &+ a_1\beta^{n-1} + \ldots + a_n \\ &= a_0\varphi(\beta)^n + a_1\varphi(\beta)^{n-1} + \ldots + a_n \\ &= 0.\end{aligned}$$

This means that $\varphi(\beta)$ is a root of the same equation—that is, φ simply rearranges the roots $\alpha, \ldots, \alpha_n$ and so determines a certain permutation:

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ a_{i_1} & a_{i_2} & \ldots & a_{i_n} \end{pmatrix}.$$

All such permutations form a certain group contained in S_n. This group is called the *Galois group* of the equation $f(x) = 0$ and is denoted by $G(f)$.

As an example of calculating the group $G(f)$, let's look at the sixth-degree equation

$$f(x) = (x^2 - x + 1)^3 - a(x^2 - x)^2 = 0.$$

We'll rewrite it in two different forms:

$$(x + 1/x - 1)^3 - a(x + 1/x - 2) = 0; \quad \text{(a)}$$
$$[x(1 - x) - 1]^3 + a[x(1 - x)]^2 = 0. \quad \text{(b)}$$

Form (a) shows that if x is a root of our equation, then $1/x$ is also its root; form (b) says that $1 - x$ will be its root as well. So if one of the roots of our equation is denoted by θ, then

$$1/\theta,\ 1 - \theta,\ 1/(1 - \theta),\ (\theta - 1)/\theta,\ \theta/(\theta - 1)$$

will also be its roots. The number a can be selected so that the equation will be irreducible; then all these six roots will be different. We see that the Galois group consists of the transformations

$$\varphi_0(u) = u,\ \varphi_1(u) = 1/u,\ \varphi_2(u) = 1 - u,$$
$$\varphi_3(u) = u/(u - 1),\ \varphi_4(u) = (u - 1)/u,$$
$$\varphi_5(u) = 1/(1 - u),$$

which constitute a group with respect to the composition operation ($\varphi_i \circ \varphi_j = \varphi_k$, if $\varphi_k(u) = \varphi_i(\varphi_j(u))$) that is isomorphic to S_3 (φ_i corresponds to p_i; check it yourself!).

The fundamental theorem

The equation $f = 0$ is solvable in radicals if and only if its Galois group $G(f)$ is solvable.

The value of Galois's theorem lies in the fact that the group $G(f)$ can be calculated, as a rule, without knowing the roots of equation $f = 0$ just by looking at its coefficients. Consider, for example, the equation

$$x^3 + bx + c = 0$$

with rational coefficients but not rational roots. Let $\Delta = -4b^3 - 27c^2$. If Δ

isn't a perfect square, then $G(f) = S_3$; otherwise $G(f) = \mathbf{Z}_3$. Both groups are solvable, as is the equation. When the coefficients $a_0, a_1, \ldots, a_n$ are chosen more or less arbitrarily, the Galois group of equation $f = 0$ will be S_n. Since group S_n is not solvable for $n \geq 5$, *a general equation of degree* $n \geq 5$ *is not solvable in radicals.*

The primary value of Galois's work didn't lie in his exhaustive answer to a question that had been a challenge to every mathematician in the world for three centuries. It was his method, in which the notions of group and symmetry are central, that was truly significant. Galois's ideas proved fruitful in all branches of mathematics and theoretical physics. The range of applications of the general idea of symmetry stretches from abstract algebra to the theory of elementary particles. In the history of mathematics one cannot find another example of a such a small work having such a tremendous impact.

Galois's contemporaries knew him only as a passionate republican revolutionary. Only after his death was he publicly called "a good mathematician" for the first time. On June 7, 1832, the *Gazette des Hôpitaux* published a police notice: "Young Évariste Galois, 20 years of age, a good mathematician, famous for his flaming imagination, died at 12 A.M. of acute peritonitis from a bullet fired at a distance of 25 paces."

. . . The day was beginning to dawn when Évariste Galois finished the letter—the last he was ever to write:

"My dear friends! I have been provoked by two patriots . . . It is impossible for me to refuse. I beg your forgiveness for not having told you. But I have given my adversaries my word of honor not to inform any patriot. Your task is simple: prove that I am fighting against my will, having exhausted all possible means of reconciliation; say whether I am capable of lying even in the most trivial matters. Please remember me, since fate did not give me enough of a life to be remembered by my country."

Circle No. 10 on Readers Service Card

IN THE LAB

Learning about (not by) osmosis

"Something there is that doesn't love a wall . . ."—Robert Frost

by Alexander Borovoy

CUT OFF A BRANCH OF A plant and it begins to fade. Put it in water and soon its leaves become smooth again, resilient, full of moisture.

Why does the branch come back to life? What forces made the moisture penetrate the plant and move inside it? What keeps the water in the cells and doesn't let it escape? Why is a plant cell permeable to water in only one direction—from the outside in?

Scientists tried for a long time to answer these questions, but a definite answer didn't come until the end of the 19th century. A short while after that, scientists managed to model (though roughly) these natural phenomena. And nowadays these phenomena are put to work in many different areas of science and technology.

Discovering of osmosis: who and how

In 1848, while studying how liquids boil, the French physicist and experimenter Jean-Antoine Nollet came upon an unknown phenomenon. In one of his experiments he hermetically sealed a glass of alcohol with the bladder of an ox and put it on the bottom of a large container filled with water. Several hours later the bladder was swollen—water had gotten into the glass and increased the pressure in it. Nollet explained this surprising fact in this way: "The bladder of an animal may be more permeable for water than for alcohol; in this case, the rate of penetration by water is greater than the rate of penetration by alcohol."

Let's repeat Nollet's experiment using more accessible materials: we'll substitute a piece of cellophane for the ox bladder and sugar solution for the alcohol.

Figure 1 depicts a simple experimental setup. A wide glass tube, hermetically sealed below with cellophane, is placed in a jar full of pure water. The cellophane is attached so as to keep water from seeping into the tube. (To create a good seal you can use a rubber band, waterproof adhesive tape, modeling clay, or any other appropriate substance.) A thin glass tube enters the wide glass tube through a rubber stopper.

At the beginning of the experiment sugar solution is poured into the tubes until the levels of liquid in the tubes and in the jar coincide. (This moment is shown in figure 1 by the dotted line.) Soon you'll notice that the height of the solution in the tubes has increased—just as in Nollet's experiment, water began moving through the barrier (cellophane, in our case).

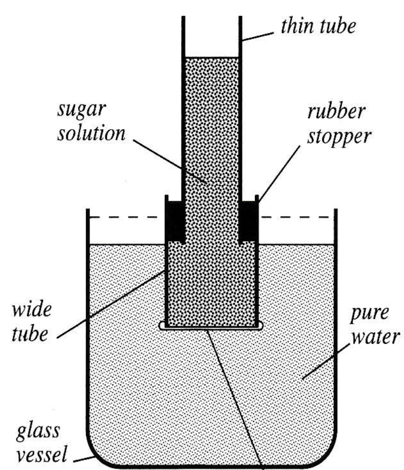


Figure 1

The phenomenon of one-way penetration of a solvent through a semipermeable membrane separating a solution and a pure solvent is called osmosis. This term comes from a Greek word that means "thrust" or "push."

Now that we've reproduced the experiment of the French scientist who discovered osmosis, let's try to understand this phenomenon.

Osmosis and one-way diffusion

Let's recall what happens when some substance is dissolved in a solvent. Molecules of the substance penetrate the solvent, and molecules of the solvent penetrate the region occupied by the solution. This mutual penetration leads to an equal concentration of the dissolved substance throughout the solution.

Now let's imagine that the solution and the pure solvent are separated by a semipermeable membrane—it allows molecules of the solvent to pass but not molecules of the dissolved substance. Clearly in this case equalization of concentrations can occur only by means of one-way diffusion of solvent.

That's what takes place in our experiment. Cellophane is not "transparent" for sugar molecules, but it is "transparent" for molecules of water;

Art by Leonid Tishkov

so at equal heights of liquid (equal pressure on the membrane) in the jar and the tubes at the beginning of the experiment, more molecules of water penetrate through the cellophane upward than downward. As a result, the level of liquid in the jar begins to decrease, and it increases in the tubes. We can express it differently: water penetrates the solution because of the force of "osmotic pressure." As soon as the hydrostatic pressure of the water column balances the osmotic pressure, the process stops. The height of this column is the quantitative measure of osmotic pressure.

Semipermeable membranes and the "artificial cell"

After Nollet's experiments many scientists—botanists, chemists, physicists—began to conduct experiments and study the phenomenon of osmosis. For semipermeable mem-

Figure 2

branes they used a wide variety of natural materials: the film from inside egg shells, ox and pig bladders and diaphragms, and others.

In 1866 the German scientist M. Traube invented the method of obtaining artificial semipermeable films made of copper ferrocyanide ($Cu_2Fe(CN)_6$). They were permeable for water but impermeable for most other substances.

We can check that by performing an elegant experiment that came to be called the "artificial cell." Don't be frightened by the film's complicated chemical formula. You can whip some up at home using blue vitriol ($CuSO_4$) and potassium ferrocyanide ($K_4Fe(CN)_6$), which are available at camera stores that sell darkroom chemicals.

Pour a weak solution of blue vitriol (approximately 3%) into a test tube and put in a small crystal of potassium ferrocyanide. The crystal must be clean, so it's better to chop it off a big crystal just before the experiment. Because of the reaction

$$2CuSO_4 + K_4Fe(CN)_6 \rightarrow Cu_2Fe(CN)_6 + 2K_2SO_4,$$

the surface of the crystal is covered by a semipermeable membrane. Water penetrates it and makes the "cell" grow. The cell "wall" expands, and at the points where it bursts from the internal pressure and some of the solution pours outside, the semipermeable cell wall forms again. In this way the cell starts branching out. So we might say that the "artificial cell" behaves like a living organism.[1]

This classic experiment requires patience and precision, and you may not obtain beautiful "plants" in your first try.

Nowadays you can perform a more effective experiment with more accessible substances, growing artificial plants in a water solution of liquid glass (sodium silicate (Na_2SiO_3), which is silicate office glue). Some people manage to grow an entire "orchard" by throwing in crystals of cobalt chloride, ferroferrous sulfate, nickel chloride, manganese sulfate, and many other substances. (See figure 2.)

Experiments with the osmometer and a theory of osmosis

A decade after Traube obtained the first artificial semipermeable membrane, the German botanist Wilhelm Pfeffer created an instrument for measuring osmotic pressure—the osmometer. The setup that we used to reproduce Nollet's experiment (fig. 1) follows Pfeffer's idea in many respects.

Using his osmometer Pfeffer showed that osmotic pressure depends only on the concentration and not on the nature of the dissolved substances (for dilute solutions) and that it increases with temperature.

[1]The experiment described calls for great care in handling the reagents. They must not soil your hands, fall on the floor, etc. If this happens, wash the chemicals off with lots of water.

His measurements also showed that osmotic pressure can be very high, reaching several atmospheres (1 atm $\cong 10^5$ Pa).

So quantitative data on osmosis were obtained, and they helped the outstanding Dutch chemist Jacobus van't Hoff formulate the theory underlying this phenomenon. In 1887 he published a paper in which he showed that molecules of a dissolved substance in solution behave like molecules of an ideal gas in a container. And the solvent in this case plays the part of . . . a vacuum!

In order to understand van't Hoff's reasoning, let's do a mental experiment. Imagine a horizontal cylinder filled with water, separated into two equal parts by a cellophane membrane that can be shifted without breaking the seal (fig. 3). If we dissolve sugar in one part of the cylinder, then, because of diffusion of pure water through the cellophane to the solution, the pressure on that side will increase and the membrane will begin to shift. The volume of the solution will increase, and the volume of pure water will correspondingly decrease. But as far as the results are concerned, this is equivalent to the supposition that molecules of sugar in the solution create a certain additional pressure on the membrane. The dissolved substance seeks to expand and shift the cellophane, just as a gas seeks to expand and shift a piston separating it from a vacuum.

On the basis of this supposition van't Hoff managed to explain all the features of osmosis discovered by Pfeffer. He obtained an equation completely analogous to the equation of state of ideal gas:

$$\pi V = nRT,$$

where π is the osmotic pressure, V is

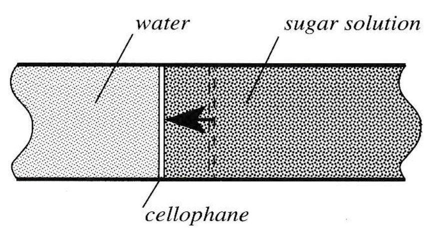


Figure 3

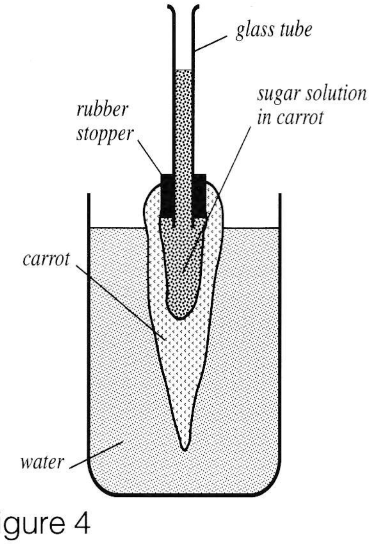


Figure 4

the volume of the solution, n is the number of moles of dissolved substance, R is the gas constant per mole, and T is the temperature. It was shown that the similarity in the behavior of a dissolved substance and a gas exists only for a very dilute solution—one in which the interaction among molecules of the dissolved substance can be ignored.

Van't Hoff's theory was corroborated by many experiments. Its author received the first Nobel prize in chemistry (1901) for "discovering the laws of chemical dynamics and osmotic pressure."

Answering the questions we posed

We began this article with a question: "What forces make moisture penetrate the cells of a plant and move inside it?" Now we can answer: basically these are forces associated with osmosis. The outer layer of the cell's protoplasm is a semipermeable membrane that allows the cell to regulate the exchange of water with the environment. If, for example, it loses moisture and the concentration of salts in the cellular fluids increases, penetration of water into the cell increases until the force of osmotic pressure is counterbalanced by the elastic forces of the stretched membrane.

To convince yourself of the semipermeability of natural cells, you can make an osmometer in which the role of the membrane is played by . . . a carrot (fig. 4)! True, this instrument won't be very precise, because the carrot's cells are partially permeable for sugar.

The evaporation of moisture from trees occurs over the huge surface area of its foliage. The concentration of salts in tree sap increases, and osmotic pressure (along with some other factors) makes water rise tens of meters. And it's not an insignificant amount of water—we're talking about dozens of liters for deciduous trees. No wonder tree branches and grass stems "cry" when they're cut.

Osmosis today

The unique properties of living organisms that allow their cells to absorb and transport many substances selectively have been investigated by many scientists. They've managed to create synthetic films, or membranes, impermeable for some substances and permeable for others. Later, "perfect filters" (or even "magic filters," as they are sometimes called in popular literature) found application in different spheres of science and technology.

These filters purify gas and oil products, freshen salt water, process milk and fruit juices, produce medicines, and do a lot more besides. Here are a few examples.

Membrane technology seems to be most widely applied in making salt water drinkable. Here the method of "reverse osmosis" is used: salt water is pressed against a semipermeable membrane at great pressure (greater than the osmotic pressure); pure water passes through, leaving behind the molecules of dissolved salts. The size and productivity of these devices vary widely, from portable units that freshen several dozen liters of water a day to the huge plants supplying Riyadh, the capital of Saudi Arabia, with fresh water (their productivity is 120,000 m^3 a day). Energy consumption with these devices is one tenth that of distillation plants (or better).

One of the first and perhaps most important applications of membranes is in the medical field—the semipermeable membrane for purifying blood in kidney machines. Now the membrane has "mastered" various other medical specialties as well. We can say that the time of the capsule-patch has arrived. This is a membrane system that introduces strictly controlled quantities of medicine to blood vessels through the skin. Capsule-patches with nitroglycerine are already available for heart patients.

Many large scientific and industrial enterprises are at work perfecting membrane technology. New methods of producing semipermeable films continue to appear. In fact, several years ago a method of producing them was developed that uses a heavy-nuclei accelerator. Under a beam of nuclei a synthetic film turns into a "sieve" with precise holes one thousandth of a millimeter in diameter.

The task in the immediate future is to devise a complete quantitative theory of the processes occurring in "magic filters" and, on that basis, develop membranes with predetermined properties.

Glittering performances

USA garners gold and silver at the International Physics Olympiad

by Arthur Eisenkraft and Larry D. Kirkpatrick

THE UNITED STATES WAS among six countries to win the thirteen gold medals at the XXII International Physics Olympiad held in Havana, Cuba, on July 1–9, 1991. Thirty-one countries brought a total of 149 competitors to the Olympiad to wind 13 gold, 10 silver, and 31 bronze medals. The competition was highlighted by the outstanding performance of the team from China, all five of whom won gold medals—an Olympiad first! The team from the Soviet Union won three gold medals, Hungary won two gold, and Canada, Norway, and the US each won a single gold medal.

The Olympiad exams consist of two parts. Students were challenged to solve three theoretical problems in five hours and one experimental problem in 4.25 hours. The first theoretical problem required the students to calculate the rebound angle for a spinning sphere that bounces to a specified height after being dropped onto a horizontal surface. The second problem required calculations of the properties of electrically charged balls moving at relativistic speeds around a square loop, which was also moving at a relativistic speed in the presence of a uniform electric field. In the third problem, the students investigated the possibility of cooling a gas using a laser beam.

In the experimental problem the students were told that a black box contained any three of the following four components: a battery, a diode, a resistor, and a capacitor. They then used various pieces of test equipment and circuits of their own design to determine which three components were in the box, how they were wired together, and the values for each component.

US Physics Team and academic directors Arthur Eisenkraft, Larry Kirkpatrick, and Avi Hauser.

Once again the US Physics Team made a fine showing, winning one gold medal, one silver medal, and an honorable mention. Derrick Bass (Florida) placed 13th overall in winning a gold medal. He was closely followed by R. Michael Jarvis (New York) in 21st place for a silver medal. Eric Miller (California), a junior, placed 63rd and received an honorable mention. Jason Sachs (New Jersey) and Theresa Lynn (North Carolina) placed 80th and 95th, respectively. Jason received a perfect 10 on the first theoretical question.

Cuban hospitality

The Cuban Ministry of Education and the Cuban Olympiad Organizing Committee were very successful in hosting this year's Olympiad. In addition to creating a successful and challenging set of exam questions, the Cubans treated the students and coaches to trips to the beach, the botanical gardens, old Havana, Hemingway's house, and Expo Cuba, as well as musical events. A highlight of the trip to the beach was observing the sun pass directly over our heads with our shadows at our feet.

Since no commercial flights exist between the US and Cuba, the team flew to Mexico City and stayed overnight before catching a Mexicana

Airlines flight to Havana. Derrick's flight from Miami to Mexico City to Havana is certainly one of the skinniest terrestrial triangles that we've ever dealt with. Try predicting and then calculating the small angle in Derrick's flight. How did you do? How does this compare with the angles used to measure distances to local stars when the Earth's major axis about the Sun is the baseline?

In contrast to the difficulties between our governments, the members and coaches of the US Physics Team were openly and warmly welcomed by the Cubans and were congratulated on our Independence Day, July 4th. In fact, there was some friendly ribbing over the results of the baseball games between the US and Cuban teams in preparation for the Pan American Games that were held near Havana in August.

Selecting and training the team

The selection of the 1991 US Physics Team began in November when invitations were sent to physics teachers and selected students across the United States. These teachers were provided with a sample preliminary exam and were requested to nominate their best student(s) who might be able to compete on the national level. A national exam was administered to these 450 candidates in February. The exam consisted of multiple-choice questions, selected by Ed Gettys (Clemson University) and the AAPT Exam Committee, and four open-response problems. The top 75 students were given a second, harder exam in March. This exam consisted of four open-response problems to be completed in 75 minutes and two additional problems to be completed in two hours. The top 20 students were then invited to a training camp during the first week in June, held on the campus of the University of Maryland (UM) and hosted by the Department of Physics and Astronomy.

During the seven-day camp the students enjoyed problem-solving sessions, rapid-fire lectures, laboratory experiments, and lots of testing on extremely difficult problems. The academic directors of the US Physics Team are Arthur Eisenkraft (Fox Lane High School in Bedford, New York) and Larry D. Kirkpatrick (Montana State University in Bozeman). They are assisted by coach Avi Hauser (AT&T Labs), Tom Kniess, and Chris Schafer (graduate students at UM). The students also had the chance to sample the frontiers of physics through guest lectures by Jim Gates, Jordan Goodman, and Ellen Williams from the UM Physics Department; take the physics IQ test à la Dick Berg (also UM); and learn some thermodynamics from Jack Wilson (Rensselaer Polytechnic Institute in Troy, New York).

A trip to Washington, D.C., included visits with Secretary of Education Lamar Alexander and the director of the National Science Foundation Walter Massey. Quick visits to the Einstein statue at the National Academy of Sciences and the Vietnam Veterans Memorial were followed by some free time at the National Air and Space Museum. A highlight of the Washington trip was a dinner hosted by IBM Watson Research with a fascinating after-dinner talk by Richard A. Webb.

The 1991 US Physics Team

The 20 members of the 1991 US Physics Team come from all over the United States, as can be seen in the following list (the members who represented the team in Cuba are marked by an asterisk):

Michael Agney, Melbourne High School, Melbourne, Florida (teacher: Carolyn Ronchetti)

***Derrick Bass**, *gold medal*, North Miami Beach Senior High School, North Miami Beach, Florida (teacher: Barbara Rothstein)

Chang Shih Chan, Northeast High School, Philadelphia, Pennsylvania (teacher: Raj G. Rajan)

Robert Hentzel, Ames High School, Ames, Iowa (teacher: Charles Windsor)

***R. Michael Jarvis**, *silver medal*, Fox Lane High School, Bedford, New York (teacher: Arthur Eisenkraft)

***Saul Kato** (alternate), Evanston Township High School, Evanston, Illinois (teacher: Robert Horton)

Irwin Lee, Naperville North High School, Naperville, Illinois (teacher: Maxine Wilverding)

Irwin H. Lee, St. John's School, Houston, Texas (teacher: Mark Kinsey)

Mark Liffmann, Phillips Academy, Andover, Massachusetts (teacher: Robert Perrin)

Lee Loveridge, Long Beach Polytechnic, Long Beach, California (teacher: Jim Outwater)

***Theresa Lynn**, NC School of Science and Math, Durham, North Carolina (teacher: H. B. Haskell)

Daniel Marks, Glenbrook South High School, Glenview, Illinois (teacher: John P. Lewis)

***Eric David Miller**, *honorable mention*, San Francisco University High School, San Francisco, California (teacher: Tucker Hiatt)

Jacob A. Morzinski, Los Alamos High School, Los Alamos, New Mexico (teacher: Leaf Turner)

Keith R. Pflederer, Libertyville High School, Libertyville, Illinois (teacher: Theodore Vittitoe)

***Jason Sachs**, Middletown High School North, Middletown, New Jersey (teacher: Edward Bechtel)

Eric Brian Shaw, Gaither High School, Tampa, Florida (teacher: Terry Adams)

Daniel Spirn, Cherry Hill High School West, Cherry Hill, New Jersey (teacher: H. K. Chatterjee)

Ryan Taliaferro, Highland Park High School, Dallas, Texas (teacher: Robert Roe)

Eric Tentarelli, Phillips Academy, Andover, Massachusetts (teacher: Robert Perrin)

All 20 team members returned home with a library of physics classics valued at over $500.

The Olympiad in 1992

The XXIII International Physics Olympiad will be held July 5–13, 1992, near Helsinki, Finland. Our hosts have promised long days and cool temperatures for the competition. Students who are interested in competing for a position on the US Physics Team and who have not received an invitation by December 1 should request an application and sample test from the US Physics Team, AAPT, 5112 Berwyn Road, College Park, MD 20740.

Sponsors

The US Physics Team is organized by the American Association of Physics Teachers (AAPT). Funding is organized by the American Institute of Physics (AIP). Sponsors of the 1991 US Physics Team who contributed $5,000 or more include AAPT, AIP, the American Physical Society, the American Vacuum Society, IBM, and the Optical Society of America. Contributors were the Acoustical Society of America, Addison-Wesley Publishing, Allyn & Bacon, Inc., the American Association of Physicists in Medicine, the American Crystallographic Association, AT&T, Beckman Instruments, Bell Core, BP Research, Inc., COMSAT Laboratories, Ford Motor Company, General Electric, GTE Laboratories, Hughes Aircraft Company (Research Lab), Janis Research Company, Inc., Lockheed Corporation, McGraw-Hill Publishing, Inc., W. W. Norton & Company, Phillips Petroleum Company, Prentice-Hall, Inc., Princeton University Press, Saunders College Publishing, Schlumberger-Doll Research, the Society of Rheology, University of Maryland, Westinghouse Foundation, John Wiley & Sons, and Worth Publishing.

Adapted from a report appearing in the AAPT *Announcer* (Sept. 1991).

Bulletin board

International Space Year

Governments worldwide have designated 1992 as the International Space Year. It was inspired in part by two historic events commemorated in 1992 whose themes have special relevance for the space age: the 500th anniversary of Columbus's voyage to the New World in 1492, with its themes of exploration and discovery; and the 35th anniversary of the 1957-58 International Geophysical Year (IGY), with its themes of scientific inquiry and global cooperation. Together, the universal perspectives of scientists and explorers capture the global outlook of the space age. The ISY in 1992 will be the first year-long worldwide celebration of humanity's future in this new, potentially transforming Age of Space.

Space agencies and educational organizations are coordinating efforts and activities related to the ISY. The January/February 1992 issue of *Quantum* is an official publication of the International Space Year and will be devoted entirely to space science and math. It will also contain information on ISY events around the country and around the world.

Space Science Involvement Program

The National Science Teachers Association (NSTA) and NASA invite you to participate in the long-running Space Science Student Involvement Program (SSIP). Since its debut in 1980–81, SSIP has involved nearly one million students in its program annually. The interdisciplinary design of the contests encourages all students to work on a science-related activity.

Students in grades 6–8 work in teams to design a moon base for human habitation, while those in grades 9–12 design and write up experiments that theoretically could be conducted on Space Station Freedom, in the wind tunnel facility at NASA Langley in Virginia, at the drop tower facility at NASA Lewis in Ohio, or with the supercomputer at NASA Ames in California. Students in all grades can enter the journalism and Mars settlement art contests.

All entrants receive a certificate of participation. Winning students visit a NASA center or attend the National Space Science Symposium, have their artwork seen in a national traveling exhibit, or have their newspaper entry printed in an NSTA publication. The top winners in the Space Station Proposal Contest compete at the national level for scholarships and other prizes.

Teachers can also participate. They receive the same travel awards as the student winners, as well as certificates and plaques. First-time winning teachers can compete for the Teacher-Newcomer Award, a week-long internship at a NASA center.

This year the National Space Science Symposium will be held in Washington, D.C., in September. Winning students will present their papers, meet with their representatives on Capitol Hill, and be honored at both an evening reception and awards luncheon.

For an SSIP poster and entry guidelines, contact NSTA–SST, 1742 Connecticut Avenue NW, Washington, DC 20009. The entry deadline is **March 15, 1992**.

"Together to Mars" finalists

A panel of judges chosen by NSTA has announced the three US finalists in the H. Dudley Wright student competition, "Together to Mars." NSTA is the US coordinator for this international competition, which is sponsored by The Planetary Society.

The competition was held to spark students' imagination and interest in space travel. Entrants had to state a problem that would be encountered on a mission to Mars, then write a research paper proposing ways to solve the problem.

The three US finalists are Michael Brush, Mission Hills, Kansas, for "The Effects of Microgravity on Cytochrome P-450"; George Zener, McLean, Virginia, for "Microgravity: Dealing with Weightlessness in a Mission to Mars"; and Noam Fast, Glen Head, New York, for "Combatting the Negative Effects of Prolonged Space Flight on the Psyche."

The finalists will proceed to the international level of the competition, where they will compete against national finalists from other countries. International winners each will receive $2,500 and an expense-paid trip to Washington, D.C., in August 1992 to accept their awards at the World Space Congress, part of the International Space Year commemoration.

Student Alternative Fuel competitions

What has a motor, chassis, gears, and tires, is guided by wires, and requires no gas to go? The race cars in the Junior Solar Sprint, the first national solar-powered model car race. Said to be the "Soap Box Derby of the 1990s," this year's race was held in Washington, DC, in July. Students from junior high schools across the country raced cars of their own design, powered entirely by solar panels. Each car was judged in two categories: its overall design, craftsmanship, and appearance and its performance in a double-elimination 20-meter sprint. Travis Talmadge, of Oakwood–Oster Junior High School south of Chicago, won this year's event.

The Junior Solar Sprint is part of the Student Alternative Fuel Competition program, designed to further student interest in science and the environment and encourage exploration of alternative energy sources. The program is operated by the US Department of Energy and the Argonne National Laboratory, a research facility southwest of Chicago. Other events planned for this program include the National Gas Vehicle Challenge, the Society of Automotive Engineers Supermileage Competition, and the National Fuel Economy Road Rally Championship. For more information on the Student Alternative Fuel competitions, call Bob Larsen or Marti Hahn at Argonne National Laboratory, 708 972-6489, or write to Argonne, Building 362-2B, 9700 South Cass Avenue, Argonne, IL 60439.

X-ray telescope in solar study

An X-ray telescope flown on a NASA sounding rocket above New Mexico at the time of totality of the July eclipse has produced high-resolution X-ray images of the uneclipsed sun that promise to complement other eclipse experiments performed throughout the world. The telescope was built by scientists at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York and at the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts. The goal of the coordinated data and image analysis is an improved understanding of the solar corona (the tenuous outer atmosphere of the sun that begins approximately 2,000 miles above its surface).

At the heart of the X-ray telescope, called NIXT (normal incidence X-ray telescope), is a multilayer X-ray mirror that overcomes the inability of conventional lenses and mirrors to work in the X-ray region of the spectrum. A multilayer mirror has the ability to enhance the low X-ray reflectivities of the materials of which they are composed.

The mirror assembly in this X-ray telescope consisted of 140 alternating layers of carbon and cobalt, designed to combine cumulatively the small amounts of X-radiation reflected from each of the interlayer boundaries. The thicknesses of those layers—only a few atoms each—were chosen so that the layers would optimally reflect only X rays with specific wavelengths. Here they were chosen so that optimal reflection would occur for X rays emitted by highly ionized atoms of iron and magnesium. Such X rays are emitted from the solar corona at a temperature of approximately 2 million degrees Celsius. The advantage of looking at features recorded at a single specific temperature is that the images are sharper—not blurred by X rays emitted at other temperatures. Researchers believe that normal incidence multilayer X-ray optics provide a valuable technique for observing coronal structure without contamination from radiation at other wavelengths produced by solar regions at various temperatures.

—Compiled by Elisabeth Tobia

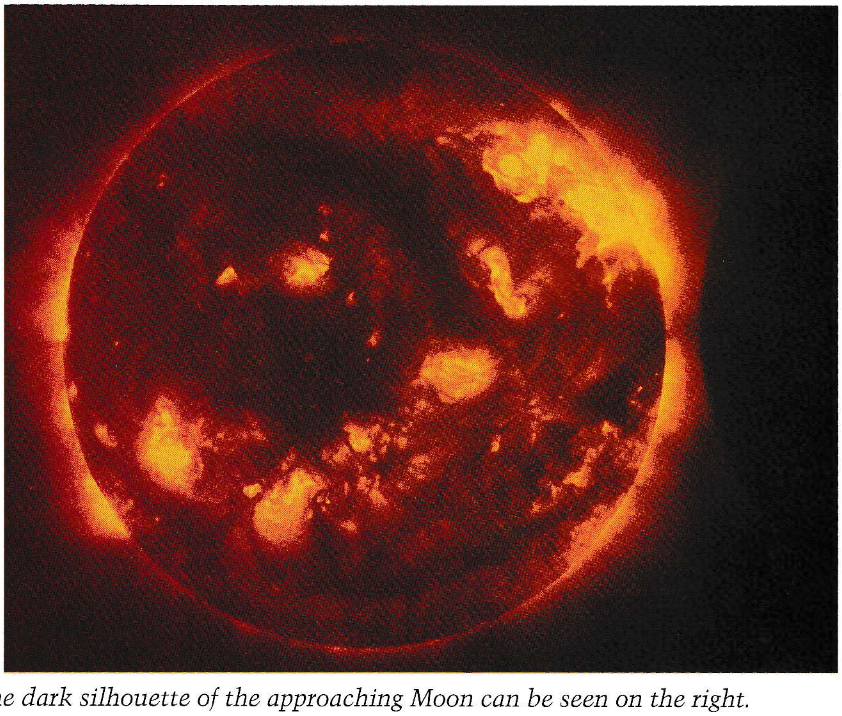


Photo courtesy of SAO/IBM Research

The dark silhouette of the approaching Moon can be seen on the right.

Circle No. 9 on Reader Service Card

ANSWERS, HINTS & SOLUTIONS

Math

M36

Divide the given square into $4 \cdot 4 = 16$ congruent squares (fig. 1). Then points L and K turn out to be nodes of the grid obtained. The rotation of the grid through 90° about point L obviously takes triangle LMD into triangle LNK. So angle DLK is equal to the rotation angle—that is, to 90° (and, in addition, $LD = LK$).

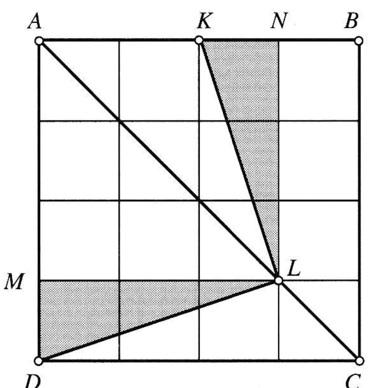


Figure 1

M37

We can represent the given table A as the sum of two tables B and C (fig. 2). If we assign minuses to numbers in all three tables in the same way according to the statement of the problem, then the sum of the numbers in every line of table B and in every column of table C will be zero. It follows that the sum of all the numbers in tables B and C, and so in table A also, will be zero.

1	2	3	...	10		0	0	0	...	0		1	2	3	...	10
11	12	13	...	20		10	10	10	...	10		1	2	3	...	10
21	22	23	...	30	=	20	20	20	...	20	+	1	2	3	...	10
. . .	. . .	. . .	. . .	. . .		. . .	. . .	. . .	. . .	. . .		. . .	. . .	. . .	. . .	. . .
91	92	93	...	100		90	90	90	...	90		1	2	3	...	10
		A						B						C		

Figure 2

M38

Let the condition of the problem hold for all sides of a pentagon $ABCDE$ except for side AE (and the corresponding diagonal BD—see figure 3). Segments AE and BD are parallel if and only if triangles ABE and ADE have the same area (the triangles have the common base AE, so the equality of their areas amounts to the equality of their heights—that is, to AE and BD being parallel). A similar argument for sides AB, BC, CD, and DE and the corresponding diagonals parallel to them yields the successive equalities of areas of triangles EAB and ABC, ABC and BCD, BCD and CDE, CDE and DEA. So the areas of EAB and DEA are the same, and this completes the proof.

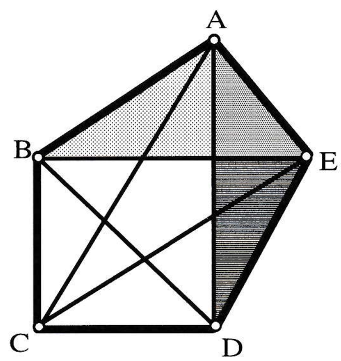


Figure 3

M39

The answer is no. Figure 4 illustrates the way to draw 14 lines separating the centers of all the chessboard squares. To prove that 13 lines won't suffice, consider the centers of 28 border squares (fig. 5). They form a square, and any 13 lines meet the sides of this square in at most 26 points. Therefore, such lines divide this square's perimeter into at most 26 pieces, so at least one of the pieces must contain two border centers. This means that 13 lines can't separate even the border centers, to say nothing of the others.

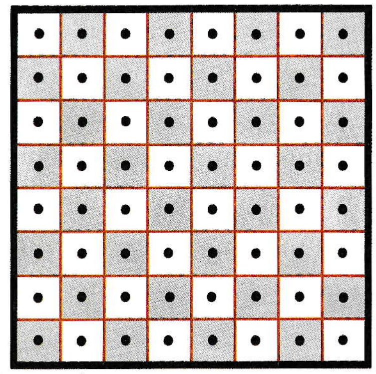


Figure 4

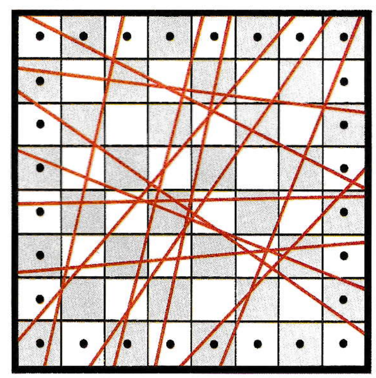


Figure 5

M40

The succession of numbers 1, 2, 3, 4 won't occur in our sequence 1, 9, 9, 1, ..., since each even number in it must be followed by four odd numbers.

Questions (b) and (c) require a more thorough examination. The answer to both of them is yes. By the definition of the sequence, knowing the preceding four numbers we can find the next number. Let's move along the sequence in the opposite direction and try to find the number x preceding the given four successive numbers a, b, c, d. Since d is the last digit of $x + a + b + c$, we can write

$$x + a + b + c = 10k + d,$$

or

$$x = 10k + (d - a - b - c),$$

where k is some integer. The number x is a digit—that is, $0 \le x \le 9$;, so x is the remainder of $d - a - b - c$ when divided by 10. This allows us to uniquely determine x. For instance, the succession 1, 9, 9, 1 must be preceded by the remainder of $1 - 9 - 9 - 1 = -18$ when divided by 10, which is equal to 2 (not 8!).

Thus, *the number preceding a given succession of four digits is always the same, no matter where and how many times this succession occurs in our sequence.*

Taking one more step back, we find that the same statement is valid for the number that stands two places before the given succession; and in general, given any four successive terms of the sequence, we can uniquely restore the entire segment preceding them.

Now we notice that at least one four-digit succession shows up twice in our infinite sequence, because the number of such successions is finite (it equals 10,000). Denote the repeated succession by a, b, c, d. Let its first occurrence be n places away from the origin of the sequence. This means that stepping n places back from the first a, b, c, d, we come to the initial four numbers 1, 9, 9, 1. But we'll also find the same numbers n places before the *second* occurrence of a, b, c, d. This means that the succession 1,9,9,1 is repeated in our sequence, giving the affirmative answer to question (b).

Finally, as we've noticed above, the digit preceding 1, 9, 9, 1 is 2, so the succession 2, 1, 9, 9 can be found in our sequence. And this is the answer to question (c).

Physics

P36

From the statement of the problem it follows that the force **F** is directed at some angle $\alpha \ne 0$ with respect to the initial velocity **v** of the body. (Otherwise—that is, if $\alpha = 0$—the change in velocity for equal time intervals would also be equal.)

Let **OA** be the vector of the initial velocity, **AB** the vector of the change in velocity for the time interval τ after the force is "activated." (See figure 6.) Then **OB** is a vector equal (in its absolute value) to $v_1 = v/2$ (the body's speed after time τ). Now let's draw a vector **BC** that is equal to |**AB**| and directed at an angle α relative to **OA**. Then **OC** is equal in its absolute value to $v_2 = v/4$ (the body's speed at the end of the time interval 2τ after the force was activated). Repeating this same procedure, draw a vector **OD** whose absolute value is equal to v_3 (the body's speed 3τ after the force was activated). Let the projections of **AB** on the x- and y-axes be Δv_x and Δv_y, respectively.

Then

$$(v + \Delta v_x)^2 + \Delta v_y^2 = v^2/4,$$
$$(v + 2\Delta v_x)^2 + (2\Delta v_y)^2 = v^2/16.$$

Taking into account that

$$v_3^2 = (v + 3\Delta v_x)^2 + (3\Delta v_y)^2,$$

we finally get

$$v_3 = v\sqrt{7}/4.$$

P37

The work done in lifting the sled is the sum of two parts: the work done against gravity $W_1 = mgh$ and the work done against friction W_2. To

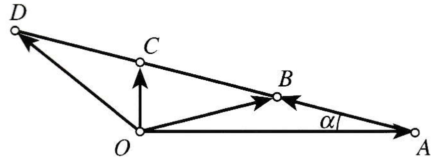


Figure 6

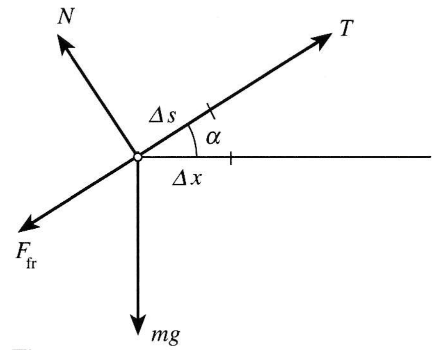


Figure 7

determine W_2 consider a small displacement Δs along the hill's surface. Let the tangent at this point be at some angle α relative to the horizon (fig. 7).

Since the child is ascending slowly and the rope tension T is directed along the tangent (that is, along Δs), we can assume the sled is in equilibrium, so that

$$F_{fr} = \mu N = \mu mg \cos\alpha.$$

The work ΔW_2 done against the force of friction in the displacement Δs is

$$\Delta W_2 = \mu mg \cos\alpha \cdot \Delta s.$$

We can see from figure 7 that $\Delta s \cos\alpha$ equals Δx minus the horizontal displacement of the sled; so it's clear that the total work done against friction in lifting the sled to the top is

$$W_2 = \mu mgl,$$

in that the total horizontal displacement of the sled is l.

So the total work W done by the child is

$$\begin{aligned} W &= W_1 + W_2 \\ &= mgh + \mu mgl \\ &= mg(h + \mu l) \end{aligned}$$

and doesn't depend on the steepness of the hill.

P38

A helicopter's lift (and an airplane's as well) depends on the density of the air. It's cooler in the morning, and the air is noticeably denser. That's why helicopters can carry greater loads in the morning. For airplanes, the increase in lift in the morning only matters during takeoff and landing; for helicopters, this is significant for the entire duration of the flight. In the thin mountain air the difference is substantial. So helicopter pilots in the mountains prefer to fly "in the thick morning air."

P39

Connect points A and B to a battery U_0 (fig. 8). It's clear that the potentials at C and D are equal; the same is true about the other pair of points E and F. This means that no current flows along the section of wire connecting points C and D, nor along that connecting E and F (the broken lines in figure 8). So we can get rid of these sections without changing the total resistance. The resistance of this simplified circuit is quite easy to calculate:

$$\begin{aligned} R_{AB} &= \frac{1}{2}\left(\frac{a}{2}\rho + \frac{a\rho\frac{a}{\sqrt{2}}\rho}{a\rho + \frac{a}{\sqrt{2}}\rho} + \frac{a}{2}\rho \right) \\ &= \frac{a\rho}{2}\left(1 + \frac{1}{\sqrt{2}+1}\right) \\ &= \frac{a\rho}{\sqrt{2}}. \end{aligned}$$

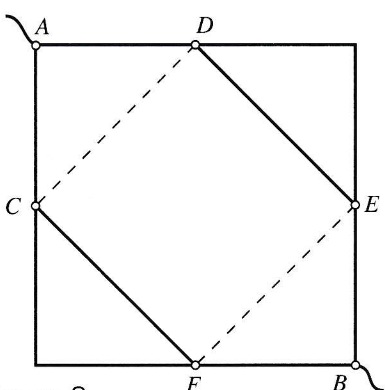


Figure 8

P40

When the light is reflected off the second mirror, it returns to the first mirror, is partially reflected, and, falling on the second, increases the total intensity of the beam of light.

Let I be the intensity of the initial light beam. (See figure 9—for illustrative purposes the incident beam is

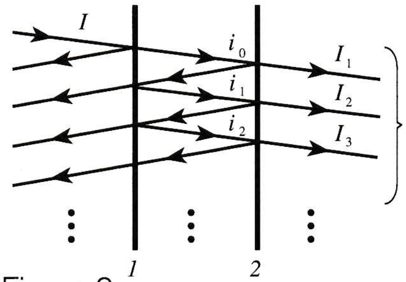


Figure 9

inclined.) The intensity of the light that has passed through the first mirror is $i_0 = I/k$ (where k is the "extinction factor" and equals 5 according to the statement of the problem). The light "trapped" between the mirrors is successively reflected by them and gradually emerges (since the mirrors are semitransparent); the amount of light emerging to the right is approximately equal to that emerging from the left.

So the total intensity of the light passed is

$$I' \cong \frac{i_0}{2} = \frac{I}{2k} = \frac{I}{10}.$$

The problem also has an exact solution. For the change in the beam's intensity after successive reflections we can write

$$\begin{aligned} i_1 &= i_0\left(1 - \frac{1}{k}\right)^2, \\ i_2 &= i_1\left(1 - \frac{1}{k}\right)^2 = i_0\left(1 - \frac{1}{k}\right)^4, \\ &\vdots \end{aligned}$$

Every time the light falls on the second mirror, it partially emerges to the right:

$$I_1 = \frac{i_0}{k},\ I_2 = \frac{i_1}{k},\ I_3 = \frac{i_2}{k}, \ldots.$$

The total intensity of the light passed is

$$\begin{aligned} I' &= I_1 + I_2 + I_3 + \cdots \\ &= \frac{1}{k}(i_0 + i_1 + i_2 + \cdots) \\ &= \frac{i_0}{k}\left[1 + \left(1 - \frac{1}{k}\right)^2 + \left(1 - \frac{1}{k}\right)^4 + \cdots\right], \end{aligned}$$

which is the sum of a vanishing geometric series with the factor $q = (1 - 1/k)^2$, and so

$$\begin{aligned} I' &= \frac{i_0}{k}\frac{1}{1-q} = \frac{i_0}{k}\frac{k^2}{k^2 - (k-1)^2} \\ &= \frac{i_0 k}{2k-1} = \frac{I}{2k-1} \\ &= \frac{I}{9}. \end{aligned}$$

For $k = 5$ this differs only slightly (10%) from the approximate solution.

Brainteasers

B36

Answer: $415 \cdot 382 = 158{,}530$.

B37

Fill the pot and tilt it as shown in figure 10. Because of the symmetry of the cylinder, 3 liters of water will be left. Now fill the jug and pour 3 liters out of it into the pot (until the pot is full). Then exactly 1 liter of water will be left in the jug.

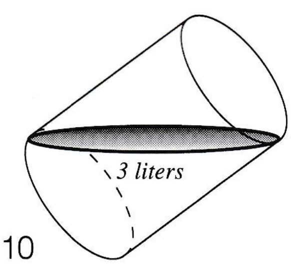


Figure 10

B38

Draw a circle whose center is at the vertex O of the given angle; label the points where it meets the arms of the

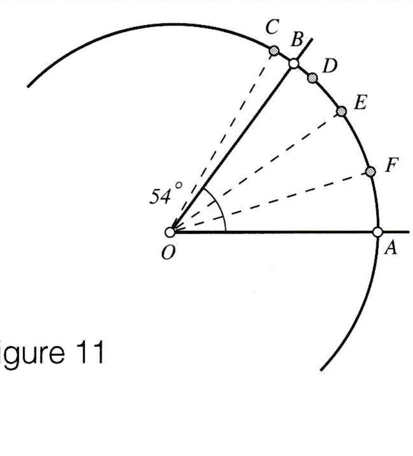


Figure 11

angle A and B (fig. 11). Using only a compass we can successively mark points C, D, E, and F on the circle such that $AC = AO$, $BD = BC$, $DE = DC$, and $EF = EB$. Then points E and F lie on the trisectors of the angle, since angle BOC = angle $DOB = 60° - 54° = 6°$, angle BOE = angle $EOF = 18° = 54°/3$.

B39

Label the coins 1, 2, 3, 4, 5, 6, 7 in clockwise order. Turn over coins 1, 2, 3, 4, 5, then 2, 3, 4, 5, 6, and so on seven times, each time shifting the first coin of the succession one position clockwise, so that the last group is 7, 1, 2, 3, 4. Each time we turn over five coins, so every coin will be turned over five times and eventually will stay upside down.

Turning four coins over at a time, one can't turn all seven of them over. To see this, write +1 on the heads of the coins and −1 on the tails. Then turning a coin over amounts to changing the sign on its upper side. When we turn four coins over, four signs are changed, so the product of all +1's and −1's on the upper sides of the coins always stays the same, whereas after turning over all seven coins it would have changed its sign.

Try to prove that with the "fivefold moves" one can turn over any given subset of the seven coins, and with the "fourfold moves" any subset having an even number of coins.

B40

While being transported overseas, cotton absorbs moisture and so gets heavier. The captains made allowance for this increase in the cotton's weight.

Polyominoes

1. $t(T) = (AB)^2A^{-1}BA^{-1}B^{-3} = (AB)^4 = e$.

2. The answer is no. Every figure made up of an even number of squares can be cut into a domino (fig. 12a) or "diagonal domino" (fig. 12b)—figures composed of two squares having a single common vertex. For both of these figures $t = e$ (with respect to D_4).

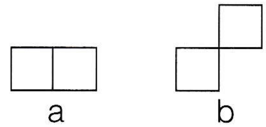


Figure 12

3. The desired group consists of all expressions of the form

$$A^{k_1}B^{l_1}\ldots A^{k_p}B^{l_p}$$

such that the corresponding polygonal curve, which may have self-intersections, is closed and bounds the figure whose "oriented area" can be divided by n. (The oriented area in this case can be calculated as follows. Take any unit square of the grid inside the polygon and find the number of full clockwise turns that are made by the vector joining the center of the square to a point on the line when this point runs along the entire line. This number can take any integer value. Then add up these numbers for all the squares.)

4. For the given polyomino P, $t(P) = (AB)^{18}$ or $(BA)^{18}$. If A and B are reflections in lines that meet at an angle of $\pi/18$, then AB and BA are the rotations by angles $\pi/9$ and $-\pi/9$, respectively, so $(AB)^{18} = e$.

5. Consider a group G generated by two elements A and B satisfying only two relations: $A^2 = B^2 = e$. Try to prove by induction over the number of squares in the polyomino P that $t_O(P) = (AB)^{2k}$, where $|k| = c(P)$ and the sign of k depends on the choice of the origin O, changing when O is shifted to the neighboring point of P's boundary.

Corrections

In the answer to Brainteaser B32 (Sept./Oct.): for "shaded" read "yellow"; for "AED" read "BCF"; for "ABF" read "ABE."

In the same issue, the first displayed equation in column 2 on page 26 contains several errors. The second instance of "$h\nu$" in the first line should read "$h\nu'$"; the numerator in the last line should read "$2h^2\nu^2$"; finally, what looks like a prime on the "ν" in the last line is actually a comma setting off the entire equation. The equation should read:

$$\begin{aligned}\Delta E &= h\nu - h\nu' \\ &= \frac{ch}{\lambda} - \frac{ch}{\lambda + \Delta\lambda} \\ &= \frac{2h^2\nu^2}{mc^2 + 2h\nu}.\end{aligned}$$

Portrait of three puzzle graces

Against the background of group theory

by Will Oakley

YOU DON'T HAVE TO BE A GROUP THEORIST to see that the three popular puzzles in figure 1 have something in common. Although they're totally different in design and appearance, all three pose the same problem: to restore the randomly scrambled puzzle to its pristine orderly state (the solid coloring of the cube faces, natural order of numbers on the pieces of the 15 puzzle, and 12 o'clock on all 18 dials of "Rubik's clock"). But the reader who's learned some elements of group theory—perhaps after reading the math articles in this issue—will see a much more profound connection between these puzzles and will appreciate them for the unique opportunity to literally "touch" groups. There are lots of group puzzles and puzzle groups. I'll just give a few examples of how groups are applied to puzzles, and how puzzles illustrate some notions of group theory.

So, what do groups have to do with puzzles? Let's look at puzzles as systems that can undergo certain manipulations (*actions* or *processes*) taking them from one state to another. Each action is a succession of elementary moves, described by the rules of the game: sliding pieces to a blank space in the 15 puzzle, turning the faces of Rubik's cube, or rotating cog wheels to make the hands of Rubik's clock go around. Two successions performed one after the other constitute a single one and determine a composite action; any succession can be undone by undoing all moves in reverse order. This means that these actions make a *group* (see the article by Alexey Sosinsky at the beginning of this issue), while elementary moves are the generators of this group. So the basic problem—*"find a chain of moves that produce the required result"*—can be reformulated as *"express the given element of the group in terms of the generators."*

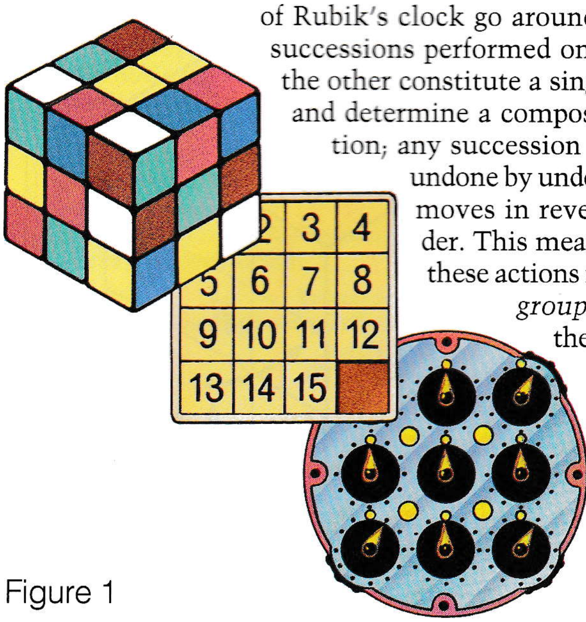


Figure 1

"What do I need all this scientific stuff for?" you may be saying. "I've never heard about groups, generators, or whatever else you want to foist off on me, but I've been doing the 15 puzzle all my life without any problems!" And basically you'll be right. But let me remind you of the craze that seized half of the world 100 years before Rubik's cube, when the ingenious inventor of the 15 puzzle, Sam Loyd, offered a $1,000 prize (in 1873!) for the first correct method of *restoring the numeric order in the box when pieces 14 and 15 are initially reversed.* Henry E. Dudeney, another great puzzle inventor, said that "it has been stated, though doubtless it was a Yankee exaggeration, that some 1,500 weak-minded persons in America alone were driven to insanity" by the 15 puzzle. But Loyd was risking nothing (which he knew full well). And the proof of the unsolvability of Loyd's problem is based on group theory, or rather, on permutation groups.

Each process permitted by the rules of the 15 puzzle can be associated with the permutation p indicating how the pieces are moved in this process. For instance, let the regular position be transformed into the "magic square" in figure 2a (where the red number 16 stands for the blank space, and the sum of the other numbers in every row, column, or diagonal is the same). The arrows in figure 2b show the displacements of all the pieces including the blank space ("piece" 16). Following the arrows, we find that piece 1 was carried to space number $p(1) = 3$ (that is, the space occupied by piece 3 in the regular position), piece 2 stayed where it was (so $p(2) = 2$), 3 went to 16, and so on, resulting in the array

$$p=\begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16\\ 3 & 2 & 16 & 8 & 10 & 11 & 5 & 12 & 6 & 7 & 9 & 1 & 15 & 14 & 4 & 13\end{pmatrix}.$$

12	2	1	15
7	9	10	4
11	5	6	8
16	14	13	3

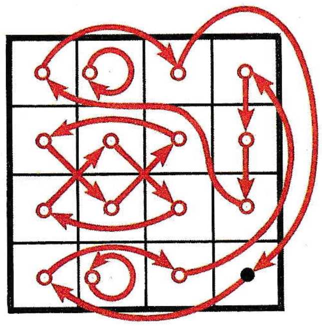


Figure 2 a b

The permutation *I* that would solve Loyd's problem is a single pair exchange of two elements (14 and 15). Each single move in the puzzle is also a pair exchange of two elements—of "piece" 16 and one of the neighboring pieces. Now, every process that brings the blank space back to the right bottom corner of the box consists of an even number of moves, since the number of up-moves in it must be equal to the number of down-moves and the numbers of right- and left-moves must also be the same. So the single pair exchange *I* must be represented as a sequence of an even number of pair exchanges. But it can be proved that, although a permutation can be represented by sequences of pair exchanges in many different ways, *the parity of the number of exchanges will be the same in all representations*. The unsolvability of Loyd's problem flows directly from this. Now try to figure out whether the "magic square" in figure 2a is attainable from the ordered position.

Sam Loyd had to come up with an unsolvable challenge to draw the attention of millions of people to his puzzle: the problems in the 15 puzzle that can be solved are too easy to excite people. Erno Rubik didn't have to resort to such underhanded means: it's much more difficult to do the cube, and lots of wretches were driven to despair by the thought that they'd overscrambled their cubes so that no one would ever set the tricky thing to rights. The solution always exists, of course, because when you play the cube you always stay within its group, and every process in the group is invertible—that is, unless you disassemble your cube. If you take it apart and put back together at random, you'll be able to restore the colors by turning faces with a probability of only 1/12 (to get the unsolvable position in the 15 puzzle, you also have to take the pieces out of the box and put them back in).

Rubik's cube is an ideal aid for a student of group theory, better than anything that ever was and, perhaps, ever will be invented. One can hardly find a notion of this theory that can't be illustrated with the cube. Take, for instance, *commutators*—elements of the form $[A, B] = ABA^{-1}B^{-1}$, mentioned in Y. P. Solovyov's article. They play the most important role in almost all algorithms for solving the cube. In the simplest case, A and B are the turns of two adjacent faces of the cube, as in figure 3, showing the effect of $C = [F^{-1}, R]$, where F^{-1} is the counterclockwise 1/4 turn of the front face, and R is the clockwise turn of the right face (these are the widely adopted notations). This commutator permutes three edges in cyclic order and at the same time performs two pair exchanges of corners. In all, seven small cubes are moved, which is not too practicable. But if we repeat process C twice, the corners will obviously come back, though twisted, so we'll get a single triple-cycle on the edges. The triple iteration of C is even more useful: it brings all the edges back to their initial positions, leaving just the two pair exchanges of the corners. (The sixfold iteration restores the initial position of all the cubes. A group-theorist would say that C is an *element of order* 6.) So our commutators generate two very useful processes. There are algorithms based completely on the commutators of two turns.

But the most impressive application of groups to "cubology" was the record algorithm of the English mathematician Morwen B. Thistlethwaite, which restores the solid coloring of the faces in not more than 52 moves, whatever the initial state is. Most of the algorithms consist of steps determined by the geometry of the cube: some of them restore the cube layer by layer; others put the corners in place first, then the edges; and so on. Thistlethwaite found an apt succession of subgroups, embedded one in the other, starting with the entire group of the cube and ending in the trivial group (which contains only the identity process, corresponding to the regular state of the cube). The subgroups are generated by gradually reduced sets of turns—for example, the next-to-last set comprises only the half-turns of all the faces. So the goal of each step is to bring the cube into the state that can be put in order by a process of the next subgroup—that is, using only the turns of the next set of generators. However strange it may seem, none of the small cubes must be put in its place until the last step! More details of this algorithm, and other remarkable applications of group theory to "cubology," can be found in

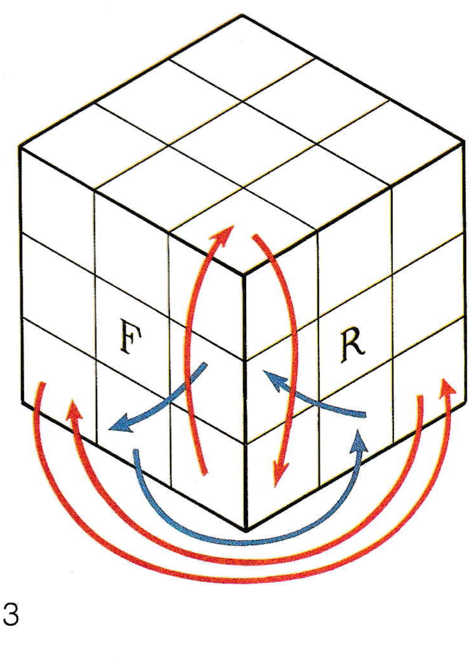


Figure 3

the *Handbook of Cubik Math* by Alexander H. Frey and David Singmaster (Enslow Publishers).

A few words about the third puzzle, Rubik's clock—the youngest of the three (it's just three years old). It differs from the other two in that it is *commutative*: the effect of a series of moves doesn't depend on their order, so it's completely determined by the number of iterations of each elementary move in the sequence. That's why Rubik's clock is much simpler than Rubik's cube (although, perhaps, more complicated than the 15 puzzle). In order to comprehend how much commutativity simplifies a puzzle, imagine for a moment that the cube has turned commutative. In this case the order of face rotations doesn't matter at all: we just have to know the total angle of rotation of each face, which can take the values $90^\circ \cdot k$, where $k = 0, 1, 2, 3$. With just six faces, we'd be able to obtain at most $4^6 = 4{,}096$ different processes. In fact, the number of cube processes (that is, the order of the cube group) equals $8! \cdot 12! \cdot 3^7 \cdot 2^{10} \cong 4.3 \cdot 10^{19}$, approximately 10^{16} times more! What makes the clock not so easy to solve is the greater number of elementary moves (31 compared to 6 for the cube) and the tricky way in which they are interlaced with each other. But for a mathematician this puzzle is in a certain sense trivial, at least theoretically. Given the readings of all the dials, one can find the angles by which the hands should be turned to indicate 12:00 o'clock. Each elementary move turns some of the hands through a certain angle x, depending on what buttons are pushed; the other hands don't move. It's convenient to measure this angle by points on the dial or in "hours" (1 hour = 30 degrees). All the different ways of choosing the "transmission" and the wheel that's turned yield 31 different ways of turning the hands. So we can denote by x_i ($i = 1, \ldots, 31$) the total angle by which the hands are turned with all the ith moves occurring in the required process, and calculate the resulting turn of each hand by summing up the turns produced by all moves. Equating the linear expressions thus obtained to the desired turn angles for each hand yields a system of linear equations that can be solved by a routine procedure used for similar (though somewhat simpler) systems in high school algebra. One thing worth mentioning is that the "numbers" x_i involved in our equations should, in fact, be treated like angles. For example, if $x_1 = 7$ and $x_2 = 8$, then $x_1 + x_2 = 3$ ($= 15 - 12$). A mathematician would say that our variables aren't integers but elements of the cyclic group of order 12 (or the group of rotations of the regular 12-gon—again see Sosinsky's article). However dull and cumbersome, the algebraic solution of the clock works all right, once again demonstrating the power of group theory.

These are just three landmarks in the vast realm of group puzzles; we'll continue to explore it in future issues of *Quantum*.

Index of Advertisers

NEW! VERSION 2

Recursive & iterative programming • Over 70 new mathematical functions • New manual, 500 examples & exercises

The Joy of DERIVE®ing.

PC MAGAZINE EDITORS' CHOICE
May 29, 1990
Derive, Version 1.6

DERIVE®– A Mathematical Assistant program.

PC Magazine says it's *"a joy to use"* and proclaims it "Editors' Choice." PC Week calls it *"fast and capable."* The *DERIVE®* program is delivered with built-in standard equipment that delights both math lovers and math phobics. After one test-*DERIVE*, you'll call it your essential mathematical assistant.

High performance.

In a test against the Big 3 symbolic math programs, PC Week rated the *DERIVE* program the highest in power in symbolic capabilities, *and* the highest in performance relative to price.

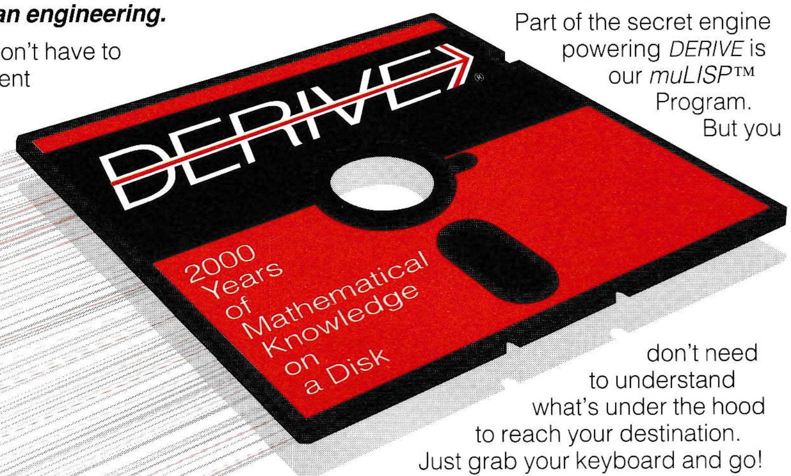
DERIVE does numeric and symbolic equation solving, exact and approximate arithmetic to thousands of digits, calculus, trigonometry and matrices. It displays formulas in comprehensible 2D format using raised exponents and built-up fractions. It plots beautiful 2D curves and 3D surfaces on monochrome or color graphics monitors.

Compact and sporty.

PC Magazine calls it a *"small wonder."* All it takes is 512K of memory and one floppy disk drive. *DERIVE* takes to the road on PC compatibles, and really gets around on laptop and even hand-held computers!

Human engineering.

You don't have to be fluent in computerese to use *DERIVE*. If you need to use math, *DERIVE* is for you. *DERIVE's* menu-driven interface and on-line help make it easy.



In fact, it's the friendliest and easiest to use of any symbolic math package on the road today. The learning curve is gentle–you'll soon be up to speed doing math, instead of trying to learn how to drive the software.

Freedom to maneuver.

DERIVE's automated expertise releases you from the drudgery of hand calculations. You are free to follow your intuition–and explore less travelled routes.

You can do problems you'd never attempt otherwise and obtain exact symbolic solutions, in addition to approximate numerical solutions.

Speed.

PC Magazine declared *DERIVE "faster, occasionally dramatically faster"* than its competitors. Goes from 0 to 600 decimal places in a fraction of a second!

Powerful engine.

Part of the secret engine powering *DERIVE* is our *muLISP*™ Program. But you don't need to understand what's under the hood to reach your destination. Just grab your keyboard and go!

Long distance DERIVEability.

This year's model has been improved to solve differential and recurrence equations. Other new features include least-square fits, partial fractions, Fourier series, Laplace transforms, and special functions from Bessel through hypergeometric. And there'll be more down the road.

No sticker shock.

DERIVE's suggested retail price is $250. Most other scientific programs for PC's start at twice that price, yet can't do symbolic math. And of course *DERIVE* doesn't require an expensive computer, a math co-processor, or even a hard disk drive.

Order *DERIVE* through your favorite software dealer or mail-order house. For a list of dealers, write Soft Warehouse, Inc. at 3615 Harding Avenue, Suite 505, Honolulu, HI 96816. Or call (808) 734-5801 after 11 a.m. PST.

And happy *DERIVE*ing!

DERIVE is a registered trademark of Soft Warehouse, Inc.

Handcrafted Software for the Mind

Soft Warehouse INC
HONOLULU · HAWAII

Circle No. 3 on Reader Service Card