QUANTUM

NOVEMBER/DECEMBER 1999

US $7.50/CAN $10.50

Springer

NSTA

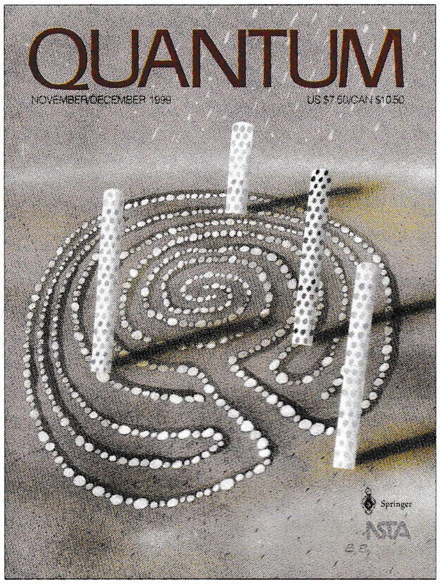
Pen and black ink with brown, blue, and pink wash on laid paper, 3 3/8 × 4 7/8, Gift of Mr. and Mrs. Neil Phillips,

A Boy on a Sled (late 1560s) by Jost Amman

AT THE SAME TIME THAT THE BOY DEPICTED above was sledding down a hill in Germany, a young Galileo Galilee (1564–1642) may have been sliding down a slope on the other side of the Alps in Italy. Perhaps it was during such an outing that Galileo became interested in the brachistochrone—the trajectory of most rapid descent from one point to another. Then again, maybe he was just having fun. Alpine athletes in general are affected by an overwhelming number of complex physical forces that are impossible to consider during their activities. You, however, can look back from the warmth and comfort of your winter lodge to consider what these downhill daredevils are up against by turning to page 20.

QUANTUM

NOVEMBER/DECEMBER 1999 VOLUME 10, NUMBER 2



Cover art by Vasily Vlasov

When students try to decipher the prose used to present a problem, they often feel as though they are stumbling through a maze in the dark. If enough light is shed on the situation, however, students can identify indicators and evaluate alternatives in order to wind their way through the problem with confidence. To discover the secrets of breaking down the language barriers present in many word problems, turn to At The Blackboard I on page 36.

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

FEATURES

4 Inner Circles
The Feuerbach theorem
by V. Protasov

10 Maxwell's Demon
On the quantum nature of heat
by V. Mityugov

14 Math Royalty
Carl Friedrich Gauss
by S. Gindikin

20 Alpine Incline
Solving for the slalom
by A. Abrikosov

DEPARTMENTS

3 Brainteasers

27 How Do You Figure?

28 Kaleidoscope
Returning to a former state

32 Physics Contest
A question of complexity

36 At the Blackboard I
Selecting the best alternative

41 In the Lab
Heating water from the top

44 At the Blackboard II
Obtaining symmetric inequalities

49 Commentary
The toy that drove the universe

51 Crisscross Science

52 Happenings
Three golds and two silvers in Italy
... Bulletin Board

55 Answers, Hints & Solutions

NSTA *Quantum* (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 10 (6 issues) will be published in 1999–2000. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; electronic mail: quantum@nsta.org. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

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

Subscription Information:
North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $49. Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Mail order and payment to: Springer-Verlag New York, Inc., Journal Fulfillment Services Department, PO Box 2485, Secaucus, NJ 07094-2485, USA. Telephone: 1(800) SPRINGER; fax: (201) 348-4505; e-mail: custserv@springer-ny.com.

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

Advertising:
Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) Jay Feinman (212) 460-1682; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Visit us on the internet at www.nsta.org/quantum/.

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

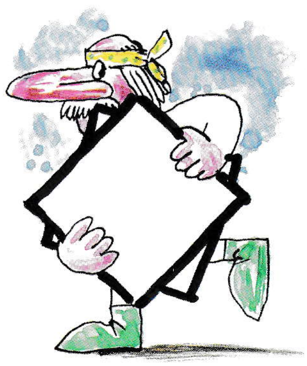
A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the Russian Academy of Sciences
in conjunction with
the American Association of Physics Teachers (AAPT)
& the National Council of Teachers of Mathematics (NCTM)

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

Just for the fun of it!

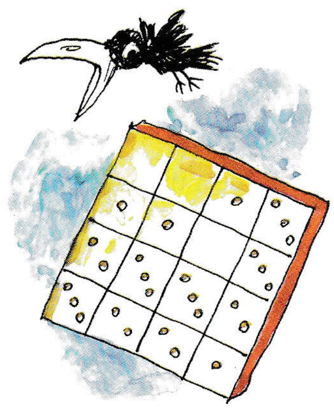
B276

Returnable bottles. A government program allows people to collect empty milk bottles and exchange them for bottles full of milk. Four empty bottles may be exchanged for one full bottle. How many bottles of milk can a family drink if it has collected 24 empty bottles?



B277

Polygonal cover-up. Is it possible to cover a polygon without right angles by squares? (Squares are not required to be equal and may intersect.)



B278

Domino theory. Eight different dominoes lie on a plane. No boundaries between the dominoes are shown in the figure. Draw these boundaries.

B279

Short triangle. All altitudes of a triangle are shorter than 1. Can the triangle's area be greater than 1000 square units?

B280

Running landscape. When you look out of the window of a moving train, it seems that everything outside is moving counter to your direction of motion. However, the farther an object is away, the more slowly it seems to move. Why?

Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON INSIDE BACK COVER

The Feuerbach theorem

Exploring the inscribed and escribed circles of triangles

by V. Protasov

IN THE FIRST HALF OF THE previous century, the German mathematician Karl Wilhelm Feuerbach proved one of the most elegant theorems of plane geometry. This article is devoted to that theorem and related topics.

The nine-point circle

Let ABC be a triangle, and A', B', and C' be the midpoints of its sides. The triangle $A'B'C'$ is called the midpoint triangle for ABC. The circle γ circumscribed about $A'B'C'$ is called the Euler circle for triangle ABC. In what follows, we will say "circle XYZ," meaning the circle passing through three given points X, Y, and Z. First, we consider a few introductory propositions. We assume that the reader is familiar with the Euler (or nine-point) circle, and with the concept of homothecy. If you don't know about these concepts, any book on advanced Euclidean geometry will fill you in.[1] The notation used is shown in figure 1.

Exercise 1. Prove that:

(I) the center of the circle circumscribed about a triangle lies at the point of intersection of the altitudes of its midpoint triangle.

(II) the radius of circle γ is half of that of circle ABC.

(III) the centroids (points of intersection of the medians) of triangles ABC and $A'B'C'$ coincide, and triangles ABC and $A'B'C'$ are homothetic with respect to their common centroid with a coefficient of -2.

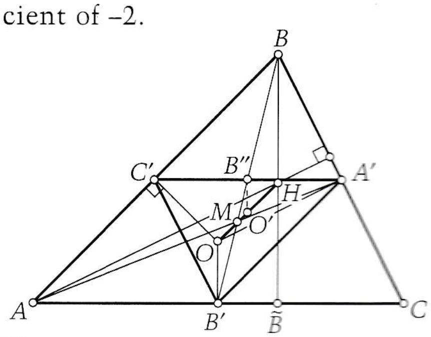


Figure 1

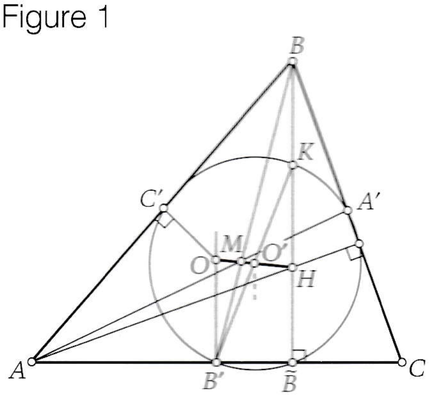


Figure 2

(IV) (Euler's theorem) for any triangle ABC, its orthocenter H, the center O of circle ABC, and the centroid M lie on a straight line (Euler's line), and $2OM = MH$.

(V) the center of the Euler circle of triangle ABC coincides with the center of segment OH.

Hint. The perpendicular to segment $A'C'$ at point B'' intersects the straight line OH at the point O' for which $OM/MO' = B'M/MB'' = 2/1$ so that $B''M = (1/2)B'M$.

Similarly, the other perpendicular bisectors of $A'B'C'$ pass through point O'.

Why is the Euler circle called the nine-point circle? The explanation is given by the following theorem, which was known to Euler.

Theorem 1. *Let γ be the Euler circle of triangle ABC. Then, the following six points lie on this circle in addition to the midpoints of the sides: the feet of the triangle's altitudes and the midpoints of the segments connecting its vertices with the orthocenter.*

You can prove this theorem by solving the following exercise.

Exercise 2. Prove that if $\tilde{B}$ is the foot of the altitude drawn from vertex B, and line $B'O'$ intersects $B\tilde{B}$ at a point K, then $BK = KH = B'O$ and

[1] For example, H. S. M. Coxeter and S. L. Greitzer: *Geometry Revisited*, Mathematical Association of America, New Mathematical Library, 1967.

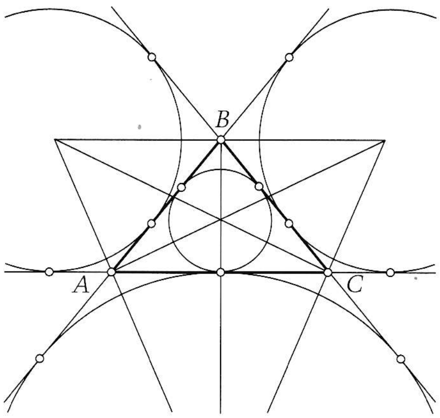
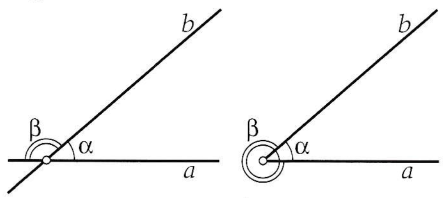


Figure 3



a b

Figure 4

$B'O' = O'K$ (fig. 2) (use Exercise 1, parts III and IV).

It follows from the result of Exercise 2 that $B'K$ is the diameter of the Euler circle; that is, points $\tilde{B}$ and K lie on this circle, which proves the theorem.

In this article we won't use the properties of the nine-point circle that are proven in Exercises 1 and 2. However, we recommend that you complete these exercises in order to get a better knowledge of the object whose deeper properties are analyzed in this article.

The Feuerbach theorem

Recall that the escribed circle of triangle ABC corresponding to side AB (or to vertex C) is the circle that is tangent to AB and to the extensions of sides AC and BC. Any triangle has three escribed circles. The center of the escribed circle corresponding to vertex C is the intersection point of the bisector of angle C and the bisectors of the external angles at A and B of triangle ABC (the reader is invited to prove this fact).

Thus, if three lines make up a triangle, there exist exactly four circles that each touch all three lines (fig. 3).

Now we can formulate the main theorem of this article.

Feuerbach's theorem. *The nine-point circle is tangent to the inscribed circle and to all the escribed circles of a triangle.*

We'll see later that in fact the Euler circle is tangent not only to the inscribed circle and all the escribed circles, but also to 60 other circles related to the triangle.

The segment theorem

To facilitate the discussion, it will be convenient to slightly redefine the way we talk about angles and arcs. We will, for the discussion below, make the following definitions.

Definition 1. *The angle between two different lines a and b that intersect at point O is the angle by which line a must be turned counterclockwise about point O until it coincides with line b.* (Such an angle is often called an *oriented* angle.)

In figure 4a, α is the angle between lines a and b, and β is the angle between b and a (it is clear that $\alpha + \beta = \pi$). This definition of an angle makes clear which of two supplementary angles is considered in any particular case. The angles between two rays a and b emanating from a common point are defined in a similar way. In figure 4b, α is the angle between rays a and b, and β is the angle between b and a (in this case $\alpha + \beta = 2\pi$).

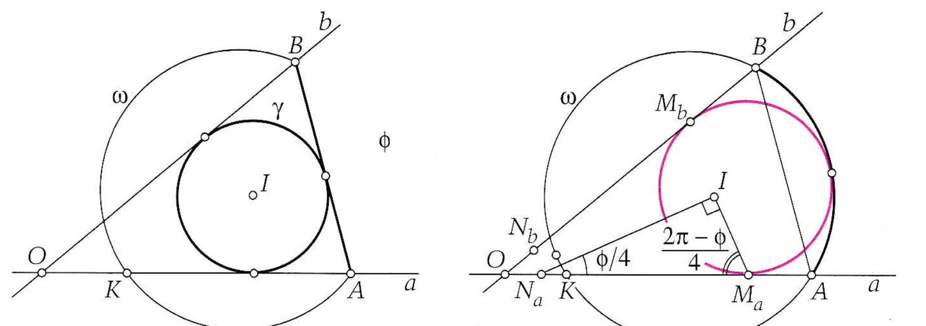
Definition 2. *An arc $\overset{\frown}{AB}$ of a given circle is the arc that is traversed by a point moving along this circle counterclockwise from A to B.*

Two points divide a circle into two arcs. With this definition, it is clear which of these two arcs is named by the symbol $\overset{\frown}{AB}$. It is clear that arcs $\overset{\frown}{AB}$ and $\overset{\frown}{BA}$ comprise the entire circle.

Finally, recall that the angular measure of arc $\overset{\frown}{AB}$ of a given circle is the angle between rays OA and OB, where O is the center of the circle, and the angular measure of arc $\overset{\frown}{BA}$ is the angle between rays OB and OA. Thus, $\overset{\frown}{AB} + \overset{\frown}{BA} = 2\pi$.

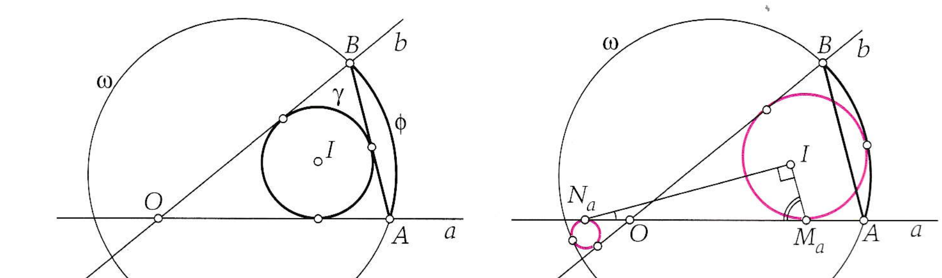
Now we formulate a theorem that is very important for further analysis and is interesting in its own right.

Theorem 2 (segment theorem). *A number ϕ is given, where $0 < \phi < 2\pi$, and a circle γ centered at I is inscribed in the angle formed by lines a and b (figs. 5–8). Arbitrary points A and B are chosen such that line AB touches circle γ (figs. 5–7), and a circle ω is drawn through points A and B for which $\overset{\frown}{AB} = \phi$. Then there*



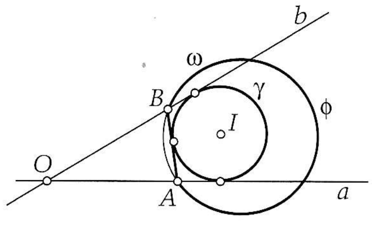
a b

Figure 5

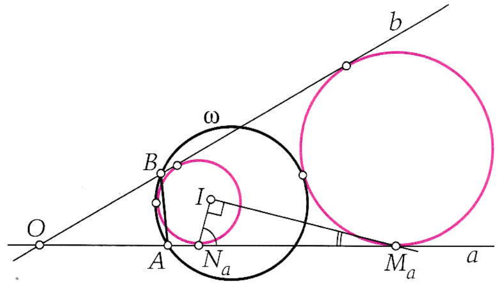


a b

Figure 6

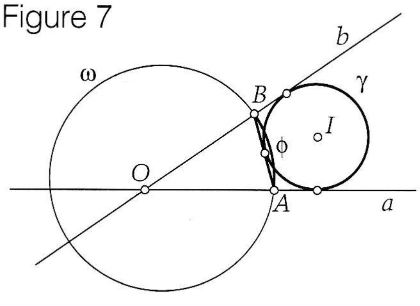


a

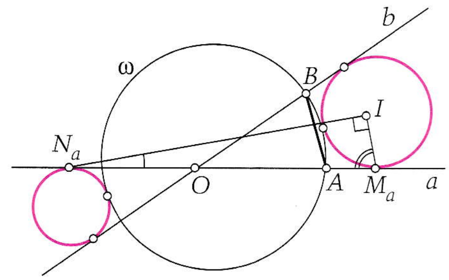


b

Figure 7



a



b

Figure 8

exist two fixed circles that are tangent to lines AO and OB and circle ω.

If one of these circles touches these lines at points N_a *and* N_b*, and the second one, at points* M_a *and* M_b*, then* $\angle AN_aI = \phi/4$ *and* $\angle IM_aO = (2\pi - \phi)/4$ *(so that triangle* N_aIM_a *is a right triangle).*

Certainly it is difficult to make sense of such a complicated proposition. To begin to think about it, imagine that points *A* and *B* move along rays *OA* and *OB* so that circle γ remains inscribed in triangle *AOB*

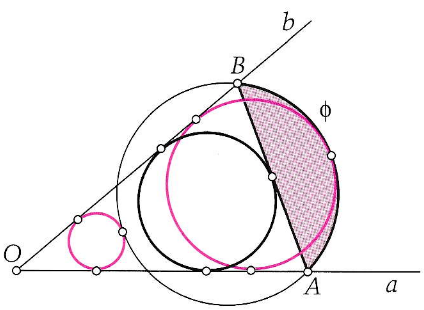


Figure 9

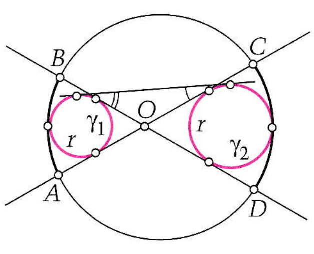


Figure 10

(fig. 5a). Let us construct an arc of a circle passing through points *A* and *B* equal to ϕ on the outer side of triangle *AOB* (this arc, together with segment *AB*, bounds the pink region in figure 9). Then, the variable circle ω containing arc ϕ touches two fixed circles, shown in red in figures 5–9.

Figure 9 illustrates one case of the segment theorem. Another possibility is shown in figures 6a and 6b. Here the angle ϕ is small (more precisely, $\phi/2 < \angle AOB$) and the point N_a "has moved" to the extension of ray *AO* beyond point *O* (why?), so that the red circle touching line *a* at point *Na* is inscribed in the angle vertical to $\angle AOB$ rather than in $\angle AOB$ itself.

Finally, circle γ can be not only inscribed, but also escribed for triangle *AOB*. This situation gives rise to two more cases (figures 7 and 8).

Our definitions of angle and arc have made it possible to put forth a unified formulation of the segment theorem that includes all these cases at once. For the time being, let us take the segment theorem for granted, without proof, and obtain several consequences.

Exercises.

3. Prove that line *BK* is parallel to the tangent to circle γ drawn from point N_a (figure 5b).

4. (I) Points *A* and *B* move on the sides of angle *O* so that triangle *AOB* has a fixed inscribed circle γ. Prove that the circumscribed circle for this triangle is tangent to a fixed circle inscribed in angle *AOB*, and that the points of tangency of this fixed inscribed circle to the angle's sides and the center of γ lie on a line.

(II) Prove proposition (I) if γ is the fixed escribed circle of triangle *AOB* corresponding to vertex *O*.

5. (I) Prove that the circle inscribed in the right angle *C* of triangle *ABC* and internally tangent to its circumscribed circle is homothetic to the inscribed circle of triangle *ABC* with respect to point *C* with a coefficient of 2:1.

(II) Prove that the circle inscribed in the right angle *C* of triangle *ABC* and externally tangent to its circumscribed circle is homothetic to the escribed circle of triangle *ABC* with respect to point *C* with a coefficient of 2:1.

(III) Prove the Feuerbach theorem for a right triangle.

6. Two nonintersecting circles γ_1 and γ_2 are given. An arbitrary circle externally is tangent to γ_1 and γ_2 and intersects their common internal tangents at points *A, B, C,* and *D* (fig. 10). Prove that

(I) the angular measures of arcs $\overset{\frown}{AB}$ and $\overset{\frown}{CD}$ are constant and equal to twice the angles formed by lines *AB* and *BD* with the common external tangent of γ_1 and γ_2.

(II) Triangles *AOB* and *COD* have fixed inscribed circles.

7. The previous exercise is just a statement of the proposition converse to the case of the segment theorem illustrated in figure 6. Formulate and prove similar propositions for the cases illustrated in figures 7 and 8.

8. Two circles γ_1 and γ_2 are given, tangent externally at point *O*. A circle ω is externally tangent to both of them. Prove that the common internal tangent to γ_1 and γ_2 (drawn at point *O*) divides ω into two arcs equal to twice the angles between this tangent and other common external tangent to γ_1 and γ_2.

9 (the lune problem). An angle with vertex at point O and two numbers ϕ_1 and ϕ_2 are given such that $0 < \phi_1 < 2\pi$ and $0 < \phi_2 < 2\pi$. Points A and B move along the sides of the angle, and two arcs passing through points A and B equal to ϕ_1 and ϕ_2 are constructed on the outer side of triangle AOB, forming a "lune". Prove that if the circle containing the first arc is tangent to a fixed circle inscribed in angle O, then the circle containing the other arc is tangent to a fixed circle inscribed in this angle.

This assertion can be made even stronger: the circle containing arc ϕ_2 is tangent to two fixed circles (however, the second circle may be inscribed in the angle vertical to the given one rather than in the given angle itself).

In a certain sense, Exercise 9 is a strong generalization of the segment theorem. The latter is a particular (or, more precisely, a limiting case), when $\phi_1 = 0$, so that one of the arcs of the lune degenerates into a line and the lune itself degenerates into a segment. Nevertheless, the solution of the general lune problem follows from this limiting case.

Proof of the segment theorem

Let us prove (using the notations shown in figure 5b) that the circle that touches lines a and b at points N_a and N_b, respectively, is tangent to circle ω. The proofs for the second circle and for other cases are quite similar. The following proposition is a strengthening of the well-known theorem about the angle between a tangent and a chord.

Lemma 1. *Let points A and B be given on a circle. Suppose line a is tangent to this circle at point A. Then, the angle between lines a and AB is equal to one-half of arc $\overset{\frown}{AB}$. Conversely, if the angle between a line a passing through point A and a chord AB of a circle (also passing through A) is equal to $\overset{\frown}{AB}/2$, then line a is tangent to the circle at point A.*

Exercise 10. Prove Lemma 1.

Lemma 2. *If M is a point inside a convex quadrilateral $ABCD$ such that $\angle CMB = \angle MAB + \angle MDC$ (fig. 11), then circles AMB and CMD are tangent to each other at point M.*

Proof. Consider a point K inside the angle CMB such that $\angle KMB = \angle MAB$ and $\angle KMC = \angle MDC$ (the statement of the lemma guarantees that such a point exists). Since $\angle KMC = \angle MDC = \overset{\frown}{CM}/2$, circle CMD is tangent to line MK at point M by Lemma 1. Similarly, $\angle KMB = \angle MAB = \overset{\frown}{BM}/2$, and thus circle MAB is tangent to line MK at point M.

Thus circles CMD and BMA are tangent to each other, since KM is their common tangent at point M.

We preface the proof of the segment theorem with several observations. On the sides of some triangle AOB (fig. 12), we lay off points N_a and N_b such that $ON_a = ON_b$. Let I be the center of circle γ, and $\angle AN_aI = \angle IN_bB = \alpha < \pi/2$. Let V be the second intersection point of circles AIN_a and BN_bI. Since $\angle IN_bB = \angle IVB$ (as angles inscribed in circle BN_bI) and, similarly, $\angle AN_aI = \angle AVI$, we see that $\angle AVB = 2\alpha$. If we now draw circle ABV, we see that $\overset{\frown}{AB} = 4\alpha$; that is, the size of arc AB is independent of the choice of points A and B.

It remains to set $\phi = 4\alpha$ and prove the following proposition: *if $\angle AN_aI = \angle IN_bB = \phi/4$, then the circle that is tangent to AO and OB at points N_a and N_b is also tangent to circle ω at point V.*

Now we can complete the proof of Theorem 2. It follows from the aforesaid that point V lies on circle ω. It remains to prove that circle N_aVN_b touches lines a, b, and circle ω.

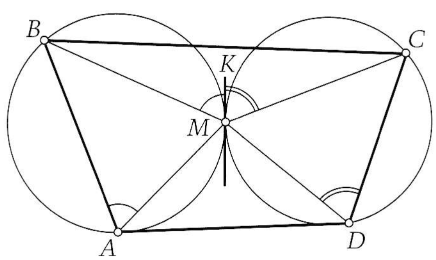


Figure 11

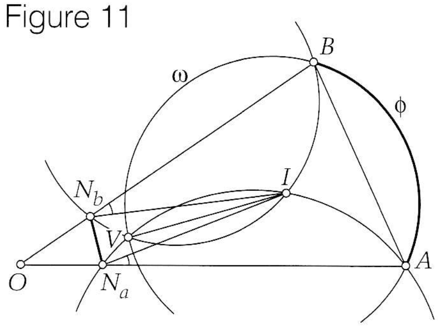


Figure 12

The rest of the proof is based on a calculation of various angles, which we are going to perform.

First of all, $\angle ON_bN_a = \angle ON_aN_b = (\pi - \angle AOB)/2$. On the other hand, $\angle N_bVN_a = 2\pi - \angle IVN_b - \angle IVN_a = \angle IAN_a + \angle N_bBI = (\angle OBA + \angle OAB)/2 = (\pi - \angle AOB)/2$. Thus, $\angle ON_bN_a = \angle N_bVN_a = \angle N_bN_aO$, and, by Lemma 1, circle N_aVN_b touches lines OA and OB at points N_a and N_b, respectively.

To prove the fact that circles N_aVN_b and ω are tangent to each other, it is sufficient to verify (Lemma 2) that

$$\angle VN_aN_b + \angle BAV = \angle BVN_b. \quad (1)$$

First of all, notice that

$$\angle BAV = \angle BAI + \angle IAV.$$

However,

$$\angle IAV = \angle IN_aV = \angle AN_aV - \angle AN_aI = \angle AN_aV - \phi/4.$$

Therefore,

$$\angle BAV = \angle AN_aV - \phi/4 + \angle BAO/2. \quad (2)$$

By Lemma 1,

$$\angle VN_aN_b = \angle VN_bB. \quad (3)$$

Therefore, the left side of equation (1) (with (2) and (3) taken into account) is equal to

$$\angle VN_bB + \angle AN_aV - \phi/4 + \angle BAO/2. \quad (4)$$

Finally,

$$\angle BVN_b = \angle BIN_b = \pi - \angle IN_bB - \angle N_bBI = \pi - \phi/4 - \angle OBA/2. \quad (5)$$

Quadrilaterals N_aVIA and N_bVIB are inscribed (each in its own circle); therefore,

$$\angle AN_aV = \pi - \angle VIA,$$

$$\angle VN_bB = \pi - \angle BIV.$$

Then,

$$\angle AN_aV + \angle VN_bB = 2\pi - \angle VIA - \angle BIV = \angle AIB.$$

Therefore, expression (4) is equal to $\angle AIB - \phi/4 + \angle BAO/2 = \pi - \phi/4 - \angle OBA/2$; that is, it coincides with the right side of equation (5), which was to be proved. Thus, the segment theorem is proved.

Exercise 11. Consider all other cases yourself.

The following exercises concern an arbitrary *curvilinear triangle* ABC consisting of segments CA, CB, and arc AB of a circle.

Exercises.

12. Using a compass and straightedge, construct a circle inscribed in a given curvilinear triangle ABC. Is such a construction always possible?

13. The inscribed circle of a curvilinear triangle ABC touches its arc $\overset{\frown}{AB}$ at a point V. Prove that the bisector of angle AVB passes through the center of the circle inscribed in the rectilinear triangle ABC.

14. The inscribed circle of a curvilinear triangle ABC touches side AC at a point M and arc $\overset{\frown}{AB}$ at a point V. Prove that circle MVA passes through the center of the circle inscribed in the rectilinear triangle ABC.

15. Let ABC be a rectilinear triangle. A circle inscribed in angle C internally touches circle ABC at a point M, T is the midpoint of arc AB containing point C, and I is the center of the inscribed circle of triangle ABC. Prove that points I, T, and M lie on a line.

Proof of the Feuerbach theorem

We will see that the Feuerbach theorem is a particular case of the segment theorem when the angle between lines a and b is $(2\pi - \phi)/2$.

Let $A'B'C'$ be the midpoint triangle of the given triangle ABC (fig. 13), N and N' be the points of tangency of line BC with the circles inscribed in triangles ABC and $A'B'C$, respectively, and I' be the center of the circle inscribed in triangle $A'B'C$.

We prove that the circle inscribed in triangle ABC touches circle $A'B'C'$, that is, the nine-point circle.

Triangles CAB and $CB'A'$ and, therefore, their inscribed circles, are homothetic with respect to point C with a coefficient of 2:1. Therefore, $CN = 2CN'$, $CN' = NN'$, and triangle $CI'N$ is isosceles.

We apply the segment theorem, setting $\phi = 2(\pi - \angle ACB)$. We have $\overset{\frown}{A'B'} = 2\pi - \overset{\frown}{B'A'} = 2\pi - 2\angle A'C'B' = 2\pi - 2\angle ACB$ (we use here the fact that $\angle A'C'B' = \angle ACB$, since $CB'C'A'$ is a parallelogram). Thus, $\overset{\frown}{A'B'} = \phi$. In addition, $\angle I'NC = (2\pi - \phi)/4$ (since triangle $I'CN$ is isosceles, then $\angle I'NC = \angle ICN = \angle ABC/2 = (2\pi - \phi)/4$ by the choice of the number ϕ).

By the segment theorem, the circle ω that passes through A' and B' for which $\overset{\frown}{A'B'} = \phi$ (this is just the nine-point circle) must touch the circle inscribed in angle ACB and line CB at point N (since for point N, $\angle I'NC = (2\pi - \phi)/4$); this is the inscribed circle of triangle ABC.

Replacing the word *inscribed* by *escribed* in the preceding reasoning, we obtain the desired result (for the escribed circles of triangles ACB and $A'C'B'$ corresponding to vertex C). Thus, the Feuerbach theorem is proved.

Notice that in the first part of the proof, only a half of the segment theorem was used (the theorem asserts the existence of two fixed circles tangent to circle ω). In this case, one of them is the inscribed circle of triangle ABC. And where is the second circle? The answer to this question is given in the following exercise.

Exercise 16. An isosceles triangle AKL with base AK is cut from an acute triangle ABC by a line tangent

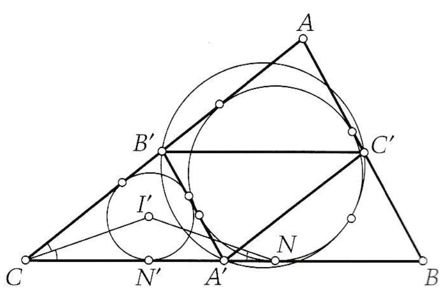


Figure 13

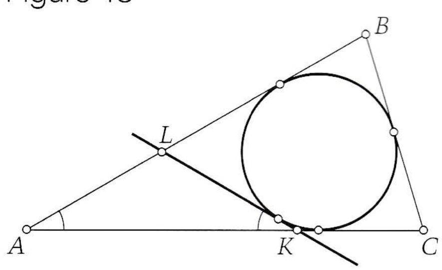


Figure 14

to the inscribed circle (fig. 14). Prove that the circle inscribed in triangle AKL is tangent to the nine-point circle of triangle ABC.

Since a small triangle like this one may be cut from each of the three angles of triangle ABC, we obtain three circles tangent to the nine-point circle (by the way, why only three and not six?).

What if we carry out this procedure for the escribed circles?

Exercise 17. Consider the extensions past vertex A of sides BA and CA of acute triangle ABC. Take points M and N on the extensions of BA and CA, respectively, such that line MN is tangent to the escribed circle of triangle ABC corresponding to vertex C, and triangle AMN is isosceles with base AM. Prove that the escribed circle of triangle AMN corresponding to vertex N is tangent to the nine-point circle of triangle ABC.

Points M and N could also be taken on the extensions of sides AB and CB beyond point B, which would give us one more circle tangent to the nine-point circle.

Exercise 18. Let the escribed circle of triangle ABC touch the continuations of sides CA and CB at points K and L, respectively. A point M is taken on segment CK and a point N on segment CL such that line MN is tangent to this escribed circle and triangle CMN is isosceles with base CM. Prove that the circle inscribed in triangle CMN is tangent to the nine-point circle of triangle ABC.

Exercises 17 and 18 assign to each escribed circle three circles tangent to the nine-point circle. Since every triangle has three escribed circles, we obtain 9 circles.

Now it's time to sum up. We have already constructed the following circles tangent to the nine-point circle: the inscribed circle, the 3 escribed circles, the 3 circles from exercise 16, and the 9 circles from exercises 17 and 18, which gives 16 circles in all.

However, these are not all the circles yet!

Exercise 19. Let H be the intersection point of the altitudes in triangle ABC. Prove that triangles ABC, AHC, AHB, and BHC have a common nine-point circle.

Each of the triangles mentioned in exercise 19 has its own set of 16 circles tangent to the nine-point circle, which gives us 64 circles tangent to the nine-point circle. This number deserves to be entered into the *Guinness Book of World Records*!

Let us conclude with a discussion of the points of tangency. In Exercises 20–22, F denotes the point of tangency of the nine-point circle of triangle ABC with its inscribed circle.

Exercises.

20. The sides of a triangle $A_1B_1C_1$ are parallel to the sides of triangle ABC and are tangent to its nine-point

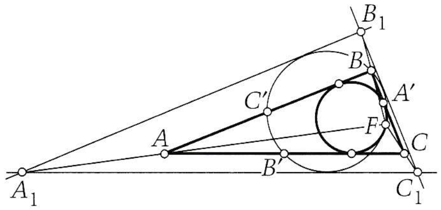


Figure 15

circle (fig. 15). Prove that lines A_1A, B_1B, and C_1C intersect at point F.

21. Prove that the nine-point circles of triangles AIB, BIC, and CIA intersect at point F, where I is the center of the inscribed circle of triangle ABC.

22. Let A', B', and C' be the midpoints of sides BC, AC, and AB of triangle ABC, respectively (the vertices are listed counterclockwise). Prove that the three lines connecting the centers of the inscribed circles of triangles $AC'B'$, $B'A'C$, and $C'BA'$ with the respective midpoints of arcs $B'C'$, $A'B'$, and $C'A'$ of the nine-point circle of triangle ABC meet at point F.

In conclusion, note that the segment theorem seems to have many other interesting consequences. There are two free parameters—the angle O between lines a and b and angle ϕ. We considered only two particular cases: $\angle AOB = \phi/2$ (Exercises 4 and 5) and $\angle AOB = (2\pi - \phi)/2$ (the Feuerbach theorem).

Message from afar

by David Arns

Once upon a weeknight dreary,
I beheld an image smeary,
Captured by a telescope that's been in space from days of yore,
As I sat with eyelids drooping,
A strange and unexpected grouping
Of celestial objects caught my eye like none had done before—
I knew I had to find out more.

I didn't know what I was seeing,
But I thought, "Another being
From another galaxy, perhaps an alien 'Signal Corps,'
Created this configuration
To confer some information
To a random listener. Yes, surely that is what it's for!"
Thus I let my fancy soar.

Then I stopped and gripped the table,
Forced my thoughts to be more stable,
Realizing I would need some proof, some evidence, and more.
So I called to book the Hubble—
To my surprise, I had no trouble
Getting seven hours' observation time, that day at four.
Now I'd give them proof galore!

So I made my observations,
Measurements, and calculations,
Disbelief and wonder nearly left me breathless on the floor.
This was proof beyond ignoring—
Sweat was from my brow outpouring—
I could see my name in scientific journals evermore!
(I'd been a no-name heretofore.)

Five weeks, and almost all was ready,
(I'd show those stuck-up folks at SETI!)
I merely had to translate all these symbols I had grabbed before.
Already I had seen a pattern: The spectrogram's bright lines were scatterin'
In ways that shocked, amazed, bewildered, stunned and shook me to the core
A *message* from a distant shore!

Methodically, I put together
Facts and data, heedless whether
Days were passing, pizza mould'ring, knocks and calls outside my door.
Finally, it was translated,
And I stood aghast, deflated:
The message from afar, for which I'd launched into my eight-week chore,
Read only, "Made in Singapore."

On the quantum nature of heat

Finding direction in chaos

by V. Mityugov

RECENTLY I CAME ACROSS the following problem given at a school Olympiad on physics: "Make and demonstrate a device that moves directionally under the influence of chaotic forces." I didn't need to rack my brains over this problem—the answer appeared immediately. I remembered the summer months of the post-war years, which I usually spent in a village with my relatives. The country boys showed me a game: when you are going to walk a long way, put a piece of wheat chaff under your shirt near the belt and forget about it. After a while the chaff can be found in different places—in the sleeve, at the back, or somewhere else. The reason is clear, but the result is always surprising.

One can think of quite a number of examples of such "mechanical rectifiers" which convert the energy of chaotic movements into translation motion. Mechanical rectification is the working principle of tidal power stations built in suitable ocean bays. At high tide the bay is closed off from the ocean by some kind of sluice gates, and during the following ebb a hydraulic turbine generates remarkably cheap energy. During the rising half period of the tides everything can be reversed, with the same result.[1]

Is this a good example of the conversion of chaotic motion into directed motion? Not really: the tidal ebbs and flows are related to the motion of the Moon, and there is nothing chaotic in them. What will change if, instead of regular tides, we have some irregular up-and-down fluctuations of the water level—say, the wind-driven surge of water in the Gulf of Finland (you may read the details in "The Bronze Horseman" by Alexander Pushkin). Evidently, in this case one would need special instrumentation that would provide the data needed to operate the gate properly.

A working model of such a power station could be constructed on a small lake or river where suitable hydrophysical conditions exist, or even in a basin or a pool. Classical mechanics and its subdivision hydrodynamics generally allow the wide use of scale modeling, so that large phenomena can be studied in small models and vice versa.

For example, before a large ship is built, its hydrodynamic stability and rolling and pitching motions under storm conditions are studied using a small model. How small can it be? It is important that the wave properties of the water surface be completely similar to those in a real storm. A model ship cannot be made arbitrarily small because at small scales the wave structure depends strongly on the surface tension. The surface tension, in turn, is caused by molecular attraction, which is described by quantum laws.

Suppose we make a very small model of a large lathe. With a suitable electric motor we can simulate all of its idle spinning motions. However, if we try to turn metal on such a model device, it will not work. The "scale invariance" principle of classical mechanics is not valid in this case because of the granular polycrystalline structure of a metal. This structure obeys the laws of quantum mechanics. The miniature cutter in a model lathe is made of a real metal, and for this reason it cannot work properly.

Let's return to chaotic (random, stochastic) motion. In 1871, the English physicist James Clerk Maxwell invented a hypothetical creature (which he called a "demon") who

[1]Tidal power stations are discussed in a recent *Quantum* paper: V. E. Belonuchkin, "Turning the tides," May/June 1998, pp. 10–14.

Art by Leonid Tishkov


Американская посылка - немного тепла для русского
Happy New Year
PHLOGISTON CO.
USA

could rectify the chaotic molecular motion of a heated gas. By manipulating a small gate, Maxwell's demon could sort the fast and slow molecules into two compartments of a vessel. The fact that Maxwell called it a demon indicates that he didn't believe in the possibility that such a microscopic physical device could exist. If he had believed it to be possible in principal, he would have called it a controller, operator, watchman, or the like. He felt intuitively that there was something wrong with such a hypothetical creature. What is wrong with it?

In Maxwell's day the kinetic origin of cold and heat was well known due to the works of Daniel Bernoulli and Michael Lomonosov. The accepted views on the nature of heat as chaotic motion of corpuscles in a heated body had driven out of science once and for all the concept of a special substance, "phlogiston," which played the role of the heat carrier. However, the language (a very conservative thing!) had preserved traces of this concept in such expressions as the "flow of heat." As to the quantum theory, there was no inkling of it at that time.

In many respects, the logical structure of classical mechanics is similar to geometry, a very "mathematical" discipline. Both have the property of scale invariance we mentioned earlier. This means that a simultaneous and arbitrary change of all spatial scales does not disturb the nature of the laws of motion and geometrical relationships. Indeed, the properties of the medians in a triangle, the Pythagorean theorem, and similar relationships do not depend on the scale. Similarly, the elastic collision of two small particles is quite similar to that of large ones. When Sir Ernest Rutherford proposed the now universally familiar planetary model of the atom, it immediately provoked fictional stories of intelligent inhabitants of the electron-planets. Thus it seems that at that level of understanding of the laws of physics, nothing may prevent one from designing a reasonable automatic device which should work as the hypothetical Maxwell's demon. The skepticism came from another direction.

As early as 1824, a French physicist named Sadi Carnot formulated the famous theorem on the limitation of the transformation of heat into useful mechanical work. Two decades later, this theorem was generalized by Rudolf Clausius and Sir William Thomson (Lord Kelvin), who raised it to the rank of a fundamental physical principle, the second law of thermodynamics. Thermodynamics turned into a self-sufficient, well-grounded, and axiomatic science seemingly independent of classical mechanics. It gave clear and reasonable solutions to important practical problems. For example, is it possible to take energy of thermal motion from a cold body and transfer it to a warm body without performing work? No, it is not. Can we obtain useful (that is, macroscopic mechanical) work due to a temperature difference of two bodies? Yes, we can.

The formal system of thermodynamics agreed with practically all observations in the natural sciences, from biological metabolism and chemical reactions to the planetary processes of re-radiation of solar energy into cold space. If the fundamental principles of Clausius, Kelvin, and Carnot were somehow cast into doubt, then much in our understanding of the living and nonliving worlds would need to be completely rethought (and in a way that is not yet known).

Actually, violation of the second law of thermodynamics would mean that we could completely extract, "free of charge," the heat (or cold, on a hot day) from the "sea" of thermal energy, and, when we were done with it, simply return this energy to its original chaotic state. This is called a *perpetuum mobile* (perpetual motion machine) *of the second kind,* and it doesn't violate the law of energy conservation!

It is doubtful that Maxwell could not see the radical consequences that the realization of his antithermodynamic demon would entail. It is more likely that this demon served to demonstrate the key scientific paradox that pointed to the logical dilemma that would confront the development of all system of natural sciences in the near future. Which way to turn? One possibility is to deny the absolute character of the thermodynamic restrictions on extracting useful mechanical work from thermal chaos. If that were the case, it would be puzzling that no living creature had learned to do it during the millions of years of evolution. The other possibility is that there exists a new fundamental law which works at the atomic level and which cannot be derived in Newtonian mechanics. An intrinsic feature of this new law would be violation of scale invariance in the transition from macroscopic motion to molecular thermal agitation.

Now is the moment to recall something. We intentionally forgot about quantum theory for a while—just to feel how badly we need it as a guide in the microcosm. Here we shall not consider the history of this wonderful discipline, however dramatic and fascinating it has been. Instead we focus on scientific paradoxes, which, like magic portals, lead us to unknown worlds. The next "magic portal" on our way is the famous Einstein-Podolsky-Rosen paradox.

Let's consider the elastic collision of two particles. In classical physics, this extremely simple model is widely used to study the conservation laws for energy and momentum. If the momenta of the colliding particles before the collision are known, then Newtonian mechanics can calculate (and thus predict) with absolute precision the respective values after the collision. It is a school problem which evidently has no hidden intrigue—at least if the particles are large enough to be considered as something like billiard balls.

The paradoxical situation is met when we try to apply quantum mechanics to describe the elastic collision of microscopic particles. Although in this case the total momentum and energy are conserved as in the classical case, it is impossible to assign individual momenta and energies to the particles after the collision. Isn't this a paradox? During the centuries of its development, the old classical physics never encountered situations in which the states of the subsystems (the particles) became indeterminate, while the state of the whole system (consisting of both particles) was completely determined from a physical standoint. Such a situation would be a nightmare for the classical physicist. Nevertheless, the logic and mathematical structure of quantum theory do provide for such a possibility.

Attempts to discuss the phenomena of the quantum microcosm with the language and images of classical physics lead to inferences which contradict elementary common sense. Werner Heisenberg repeatedly pointed out the necessity of developing a new physical intuition and a corresponding language of quantum mechanics. This doesn't mean that the new quantum images are inherently alien to conventional human reasoning. Even the good old classical physics is far from being "natural" for the human brain—it looks self-evident just because we adapted to it over the course of centuries. A certain old lady in Ilf and Petrov's famous novel *The Twelve Chairs* did not believe in electricity, and for this reason alone she did her cooking on a kerosene burner. And think how long it took before humankind accepted the idea that the Earth is round!

It may seem that we have gone astray from our original course. But in fact we have reached the crucial point in the story: evidently, the idea of scale invariance should be rejected when entering the precincts of the microcosm. There is a specific indeterminacy that is inherent in quantum mechanics. In the case of colliding particles it "disperses" the information about their prehistory among the "degrees of freedom." It is clear now how naive the idea of Maxwell's demon appears after the overthrow of the postulate that physical laws are independent of the spatial scale.

Without too much detail, violation of scale invariance can be illustrated by a dimensional analysis of the physical quantities. To make such a violation possible in the case of colliding particles, there must be one more combination of parameters of the motion with dimensions of length—in addition to the size of the particles and the extent of the effective interaction region in the collision. This is possible only if an additional dimensional constant of a universal nature is introduced into mechanics. By the way, the three laws of Newton have no such constant: in classical mechanics the mass, size, energy, and velocity are measured in arbitrary units. Accordingly, the choice of space-time scale (rulers and watches) is also arbitrary. This is the very reason why classical mechanics imposes no constraints on scaling.

In contrast, quantum mechanics has the necessary constant: this is the famous Planck's constant $h = 6.63 \cdot 10^{-34}$ J · s. Instead of the precise coordinates used in Newtonian mechanics, in quantum theory the motion of a particle is described with the help of a wave function. The free motion of a particle with mass m and velocity $\mathbf{v}$ is described by the propagation of a wave packet in which the spatial oscillations are characterized by the de Broglie wavelength $\lambda = h/(mv)$. This is what we were looking for! The ratio of λ to the purely geometrical length gives the necessary dimensionless parameter which tells us whether classical or quantum mechanics should be applied in a particular case. For example, if a 9-g bullet has a speed of several hundred meters per second, then λ is about 10^{-34} m. Clearly, in this case Newton's laws will work with tremendous accuracy. In contrast, the de Broglie wavelength for an electron with an energy of a few electron-volts is about the size of the hydrogen atom or of the characteristic interaction length for two colliding electrons.

Although the quantum collision process is wonderfully simple, it is very unusual. On the whole, the system of two colliding particles is not subjected to a kind of "chaotization" in the mechanical or any other sense. At the same time, the state of an individual particle becomes uncertain, and during subsequent collisions and contact with other objects (or with an observer's devices) it will behave as a child of chaos.

It is at this very point that mechanical motion can be joined with thermodynamic principles. One cannot help admiring the beauty of this physical picture. Indeed, some specific "nonredundancy" of the quantum description of matter underlies the nature of heat. There is nothing extra in the description beyond that which is sufficient to describe everything.

The Chinese are famous for their skill in inscribing poems on the surface of a tiny grain of rice. In contrast, it is impossible to "write" on an electron any information which is not equivalent to the parameters of the electron's state. In acquiring new information about the state of the partner particle in the collision, the electron "forgets" information about its own previous state. Therefore, an individual electron is an inherently unpredictable object. Let's point out once again that although there are no words in the common language to describe this quantum phenomenon, the mathematical formalism copes with it quite nicely.

Using the new principles and the respective mathematical formalism of the quantum theory of open systems, scientists have provided a strict mathematical description of the thermodynamic laws which were cleverly inferred in the last century. However, the previous picture of the "thermal chaos" of molecules has acquired some unusual features. The original mechanistic views considered thermal motion as a destructive phenomenon. It looked as if the chaotic collisions irreversibly destroy the traces of previous states of a physical system, irretrievably annihilating any information about them.

CONTINUED ON PAGE 30

Carl Friedrich Gauss

Nothing can be considered completed if anything remains to be done.

—Gauss

by S. Gindikin

IN 1854, GAUSS' HEALTH took a turn for the worse, and the Privy Councilor, as Gauss had been dubbed by his colleagues at the University of Göttingen, could no longer take the daily strolls from the observatory to the literature museum that had been part of his routine for twenty years. The professor, who was approaching his eightieth year, was finally persuaded to consult a doctor. In summer, he felt better and even attended the opening of the Hannover-Göttingen railway. In January of 1855, Gauss agreed to sit for the artist Heseman for a medallion. After Gauss' death in February of 1855, a medal was struck in his honor from this medallion, with the inscription *Mathematicorum princeps* (Prince of Mathematicians) under his bas-relief. The history of every real prince begins with a childhood surrounded by legends. Gauss was not an exception.

Brunswick, 1777–1795

Gauss didn't inherit his title, although mathematics was not altogether alien to his father, Gerhard Diederich. A jack of all trades, he was an expert on fountains and also worked as a gardener. He was also known for his art in calculating. During fairs in Brunswick and even Leipzig, merchants used his services, and he had a steady job at the largest mortuary in Brunswick (a job which he would hand down to his son by his first marriage, Georg, a retired soldier).

Carl Friedrich was born on the 30th of April, 1777, in house 1550 on the Wendegraben canal in Brunswick. It is believed that he inherited good health from his father and outstanding intellect from his mother. His uncle Friederich, a skillful weaver, was closest to the future scientist. According to Gauss, his uncle was "a born genius." About himself, Gauss said that he could count before he could talk. The earliest mathematical legend about him tells that at the age of three he watched his father settling accounts with some masons who had been working for an hourly wage, when suddenly he corrected his father and turned out to be right.

At the age of seven, Carl Friedrich entered Catherine's School. In that school students were not taught how to count until the third grade, so for the first two years nobody paid attention to little Carl.

The children usually got to the third grade at the age of 10 and stayed in that grade until confirmation (at the age of 15). The teacher Büttner had to devote himself simultaneously to children of different ages and knowledge. For this reason, he often gave some of the students long exercises in calculation in order to be able to talk to other students. Once, he asked a group of students, among them Gauss, to sum up all natural numbers from 1 to 100. As a student finished the calculations, he would place his slate on the teacher's desk. The order of the slates was taken into account when giving marks. Ten-year-old Gauss turned in his slate as soon as Büttner had finished assigning the task. To everybody's surprise, only Gauss' answer turned out to be correct. The explanation was simple: as the teacher had been dictating the task, Gauss found a trick for summing a general arithmetic progression! The fame of the infant prodigy spread all over Brunswick.

A certain Bartels was the teacher's assistant at the school. His main duty was to sharpen pens for junior schoolboys. Bartels was interested in mathematics and had some mathematics books. Gauss and Bartels began to study mathematics to-

Art by Vadim Ivanyuk

$a_1 = \frac{a_0+b_0}{2}$
$b_1 = \sqrt{a_0 b_0}$
$a_2 = \frac{a_1+b_1}{2}$
$b_2 = \sqrt{a_1 b_1}$
$\frac{2\pi}{M(\sqrt{2},2)}$
y
x
O
1777
1791
1795–1798
1801
1812
1816

gether. They got to know Newton's binomial theorem, infinite series, and so on.

It's a small world! Later, Bartels became professor of mathematics at the University of Kazan and would have a student named Lobachevsky.

In 1788, Gauss entered the Gymnasium. However, mathematics wasn't taught there—students studied classical languages. Gauss learned languages with enthusiasm and made considerable progress, and he even entertained the idea of becoming a philologist.

Gauss became known at Court. He was introduced to Carl Wilhelm Ferdinand, Duke of Brunswick. The boy came to the palace and amused the courtiers by his mastery of the art of calculating. Thanks to the Duke's sponsorship, Gauss was able to enter the University of Göttingen in October 1795. First, he attended lectures on philology and hardly attended mathematics lectures at all. However, this does not mean that he didn't study mathematics.

The favorite science of the greatest mathematicians

This was one of the numerous epithets that Gauss bestowed on arithmetic (number theory). By that time arithmetic had become a branch of mathematics rather than a collection of uncoordinated propositions and observations.

Later, Gauss would write: *"Most important, we are indebted to recent research—certainly not extensive, but deserving of great praise—by mathematicians such as Fermat, Euler, Lagrange, and Legendre, who have discovered an entryway to the treasury of this divine science, and shown us its riches."* However, the boy of Brunswick didn't know all this yet, and was rediscovering at an astonishing rate the facts which had taken his great predecessors many years to find. Here are some topics in which Gauss was interested at that time.

Gauss noticed that the remainders upon division of squares of integers by a prime p cannot take on arbitrary values. For example, if the divisor is 3, the remainder can only be 0 or 1; if the divisor is 5, the remainder can only be 0, 1, or 4. However, we may consider this situation from another standpoint. We can ask for which primes p do numbers n^2 exist such that when divided by p they give a remainder of q? For $q = 1$, any p will do ($n = p + 1$). However, some values of p can give a remainder of $p - 1$, while others cannot. For example, when the square of an integer is divided by $p = 3$ a remainder of 2 does not occur, whereas $p = 5$ can give a remainder of 4 (when $n = 3$). The values $p = 7$, 11, and 19 do not produce a remainder of $p - 1$, but $p = 13$ and 17 do give such a remainder. Observations suggest the following hypothesis: if $p = 4k - 1$, then no n^2 can give a remainder of $p - 1$, but if $p = 4k + 1$, such an n^2 does exist. Gauss didn't know that this hypothesis had been formulated by Fermat and proved by Euler.

"I have accidentally run across an amazing arithmetical fact. Since it not only seemed elegant in itself but also suggested that it was connected with other outstanding facts, I got down to proving it and finding the principles on which it is based. After I had at last achieved this goal, the beauty of these studies captivated me, and I could no longer do without them."

Gauss tried to determine the values of p for which there exist numbers n^2 giving remainders of $q = 2$, $q = 3$, and so on. For $q = 2$, he guessed that the matter hinged on the remainder upon division of p by 8 (Euler had failed to prove this fact, but it was proved by Lagrange). For $q = 3$, it depends on the remainder upon division of p by 12. The general proposition asserts that all primes p that have the same remainder upon division by $4q$ can either simultaneously give a remainder of q when divided into a certain n^2, or they cannot. Gauss called this proposition the "golden theorem"; now it is called the law of *quadratic reciprocity*.

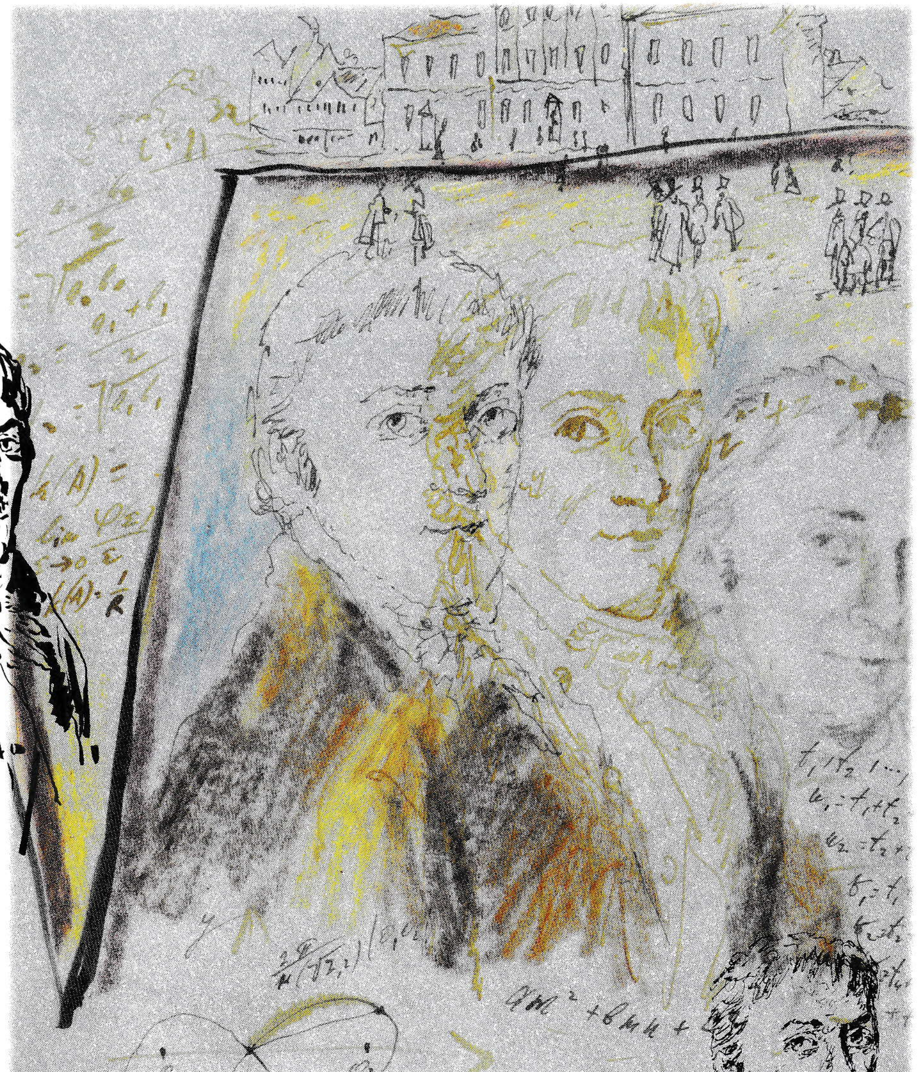
The "golden theorem" didn't yield to the first attack of young Gauss. He wrote that this theorem had tormented him an entire year but had refused to yield to the most intensive efforts. However, this was the point when Gauss had caught up with the mathematics of his time: every effort of the most prominent mathematicians to prove the law of quadratic reciprocity had failed.

Here is another topic of Gauss' studies. He noticed that if 1 is divided by p, the decimal digits repeat, which gives an infinite periodic decimal fraction. It is not difficult to prove the periodicity, but how can we find the length of the period? Gauss investigated prime numbers one after another and wrote out the corresponding periods. It seems to be a very tedious job (for example, for 97 the period is 96 digits long). However, Gauss investigated all $p < 1000$. He found that the length of the period always divides $p - 1$. (This fact can be derived from Fermat's "little" theorem, which Gauss proved independently.) Gauss was interested in those p for which the period is exactly $p - 1$. For this to be true, it is sufficient that the set of remainders upon division of 10, 10^2, ..., 10^{n-1} by p contain all the nonzero remainders. We don't know even now whether the number of such primes is infinite or not.

Gauss noticed that the set of remainders upon division of the numbers 3, 9, 27, ..., 3^{16} by 17 contains all of the possible nonzero remainders 1, 2, ..., 16. This observation stimulated the first great discovery made by Gauss—the construction of the regular 17-gon.

Gauss knew that for $n = 3$, 4, and 5, regular n-gons can be constructed with compass and ruler (or, equivalently, a circle can be divided into n equal parts). Certainly, this is also possible for $n = 2^k$, $2^k \cdot 3$, $2^k \cdot 5$, and 15 ($1/3 - 1/5 = 2/15$). Apparently, Gauss knew that the ancients had been unable to construct n-gons for any other n.

The mathematics of the modern era has made it possible to reduce the problem of construction of a regular n-gon to an algebraic problem. The possibility of constructing such a polygon with straightedge and compass reduces to representing the roots of the equation $z^n - 1 = 0$ in terms of quadratic irrationalities. If we ignore the root $z = 1$, we can limit our investigation to the equation $z^{n-1} + z^{n-2} + \dots + z + 1 = 0$, and we need to express these roots using integers, and repeatedly applying arithmetic operations and the operation of extracting square roots (but no other roots).



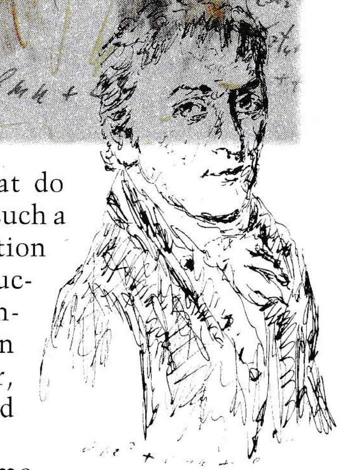
This reduction allows for a uniform consideration of all cases known to the ancients, while the ancients themselves had to find an original method for each individual case.

Gauss pondered over the *cyclotomic equation* (which is what the above equation is called), and simultaneously studied the divisibility of numbers. On the 30th of March, 1796, when awakening, he suddenly recognized the relationship between these two problems.

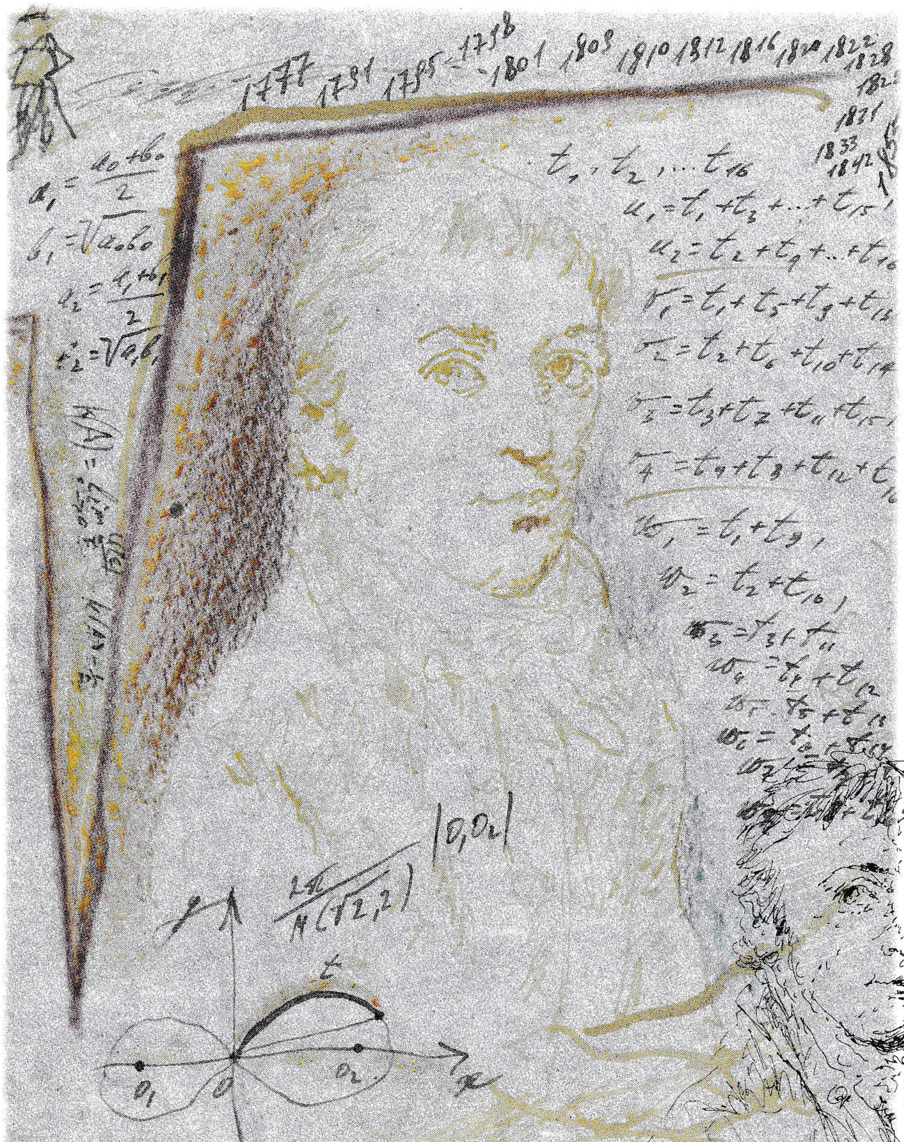
Let $n = 17$. If ε is a root of the equation $z^{16} + z^{15} + \dots + z + 1 = 0$, then $\varepsilon^2, \varepsilon^3, \dots, \varepsilon^{16}$ are its other roots. Gauss rearranged the roots in such a way that the root ε^l was assigned a number k if 3^k has remainder l upon division by 17. Thus each of the 16 roots was assigned a number. We denote the roots enumerated in such a way by $t_1, t_3, \dots, t_{16}$. Define $u_1 = t_1 + t_3 + \dots + t_{15}$, $u_2 = t_2 + t_4 + \dots + t_{16}$, $v_1 = t_1 + t_5 + t_9 + t_{13}$, $v_2 = t_2 + t_6 + t_{10} + t_{14}$, $v_3 = t_3 + t_7 + t_{11} + t_{15}$, $v_4 = t_4 + t_8 + t_{12} + t_{16}$, $w_1 = t_1 + t_9$, $w_2 = t_2 + t_{10}$, ..., $w_8 = t_8 + t_{16}$. It turns out that u_1 and u_2 are the roots of a quadratic equation with integer coefficients; v_1, v_2, v_3, and v_4 are the roots of the quadratic equation with coefficients that are expressed in terms of u_1 and u_2; $w_1, w_2, \dots, w_8$ can be similarly expressed in terms of v_1, v_2, v_3, and v_4; and finally, $t_1, t_3, \dots, t_{16}$ are expressed in terms of $w_1, w_2, \dots, w_8$. As a result, we obtain the desired representation of the roots of our equation.

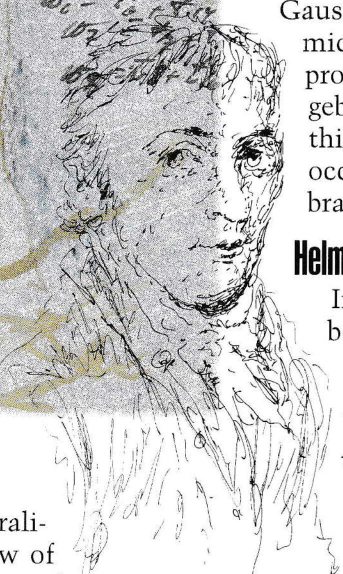
Thus, Gauss made progress in solving a problem about which nothing new had been done since Euclid. Since the formulation of the problem was quite elementary, this discovery was reported in the newspapers.

Later, Gauss gave a complete solution to the cyclotomic problem. It reduces to the case where n is prime. For primes n of the form $2^{2^k} + 1$ (these are the so-called Fermat's primes), the construction proceeds by the same scheme as for $n = 17$. The next n of this form is 257. There is evidence that Gauss performed the detailed construction for this case as well. Gauss also proved (but didn't publish his proof) that for primes that do not admit such a representation the construction is impossible (in particular, for $n = 7$ and 11).



It became clear that Gauss' destiny was to be a mathematician, not a philologist. In the later years of his life, Gauss recalled how at that time ideas had boiled up in his head. He hardly had time to make fragmentary records of them. Gauss began to keep a diary. The first record dated the 30th of March, 1796. It concerned the construction of the regular 17-gon. The second record (dated the 8th of April) shows that Gauss proved the law of quadratic reciprocity. We know from the diary that the great math-





ematician continued studying, accurately completing "student" exercises.

In 1798, Gauss graduated from the university and returned to Brunswick to systematize his results in number theory. In 1801, he published his famous *Disquisitiones Arithmeticae* (Arithmetic Investigations). This huge book (over 500 large-format pages) contained Gauss' main results: the law of quadratic reciprocity, the cyclotomic problem, and the question of representability of integers in the form $a^2 + bmn + cn^2$ (in particular, as a sum of two squares). The book was dedicated to the Duke, who also provided the means for publishing it. The printed version of the book contained seven parts (there was not enough money to print the eighth part). This eighth part was to contain a generalization of the law of quadratic reciprocity for powers greater than two, and in particular, the law of biquadratic reciprocity. A complete proof of this law was obtained by Gauss only on the 23rd of October, 1813. Gauss recorded in his diary that this event coincided with the birth of his son.

Aside from *Disquisitiones Arithmeticae,* Gauss did practically no more work in the field of number theory. He mainly systematized and completed what was conceived in those years. For example, he gave six more, proofs of the law of quadratic reciprocity. The *Disquisitiones Arithmeticae* was far ahead of its time. When writing this book, Gauss had been out of contact with contemporary mathematicians, and for a long time the book hadn't been available to German mathematicians. In France, where the book might have interested such scientists as Lagrange, Legendre, and others, a misfortune occurred. The bookseller who was to sell it went bankrupt, and more than half of the copies were lost. As a result, Gauss' students had to copy parts of the book by hand. The situation in Germany began to change only in the 1840s, when Dirichlet thoroughly studied *Disquisitiones Arithmeticae* and lectured on it. However, the book reached Bartels and his disciples in Kazan as early as 1807.

Disquisitiones Arithmeticae had a great impact on the development of number theory. Starting from Gauss' treatment of the cyclotomic problem, Galois studied the problem of the solvability of algebraic equations in radicals. To this day the laws of reciprocity occupy a central place in algebraic number theory.

Helmstadt dissertation

In Brunswick, mathematics books were scarce, so Gauss often went to Helmstadt, where there was a good library. In 1798, working in this library, Gauss wrote a dissertation devoted to the proof of the fundamental theorem of algebra, which asserts that any polynomial with complex (and in particular, with real) coefficients has at least one root (which in the general case is complex). If we do not want to go beyond the field of real numbers, the fundamental theorem of algebra can be formulated in the following way: any polynomial with real coefficients can be factored into a product of polynomials of the first and second degree. Gauss analyzed all previous attempts at proving this theorem and thoroughly implemented an idea of d'Alembert. However, the proof wasn't flawless, since the rigorous theory of continuity had not yet been developed. Later, Gauss suggested three more proofs of the fundamental theorem of algebra (the last one in 1848).

The Lemniscate and the arithmetic-geometric mean

In this section, we tell about yet another line of Gauss' research that began as early as in his childhood.

In 1791, when Gauss was 14, he played the following game of numbers: he took two numbers a_0 and b_0 and calculated their arithmetic mean

$$a_1 = \frac{a_0 + b_0}{2}$$

and geometric mean $b_1 = \sqrt{a_0 b_0}$. Then he calculated the means of a_1 and b_1:

$$a_2 = \frac{a_1 + b_1}{2}$$

and $b_2 = \sqrt{a_1 b_1}$, and so on. Gauss calculated both sequences with high accuracy. In a few steps, a_n and b_n were impossible to differentiate: all the decimal digits computed coincided. In other words, both sequences rapidly converged to a common limit $M(a_0, b_0)$, called the *arithmetic-geometric mean.*

At the same time, Gauss studied a curve called the *lemniscate* or the *lemniscate of Bernoulli.* This is a set of points such that the product of their distances to two fixed points O_1 and O_2 (*the foci*) is constant and equal to

$$\left(\frac{1}{2}|O_1O_2|\right)^2 \text{ (fig. 1).}$$

In 1797, Gauss began a systematic study of the lemniscate. He had been trying to find its length for a long time, until he guessed that it was equal to

$$\frac{2\pi}{M\left(\sqrt{2}, 2\right)}$$

(where M denotes the arithmetic-geometric mean defined above).

We don't know how Gauss guessed this. However, we know that he did it on the 30th of May, 1799. At first, Gauss had no proof of this formula, so he calculated both values to 11 decimal places! He also defined functions for the lemniscate similar to trigonometric functions of the circle. For example, for the lemniscate in which the distance between the foci is $\sqrt{2}$, the *lemniscate sinus* $\mathrm{sl}(t)$ is simply the length of the chord corresponding to the arc of

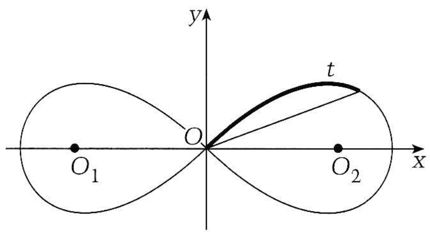


Figure 1

length t (fig. 1). Gauss spent the last years of the eighteenth century on the development of the theory of lemniscate functions. For them, he obtained addition and reduction theorems similar to those for trigonometric functions.

Then Gauss began to study elliptic functions, which are a generalization of lemniscate functions. He realized that it was a quite new branch of mathematical analysis. After 1800, Gauss didn't devote enough time to the theory of elliptic functions and didn't develop it to the degree that would satisfy him in completeness and rigor. From the very beginning, he decided not to publish intermediate results, hoping to publish everything in a final book as was the case with his arithmetic research. However, he never had time to carry out this plan.

In 1808, he wrote to his friend and student Schumacher: *"We can easily manipulate trigonometric and logarithmic functions; however, a magnificent golden spring that hides the secrets of higher functions remains almost* terra incognita. *I worked hard on this topic, and I am going to publish a major work, which I have already hinted at in my* Disquisitiones Arithmeticae. *One is amazed by the wealth of extremely interesting facts and relations presented by these functions."*

Gauss was sure that there was no need to hurry with the publication of his results—it had been that way for 30 years. But in 1827, two young mathematicians—Jacobi and Abel—published many of the results that had been earlier obtained by Gauss. He wrote:

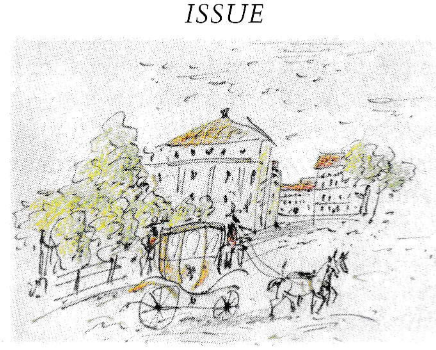
"The results obtained by Jacobi are a part of my own large work, which I hope to publish if the Almighty grants me strength and peace of mind." (A letter to Schumacher)

"By presenting the results with great rigor and elegance, Abel anticipated a lot of my own thoughts and facilitated my task by about a third. Abel went the same road as I did in 1798; thus it is not surprising that we obtained similar results. This similarity extends even to the form and sometimes to the notation, so that many formulas seem to be rewritten from mine. However, this fact shouldn't be interpreted in the wrong way, I don't remember a single instance when I discussed these questions with outsiders." (A letter to Bessel)

At last, in a letter to Crelle (May, 1828), Gauss wrote: *"Since Abel demonstrated such an insight and elegance in his presentation, I feel that I can refrain from publishing my results."*

It must be noted that a remark in *Disquisitiones Arithmeticae* that cyclotomy theory can be extended to the lemniscate had a great impact on Abel. He wrote: *"I had been pondering over these questions for a long time, and at last I was able to lift the veil of mystery over Gauss' cyclotomy theory. Now his reasoning is quite clear to me."*

CONTINUED IN THE NEXT ISSUE



Solving for the slalom

Once you understand the forces, it's all downhill from there

by A. Abrikosov

IN THE WORLD OF WINTER sports, downhill (or alpine) skiing, the luge, and the bobsled are in a class by themselves. Watching competitions in these sports, you may wonder if there is a reasonable way of descending a hill the fastest. Maybe the time of descent in such a competition is purely stochastic.

By now you have all solved the problem of a body sliding down an inclined plane (figure 1). The system of equations describing this motion yields the acceleration of the body at the start, the speed of the steady-state motion, and the duration of the motion from start to finish. Using this analogy we arrive at a paradoxical result: the time of descent doesn't depend on the actions of the participant. However, the skill of the participant has been omitted from these equations. Let's see how we might factor in the skill.

Before we begin, we should think about friction and air resistance. The coefficient of friction is determined by the choice of wax. The role of aerodynamics should not be underestimated for lugers and ski jumpers (and for the downhill and super G events, too). Not only is it vital to have the correct stance, but even the material and cut of the ski suit are important.

Art by Pavel Chernusky

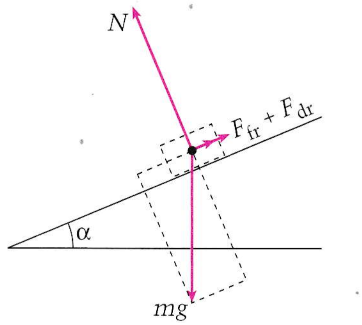


Figure 1. *A body on an incline:* $m\mathbf{g}$ *is the force of gravity,* $\mathbf{N}$ *the normal force,* $\mathbf{F}_{fr}$ *and* $\mathbf{F}_{dr}$ *the forces of friction and air resistance (drag). The equation of motion of the body is*

$$m\mathbf{a} = m\mathbf{g} + \mathbf{N} + \mathbf{F}_{fr} + \mathbf{F}_{dr}.$$

An impressive example is the legendary victory of the French downhill racing team. The French were the first to realize that wearing flapping numbers on their chests is a luxury at a speed of 100 km/h. Under a barrage of jokes they glued the numbers onto their ski suits. Nobody laughed when they were first at the finish line.

Nowadays hundredths of a second are at stake—so skaters put on "ultrastreamlined" suits, and downhill racers are even tested in wind tunnels. Bent ski poles help skiers take the optimal stance. New ski waxes and other coatings are tested, and novel alloys for bobsled runners are composed. The theft of a ski wax, described by the popular children's writer Leo Kassil in the novel *The White Queen's Move,* is child's play compared to the competition for Olympic gold.

However, the modern equipment of all the participants is more or less equivalent. The key role is the individual qualities of the athlete: a strong will to win, physical conditioning, and special training.

Let's see how the laws of mechanics can turn one's physical attributes and the desire for victory into precious seconds. In analyzing the principles of skiing we shall not discuss the high-speed downhill event, where the main roles are played by aerodynamics, control, and the choice of trajectory, but rather the slalom—a downhill race along a zigzag course. In such a race an athlete must succeed by his or her own efforts and at times display an almost acrobatic dexterity.

Forces

Let's return to the forces that affect a skier. As spectators, we observe the skier in an inertial (laboratory) reference frame fixed to the hill or the television camera. By contrast, skiers observe the world from their own noninertial reference frames, which are fixed to them. Although it is not a simple system for calculations, let's view the course from the skier's eyes.

Assume that the skier slides along an arc at a constant speed. In addition to the real forces such as gravity, friction, the normal force, and air resistance, there is another force in the skier's noninertial reference frame: the centrifugal force F_{cf}, which is directed from the center of the arc and is given by

$$F_{cf} = mv^2/R,$$

where v is the speed of the skier and R is the radius of the arc (figure 2).

The skier's center of mass is fixed in the moving reference frame. Therefore the sum of all forces applied to the skier (the resultant force) is zero at any point in time. Thus the reactive force of the snow must be tilted toward the center of the arc, since this is the only force that can counterbalance the centrifugal force F_{cf}. In bobsled the necessary tilt of the reactive force is provided by tilting the runners of the sled. Skis have metal edges to improve their bite on the snow. When turning, a skier turns the skis on edge to "grip" the snow—just as skaters do on ice. In the inertial (laboratory) reference frame the horizontal component of the reactive force imparts a centripetal (directed toward the center) acceleration to the skier. To provide a sure grip on treacherous hard-packed and icy slopes, the edges should be sharpened regularly, especially before competitions.

What loads act on a skier? Let's estimate them. The mean speed of a slalom racer is about 10 m/sec, and the radius (radius of curvature) of the arc is about 5 m; therefore $F_{cf} = mv^2/R \approx m \cdot 20$ m/sec^2, which is twice as large as the weight of the skier. This force should be added to the component of the force of gravity normal to the slope of the hill, which has a value of $mg \cos \alpha$ (usually $\alpha < 30°$, so $\cos \alpha > 1/2$). Thus the total load is larger than $2g$ and is applied predominantly to the "outside" leg (an attempt to "stand" on the inside leg usually results in a fall). The character of these loads is very similar to the loads produced by a vibrational testing machine. Now you see why professional downhill skiers train their muscles even in summer (for example, they do squats with weights).

Trajectory

Let's decompose successful racing into individual components. Why can't a skier be treated as a bead sliding along a smooth curved wire? First, because the skiers

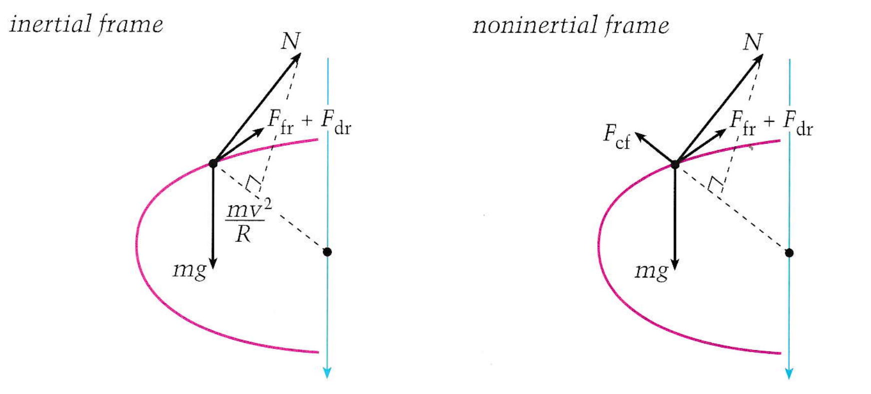


Figure 2. *Inertial and noninertial reference frames. A centrifugal force acts on the skier in the noninertial reference frame. Here and in the following figures the blue arrows denote the direction of the fall line.*

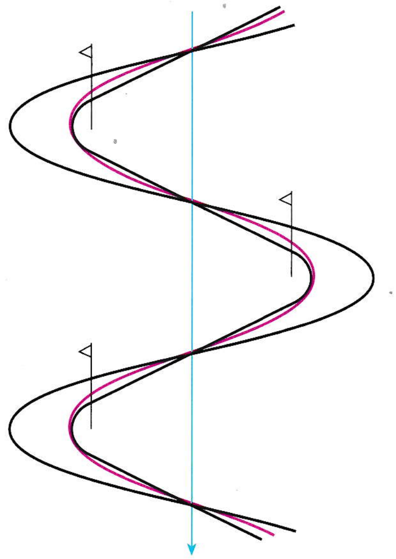


Figure 3. *Variants of a slalom track. The red line is the optimal track.*

choose their own paths within the corridor set by the flags (or by the chute in bobsled). From a physical viewpoint the slope of a hill is a two-dimensional space where a skier (even if considered to be a material point) must find the optimal trajectory. In contrast, the motion of a bead on a wire is one-dimensional. It may simulate bobsled racing (the most "one-dimensional" kind of racing), where the trajectory is more or less fixed and the racing time depends predominantly on how well the crew accelerates the bobsled at the start.

The optimal trajectory is determined by a combination of several factors. First of all, it is desirable to ski the shortest path, minimizing the deviations from the fall line (figure 3). Here the benefit results not only from the path length but also from an increase in the mean steepness of the trajectory: the steeper the slope, the greater the motive force and the smaller the friction. Therefore, slalom racers try to ski as close as possible to the flags: they even touch them with their shoulder or torso.

We can estimate the loss of time due to lengthening the path traveled. Let the deviation from the optimal curve be only 10 cm. A slalom course usually has 50 gates. At a mean speed of 10 m/sec, the lost time will be quite noticeable:

$$\Delta t \approx (50 \cdot 0.1\ \text{m})/(10\ \text{m/sec}) = 0.5\ \text{sec}.$$

By contrast, in the high-speed downhill event or in the giant slalom, where the number of gates is smaller and the mean speed is larger, small deviations from the optimum path are not very significant. However strange it may seem, to "race to the flag" (that is, to straighten the trajectory between the flags) is also disadvantageous. First, skiers must decrease their speeds to carve sharper turns. Second, the time of travel along a straight line is not necessarily the shortest.

Let's consider a simple example. A bead slides with zero initial speed from point A to point B along two trajectories: first along an arc, then along the subtending chord (figure 4). If the angular size of the arc is small, the duration of the motion in the first case is one-fourth of the period of oscillation of a simple pendulum of length R (we neglect friction):

$$T_1 = \frac{\pi}{2}\sqrt{R/g}.$$

The length of the chord is $l = 2R\sin(\theta/2)$, and the acceleration of the bead in the second case is $a = g\sin(\theta/2)$, so the corresponding duration of the descent will be

$$T_2 = \sqrt{2l/a} = 2\sqrt{R/g}.$$

Since $T_1/T_2 = \pi/4 < 1$, the bead that slides along the arc will be first to the finish.

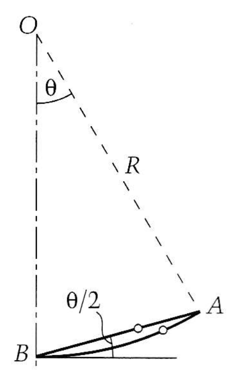


Figure 4. *Beads on wires. The time required to slide down along the arc* AB *is shorter than that along the subtending chord.*

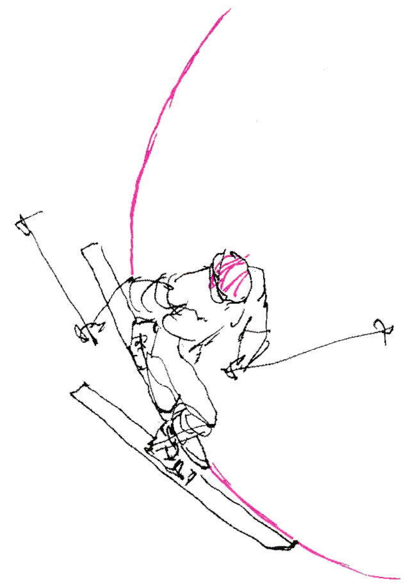


Figure 5. *Telemark turn: The skier takes a step forward with the outside ski and uses the arms for balance. However, modern downhill equipment doesn't allow one to perform this kind of turn.*

This is not a miracle: although the path along the arc is longer, it starts with a steeper slope. Therefore, the bead accelerates more rapidly, and in this case the advantage gained in speed is more important than the disadvantage suffered in total path length. We may guess that a trajectory composed of two smoothly joined arcs is also better than a trajectory with alternating drastic turns and straightened segments.

Even in his day Galileo was interested in the shape of the *brachistochrone*, which is the name given to the trajectory of most rapid descent from one point to another. He believed that the curve in question is a circular arc (as in our example). However, in 1697 Johann Bernoulli showed that in the absence of friction this "magic" curve is not a circular arc but a cycloid. The equation of the brachistochrone is used to design bobsled runs and roller coasters, but it is impossible to calculate the optimal trajectory of a slalom skier by purely theoretical means. Ski racers need intuition and experience. They must carefully examine the arrangement of flags on the course. As the famous French

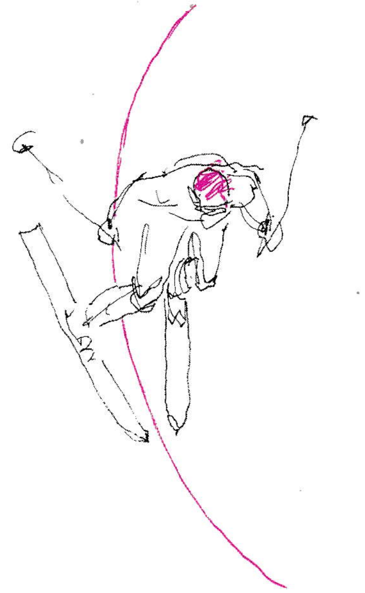


Figure 6. *Snowplow (or wedge) turn: Standing on the inside ski, the skier pushes against the snow with the outside ski.*

alpine skier Jean-Claude Killy advised, they must think five gates ahead.

Skis

In alpine skiing, as in many other sports, progress in sports technique goes hand in hand with improvement of the equipment. Just as one couldn't dream of 6-meter pole vaults before the invention of the fiberglass pole, so it was impossible to imagine the style and technique of modern alpine skiers with the old German hickory skis and "Kandahar" bindings.

Fast skis must not only slide with minimal friction, they must hold the slope and not slide in the transverse direction. Indeed, the formula for centrifugal force says that the speed in a turn is proportional to square root of the transverse reaction force F of the snow:

$$v = \sqrt{FR/m}\,.$$

However, this is not the only reason to avoid side-slipping, or skidding. More importantly, skidding takes energy — precious kinetic energy is spent scraping snow from the slope. Usually skiers employ skidding when they need to slow down in order to negotiate a steep place, where novices are not confident of

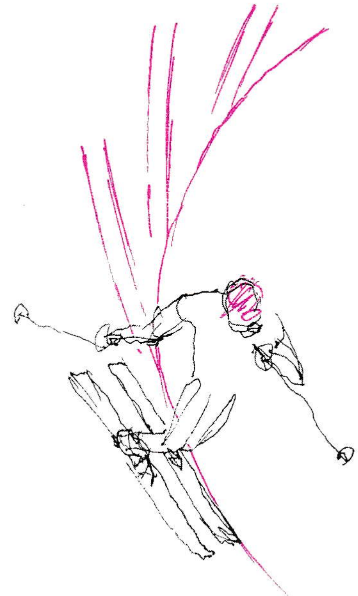


Figure 7. *A parallel turn performed by skidding (Christy). The parallel skis sweep less area in comparison with the snowplow turn, resulting in less braking.*

their skill, or just to smooth out the course after training. But if a skier skids on a turn in competition, the stopwatch will register it immediately. Now we see why skiers set their skis at as large an angle to the slope as possible and to cut the edges into the snow. (Nowadays this reveals the trademark of the manufacturer on the bottom of the skis.)

Downhill skiing owes its origins to Fridtjof Nansen, a great Norwegian polar explorer, politician, and Nobel Price winner, who was also the author of the first book on downhill skiing.

In the time of Nansen, skiers were equipped with soft leather boots attached to rigid skis without metal edges. The first tool of the downhill skier was the telemark turn (figure 5). This beautiful turn requires skill and carries the risk of a fall.

The telemark turn was replaced by the more popular wedge or snowplow turn (figure 6). Nowadays most skiers start out by learning this turn. It is the simplest way to perform a turn, but unfortunately it is also the slowest.

The "last word" in turns is the parallel turn, in which the skis are kept parallel. This turn was continu-

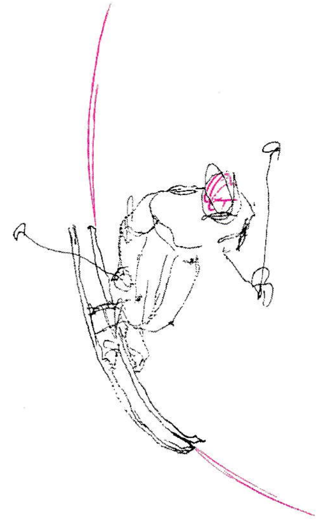


Figure 8. *A "carved" turn. The skis are bent and they do not scrape snow, allowing minimal braking.*

ally improved, and in time it firmly took hold as the main element in the technique of a downhill skier. At first, the parallel turn was performed with skidding (the Christy; figure 7). The inside ski is unloaded; the skier's body is a little bit forward, so the tip of the ski holds the snow better than the heel. Therefore the heels skid slightly in performing the turn. This turn was faster and safer than the previous ones.

Further progress in turning technique comes naturally. It is the so-called "carved" turn (figure 8), for which the resistance to motion is minimal. As speeds increased, the old requirements on ski equipment were replaced by just the opposite ones. Modern ski boots are made of rigid plastic with high tops. They transfer forces directly from the lower leg to the ski and permit good edging. The ski itself is flexible, so it can be bent along the arc to inscribe a turn, leaving a narrow carved track. To make it easier for the ski to assume an arc shape, it is made somewhat narrower just under the boot (figure 9). When edging, the skis hold the snow more strongly at their ends, so the pressure curves them in the proper way.

It is not easy to make good skis. A manufacturer must meet at least two conflicting requirements. First,

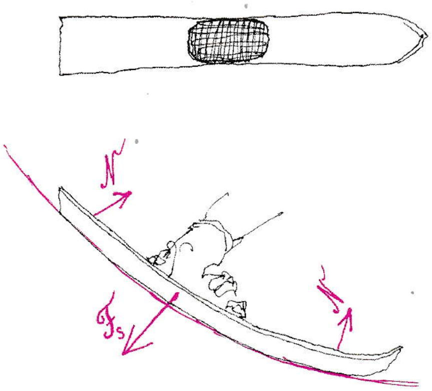


Figure 9. In order to make it easier for a ski to bend on the turns, it is made with a "waist," which is a small narrowing in the middle. When performing "edging," the pressure of the foot will bend the ski as needed.

the ski must bend elastically along an arc in the longitudinal direction. Second, to keep it from skidding, the ski should be rather rigid to resist twisting into a "propeller" shape. The skis should also "maintain direction" and not curve in the plane of the slope (that is, they must not curve into a saber shape).

To find a happy combination of just these two features alone is a very difficult problem for the manufacturers, and there is no clear-cut solution. In the standard slalom with its steeper turns but lower velocities, skiers use more flexible and shorter skis than those used in the giant slalom. Snow and weather conditions also make their demands. Therefore, just before the start of the race the future champion will choose a particular pair of skis from a set carefully prepared beforehand.

Fortunately, amateur downhill skiers do not need to worry about such things. Those special skis we see on TV represent only the very tip of the iceberg. While professional sports equipment is very demanding and unforgiving of minor mistakes, millions of simpler and more comfortable skis serve faithfully for the many amateur athletes who are the fans of this major sport.

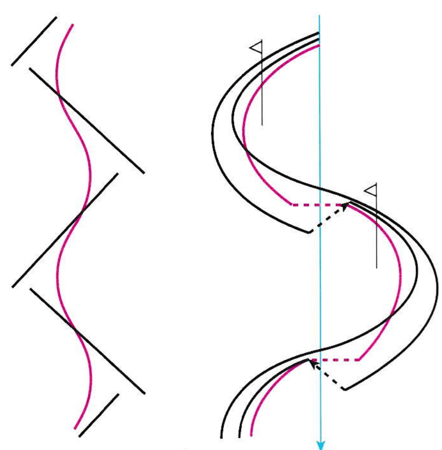
Some tricks

Now we consider the most important element, without which the technique of modern slalom and giant slalom is impossible. Although the basic physical principles of this element are clear, some of what we say may be controversial. Some specialists do not believe that downhill skiers can increase their speed by their own maneuvers, even though films of outstanding skiers prove it. The hidden possibilities of active skiing were demonstrated by Ingemar Stenmark at the very beginning of his fantastic career. The "Swedish Hurricane" outstripped his competitors by more than one second, while they were desperately fighting for mere tenths of a second.

What we have to say next may sound unbelievable to those who know how to ski. While recognizing the validity of their experience, here we shall describe the process from our own physical point of view.

Is there a clue to the phenomenon of acceleration in downhill skiing to be found in some other sport? Yes, in ordinary cross-country skiing! Nowadays the "skating" technique is very much in fashion (figure 10a). This technique has led to new racing records. Can slalom racers adapt it to their own needs?

The idea is not new. The skating technique was actually first tried in alpine skiing before it spread to the flats. On a hillside slope it looks like this: every turn is taken with a single push of the outside ski, so the tracks diverge slightly from the beginning to the end of the arc (figure 10b). The origin of the speed gain is obvious, but there are paybacks to be made. First, the skating technique requires very fine coordination of the legs and keeping one's balance at the moment when the load is shifted to the inside leg. Second, it prolongs the transition from one turn to the next. At the crossing points of the trajectory and fall line the skier changes one arc for another and carries his center of mass over the skis. Then comes the next turn. If the skis are wide apart, the transfer of center of mass will take longer, so the linear portion of the trajectory will grow. In addition, the skis will be "unedged" for a longer period, so the skier will be slowed by side-slipping. This effect manifests itself on trails where it is very difficult to hold on to the slope. Thus the skating technique is not the fastest way to go.



a b

Figure 10. *The skating technique on a flat site (a) and on a slope (b). The red curves are the trajectories of the skier's center of mass, while the black curves are the tracks of the skis.*

It would be ideal both to accelerate and to keep the skis closer together. Can this really be done? Let's analyze the skating technique once again (figure 10a). The center of mass of a cross-country skater traces a wavy trajectory (the red line in figure 10a). The athlete's body moves ahead of the supporting ski and at some angle to it. It is at this period, and not during the change of the supporting ski, that the skier performs work and gains speed.

Haven't we encountered this kind of motion previously? Yes, indeed. The slalom skier also has a zigzag motion, and his body doesn't slavishly follow the skis. At the end of a turn the skier's center of mass overtakes the skis, passes over them and goes ahead, that is, to the inside of the subsequent arc. Can a push be added to it? The answer is yes.

However, this push looks quite different from the skating technique and is rarely seen by the inexperienced eye. In reality, it is a particular type of motion that has two phases: *bending* and *unbending*. Initially, when the skiers come to the point at which the arcs are joined (that is, the point where the trajectory crosses the fall line and where the body overtakes the skis) they bend their knees as if absorbing the shock of a bump. This bending (sometimes it is rather abrupt) makes it possible to retain the speed gained and to avoid side-slipping. The knees are straightened immediately after passing the point where the arcs are joined, and this move pushes the body slightly forward, imparting an extra impulse in the direction of the fall line. At the end of the arc the skis again run ahead of the skier, and the cycle of bending (absorbing shock) and unbending (pushing off) is repeated. All this is performed on narrowly set skis.

By the way, when observing downhill skiers, you may have seen that in many cases the skiers take side steps from ski to ski. This doesn't contradict our reasoning: first, the ski course is not perfectly even; second, these small steps help them keep their balance, and third, even great masters can make a mistake, after all.

To clarify the physics of skiing, we now turn to quite another entertainment that is far from winter sports.

Summer analogy

Let's leave downhill skiing for a while and recall the good old summer time and large park swings. After passing the lowest point, the swing goes up with gradual deceleration. At the moment when it stops at the highest point, we squat down and rush earthward, the wind whistling in our ears. At the lowest point, where the overload is maximum, we stand up and again fly upwards with a fluttering heart—and this time we go a little bit higher than before. During this swinging, the center of mass of the system describes a "figure eight" (figure 11).

Increasing the amplitude in an oscillatory system due to changes of its parameters (in our swing it is the distance from the suspension to the center of mass) is called parametric resonance. By standing up at the lowest point, we perform positive work against the combined centrifugal and gravitational forces (considering the motion in the noninertial reference frame). At the uppermost point the centrifugal force is zero, and the only acting part of the gravitational force is $mg \cos \alpha$. Thus the negative work performed in squatting (with the same amplitude) is smaller in absolute value. The total work performed during a cycle is positive, so the energy of the system continually grows.

In a similar way we may estimate the energy balance of a downhill skier. Here a surprise is waiting. At first glance, everything is similar to what takes place on a swing. The centrifugal and gravitational forces also act on a skier who moves along an arc (figure 12). The angle between these forces varies, so the resultant force is minimal at the beginning of the arc (F_{r1}), while it attains its maximum value (F_{r2}) at the end of the arc. In squatting down, the skier performs negative work, while in standing up again he performs positive work. However, when he straightens up at the beginning of the arc, he is affected by a smaller force than when he bends his legs at the end of the arc. Thus the total work performed during the bending-unbending (or absorb-and-push) cycle is negative! Isn't this a paradox? It would make more sense for the skier to do positive work to increase his kinetic energy.

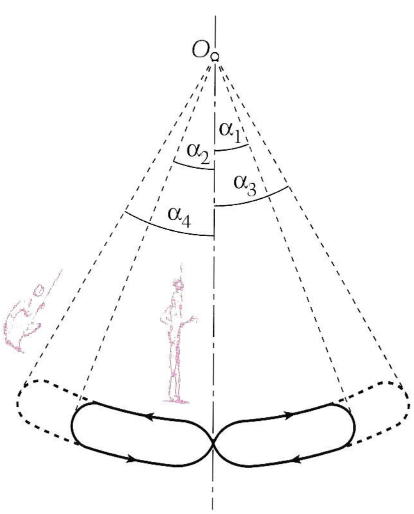


Figure 11. *Trajectory of the center of mass during swinging on a swing.*

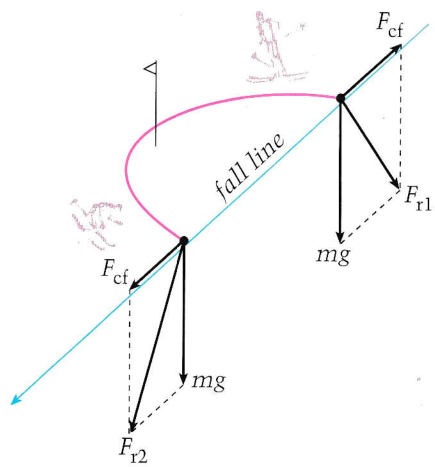


Figure 12. *The resultant of the gravitational and centrifugal forces at the beginning and the end of an arc:* $F_{r2} > F_{r1}$.

There is no mistake here—indeed, the skier performs work to damp speed! Up till now we haven't concerned outselves with the energy balance but have thought only about gaining speed and minimizing the losses. In this theoretical haste it is no wonder that we have stubbed our toe. Now it is time to dot the i's and cross the t's. We write the law of energy conservation on some portion of the path as

$$\Delta E = mg\Delta h + \frac{\Delta\left(mv^2\right)}{2} = W_{fr} + W_{dr} + W_{skier}.$$

On the left-hand side of this equation are the changes in potential ($mg\Delta h$) and kinetic ($\Delta(mv^2)/2$) energies of the skier, while on the right are the work performed by the skier, by friction, and by the air resistance or drag. What are the comparative values of these components?

Let's start our analysis on the left-hand side of the equation. The mean speed of a skier does not vary much along the course. Thus the second term is not particularly important, and we may drop it from the energy balance equation, that is, we set $\Delta(mv^2)/2 \approx 0$. By contrast, the first term is large. Indeed, to gain the speed typical for such an event ($v \sim 10$ m/sec), a slalom racer needs to descend only $\Delta h = v^2/2g = 5$ m. This is a tiny amount on the slalom courses where the drops in elevation are counted in the hundreds of meters.

Now look at the right-hand side of the energy equation. The braking of the skis during a carved turn is very small. The air resistance is larger, and it depends on speed. However, both of these decelerating forces will not prevent a skier from gaining a speed of about 100 km/h (28 m/sec). (Although the friction of the skis in a slalom turn is greater than in the downhill event, the difference is not enough to alter our conclusion.) At such a speed the flags along the course would look like a picket fence, and the course would be impassable. Thus, the main concern of a skier is not to gain but to lose speed! Therefore, the negative work performed by the skier in the bending-unbending cycle is a necessary condition to meet the requirement of energy balance:

$$W_{skier} < 0.$$

In order to win a skier must do work, and the negative sign by no means makes this work easier. We see that physics conforms to the rule: one must work hard to win.

Why is this mode of decreasing energy more efficient than the gradual energy dissipation during a skidding turn? First, the static loads are replaced by less tiring dynamic ones. Now one does not brace with all his might against the snow with the ski edges and lose speed at the end of an arc. Second, recall the example of the two beads on wires (figure 4). If the skier's energy is spent to overcome friction, the motion along an arc can be roughly considered as uniformly accelerated like that of a bead on a straight wire. When executing a turn with acceleration, a skier performs work and gains speed at the beginning of the arc and loses speed at the end of it. Consequently, the mean speed is greater, and a shorter time is needed to make a turn (just as for a bead on a wire). Thus, the bending-unbending mode of racing provides an additional way to control speed.

In contrast to bobsled and luge, downhill skiing is distinguished by freedom of movement and unique dynamic possibilities. In this respect it is similar to skateboarding. By the way, skateboarding is a good example of the fact that the "skating" technique is not the only way to accelerate oneself—where can one step when both feet are standing on a single skateboard? Acceleration gained by parametric resonance makes it possible to climb small elevations on a skateboard. Its working mechanism is the same: the body is inclined inside the arc, so bending at the beginning of the arc is followed by unbending at the end, thereby yielding a resultant impulse in the direction of motion. One should not be misled by the fact that a skateboarder accompanies bending with a vigorous swivel at the waist: that helps him hold the arc.

From the external observer's point of view, the main role is played by the normal (to the axis of the skateboard) component of the frictional force, which is similar to the component of the reactive force of the snow during edging of downhill skis. This force alternates in value and direction during a zigzag motion, but on average it is directed forward.

And so the whole secret is revealed. But do not think that you know everything there is to know about downhill skiing: our theoretical "sit down-stand up" model is greatly oversimplified. Swimming must be learned in water! There is no downhill skier who has not plowed a couple of large snowdrifts with his body.

Conclusions

Downhill skiing is accompanied by an incomparable feeling of wonderful freedom, with the whole world rushing toward you, sparkling with frosty snow. This can not be described by formulas. Small children learn to ski by imitating elders while knowing nothing of the physics of skiing. However, knowledge is power, and so Newton's laws show us another way to master skiing—from the head to the legs, so to speak.

There are some points we did not even touch on in our analysis. For example, a skier is affected not only by forces, but also by their moments. In other words, a skier is not a material point, but a rotating physical body. Perhaps in this avenue the lovers of downhill skiing will find the answers to some unsolved problems.

What if you are not a ski enthusiast? Now that you know something of skiing theory, wouldn't you like to apply your knowledge to some other sport? Why not try? As for me, if downhill racing is not happiness itself, it will do nicely as a substitute.

Quantum articles on friction, energy conservation, and parametric resonance:

I. Slobodetsky, "Squeaky Doors, Squealing Tires, and Singing Violins," November/December 1992, pp. 46–50.

J. Wylie, "Friction, Fear, Friends, and Falling," July/August 1992, pp. 4–8.

Y. G. Pavlenko, "The Satellite Paradox," March/April 1993, pp. 50–52.

A. Buzdin, "Keeping Cool and Staying Put," May/June 1993, pp. 17–20.

A. Chernoutsan, "Swinging Techniques," May/June 1993, pp. 64–65.

A. Eisenkraft and L. Kirkpatrick, "Pins and Spin," July/August 1995, pp. 34–36.

A. Chernoutsan, "A Gripping Story," March/April 1996, pp. 40–42.

A. Eisenkraft and L. Kirkpatrick, "Moving Matter," November/December 1996, pp. 32–33.

A. Chernoutsan, "Why Doesn't the Sack Slide?," May/June 1997, pp. 50–52.

L. Turner and J. Pratt, "Does a Falling Pencil Levitate?," March/April 1998, pp. 22–25.

A. Stasenko, "Hurling at the Abyss," November/December 1998, pp. 43–44.

HOW DO YOU FIGURE?

Challenges

Physics

P276

Dummy Earth. A dummy Earth is made for one of Spielberg's films. It has the same size and mass as the real Earth, but its construction is different: there is a small ball of extremely dense matter inside a very light outer plastic ball. Due to some inaccuracy during assembly, the center of mass of the heavy ball is shifted in the equatorial plane by a distance d = 100 km from the center of the outer shell. Find the minimal period of revolution of a satellite orbit in the equatorial plane. (A. Zilberman)

P277

Soapy bubble. A soap bubble is inflated with gaseous nitrogen. At what diameter will it float in atmospheric air of the same temperature? The surface tension of the soap solution σ = 45 mN/m, the molar mass of air M_a = 29 g/mol, the molar mass of nitrogen M_N = 28 g/mol, and the atmospheric pressure $p_0 = 10^5$ Pa. Neglect the mass of the soap film. (A. Sheronov)

P278

Electrical sandwich. A plane or sandwich capacitor consists of three parallel metal plates of area S. The space between the plates is filled with dielectrics characterized by dielectric constants κ_1 and κ_2 and resistivities ρ_1 and ρ_2. The thicknesses of the dielectrics are d_1 and d_2. The capacitor is connected to a constant voltage source V. Find the charge on the middle plate when the current in the circuit has reached its steady state.

P279

Toroidal transformer. Three identical coils are wound symmetrically on a toroidal core. One of the coils is connected to an AC source, the second is left open, and the third is connected to a voltmeter. In this case the voltmeter reads 1/2 of the source voltage. What will it read if the second coil is short-circuited? Neglect the resistance of the coils, and consider the voltmeter and the voltage source to be ideal. The magnetic permeability of the core doesn't depend on the magnetic flux. (A. Andrianov)

P280

News and views. You are given a prism with the cross-section shown in figure 1 and index of refraction n. $\angle ABC$ is a right angle and $\angle BAC = \angle ACB$. Faces AC and BC are smooth and face AB is frosted (figure 1). The prism is placed on a newspaper with face AC down. What fraction of the text (by area) can be read through the smooth face BC? (S. Gordyunin, P. Gorkov)

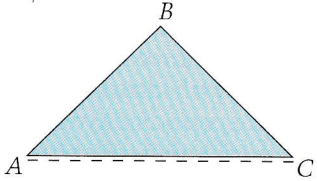


Figure 1

Math

M276

Pyramidal structure. Five edges of a triangular pyramid are of length 1. Find the sixth edge if it is given that the radius of the sphere circumscribed about this pyramid is 1.

M277

Algebra quiz. Solve the following system of equations:

$$\frac{x-1}{xy-3} = \frac{y-2}{xy-4} = \frac{3-x-y}{7-x^2-y^2}.$$

M278

Trigonometry quiz. Solve the equation

$$|\cos 3x - \tan x| + |\cos 3x + \tan x| = |\tan^2 x - 3|.$$

M279

Abacus workout. Without using a calculator, find which of the following numbers is greater: $29^{200}2^{151}$ or $5^{279}3^{300}$.

M280

Smallest angle. Let M be the midpoint of side BC of a triangle ABC and let Q be the point of intersection of its bisectors. It is given that $MQ = QA$. Find the minimum possible value of angle MQA.

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

Returning to a former state

by A. Savin

IN MY CHILDHOOD, I WAS fond of the kaleidoscope. Look into the magic tube and you see a magnificent mosaic. Rotate the kaleidoscope a little, and a new pattern appears; another turn brings yet another pattern, and so on.

When I first saw Rubik's Cube, I just rotated it as I would a kaleidoscope to admire the play of colors on its faces. Soon, I was tired of aimlessly rotating it and tried to bring the cube back to its original arrangement. After several hours of unsuccessful attempts to find an algorithm to bring the cube into order, I looked up a magazine article and, in about an hour, learned how to do it.

Then I began to rotate the cube more carefully so as to be able to return it to its original state. I rotated a face once, then three more times (fig. 1), and the cube returned to its original state.

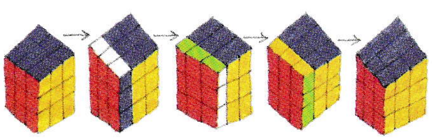


Figure 1

What if we rotate two neighboring faces in turn, say, in the same direction? First face, second face—the first pair of operations; first face, second face—the second pair; first face, second face—the third pair; and so on (fig. 2). At the time, I already knew that the cube would return to its original state after a certain number of pairs of rotations. My fingers grew tired, I lost track of the count, and still the cube wouldn't return to its original state. Only after 105 pairs of rotations did the colors come back into order.

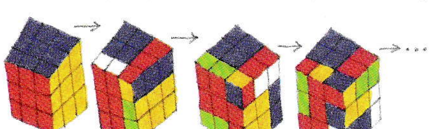


Figure 2

Why was I so sure that this moment would necessarily come? I just thought about the problem a bit. The cube has only a finite number of states, although this number is very large. Therefore, it cannot come to a new state every time—sooner or later, the cube will come to a state that has already occurred. We call this state *A*, and the original state of the cube will be called *E* (fig. 3).

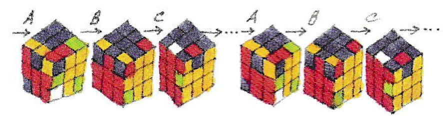


Figure 3

A pair of rotations brings the cube to a state *B*, then from the state *B* to a state *C*, etc. until it comes to the state *A* again; then everything repeats. Mathematicians say that *the states of the cube repeat periodically*. We now formulate this result for objects of any sort, rather than only for Rubik's Cube.

Periodicity Theorem: *Assume that an object can be in a finite number of states and an operation is defined that unambiguously brings each state to another state. Then, successive application of this operation gives a periodically repeated sequence of the object's states.*

This sounds like a law of nature or a mathematical theorem, and indeed it is. Moreover, this proposition makes it possible to obtain very interesting results.

However, let us first complete our reasoning concerning Rubik's Cube. Notice a property of our operation: for any state the previous state can be unambiguously determined. To get it, one must rotate first the second face and then the first face of the cube in the opposite direction as before. Look at figure 4.

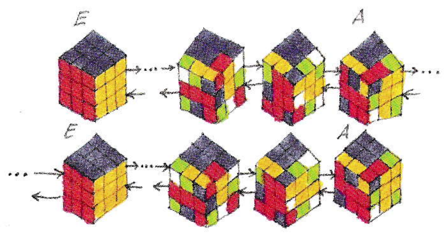


Figure 4

Moving backward from the state *A* obtained for the first time, we come to the state *E* in a certain number of steps. Therefore, from the state *A* obtained for the second time, the same number of backward steps brings us to the state *E*. Thus between the first and the second states *A* the cube must necessarily "visit" state *E*.

Such was my reasoning before I began to rotate the cube according to the above rule. Of course, the number of pairs of rotations could be very large, since the total number of states of Rubik's Cube is huge—43 252 003 274 489 856 000. I tried rotating the faces in two different directions—one face clockwise and the other face counterclockwise. In this case, the cube returns to its original state after 63 pairs of rotations; if three faces are

Art by Vadim Ivanyuk

rotated in turn in the same direction, the cycle terminates in 80 triples of rotations. If you have a Rubik's Cube at hand, you can experiment with other combinations of rotations. If not, take a sheet of paper and a pencil, and start experimenting with numbers. A calculator or personal computer can be of use; however, we can do without them.

Repeating digits

Let us divide 136 by 11. By the way, do you know a test for divisibility by 11? Here it is: *a number is divisible by* 11 *if the difference between the sum of the digits in its even decimal places and the sum of digits in its odd decimal places is divisible by* 11. In our case, $(1 + 6) - 3 = 4$, and therefore 136 is not divisible by 11. What if we try to divide 136 by 11 (fig. 5)? It is easily seen that

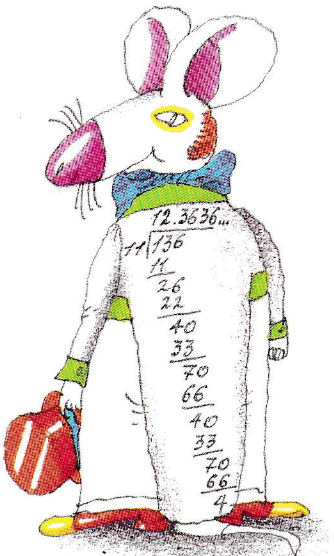


Figure 5

digits after the decimal point repeat periodically. Such a decimal fraction does not terminate and is called an *infinite periodic decimal fraction.* It is written as $12.\overline{36}$, where the group of repeating digits is written in parentheses.

What other integers give periodic decimal fractions when divided by other integers? In fact, any integers will, provided that the divisor is not zero (we can consider any finite periodic decimal fraction as an infinite one by continuing it with zeros after the last decimal place). Why is this so? To prove it, we will use the Periodicity Theorem formulated above. Consider the difference under the bar as the object that appears in the theorem. In our case, this difference is 2, then 4, then 7, then 4 again, 7 again, etc. Consider the operation that is performed on this object. First, we add a digit to the right of it, then subtract the maximum multiple of the divisor not exceeding this new number, and the difference obtained is the result of the operation.

At first, the digit that is added on the right depends on the dividend; however, after a certain number of steps (in our case, beginning with the second step), the digit that is added on the right is always zero. Now the conditions of the proposition are satisfied: the object (the difference obtained when we subtract) can be in a finite number of states (nonnegative integers less than the dividend). The operation is well-defined: it takes a number to another number. Therefore, the sequence of numbers obtained will repeat periodically. Therefore, the digits in the quotient will also repeat periodically. Thus, the proposition is proved.

Repeating sums

What if we take the sum of the digits of a number, then the sum of the digits of the number thus obtained, and so on? For example, for the number 1987, everything is clear:

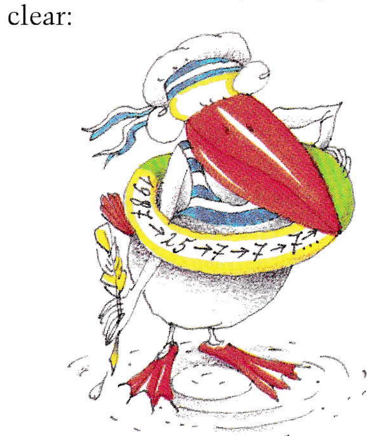


Figure 6

What about other numbers? It is easily seen that any number is greater than the sum of its digits (except for one-digit numbers); thus the operation of taking the sum of digits makes the number smaller until it becomes a single digit, and our operation takes any one-digit number to itself. Therefore, we have proved that the sequence thus obtained is periodic (with period 1), beginning at a certain place.

What is the digit that is repeated in this sequence? This question is easy to answer without writing out the sequence. *If the initial number is not divisible by 9, then it is the remainder upon its division by 9, and otherwise it is 9.* Why is this so? In fact, because of the test for the divisibility by 9. However, you most probably know this test in a "shortened" form: *in order for a number to be divisible by 9, the sum of its digits must be divisible by 9.* However, a stronger proposition holds: *any number and the sum of its digits have the same remainder upon division by 9.*

Now it is clear that every number in the sequence has the same remainder upon division by 9 as does the initial number. All one-digit numbers, except for 0 and 9, have different remainders upon division by 9. However, the sum of the digits of any nonzero number is greater than zero, and so our proposition is proved. A periodic sequence with period 1 is called a sequence with a *fixed point.*

Squares of digits

What if we consider the sum of squares of the digits of a number rather than the sum of the digits themselves? Again, consider the number 1987 (fig. 7). We obtain the

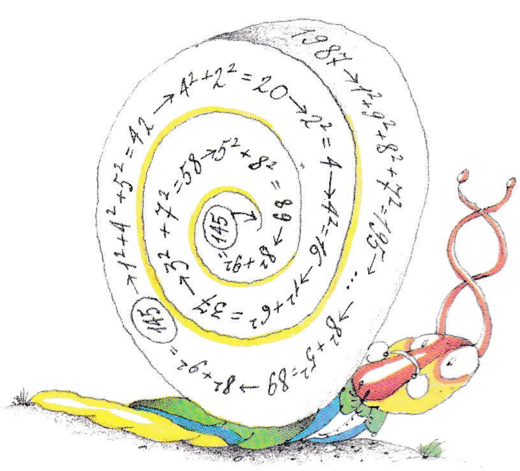


Figure 7

following periodically repeating sequence: 145, 42, 20, 4, 16, 37, 58, 89.

For the number 133, we obtain a sequence with a fixed point of 1 (fig. 8).

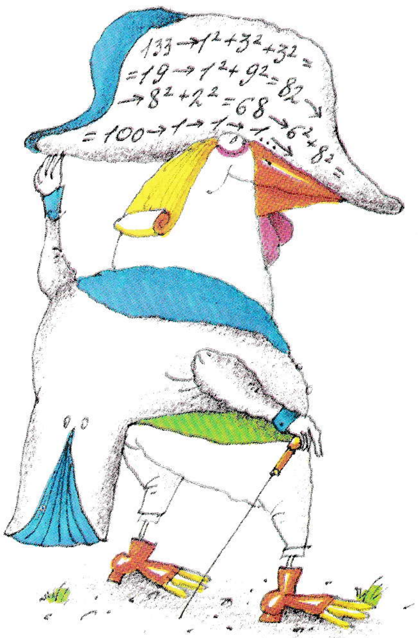


Figure 8

Does the operation of taking the sum of squares of the digits always give a periodic sequence? If yes, then what sequences can be obtained?

For the operation under consideration, the following assertions hold:

1. *Any number less than* 200 *goes into a number less than* 200.

2. *Any number greater than or equal to* 200 *(or even* 100*) becomes smaller under this operation.*

The first assertion is easily verified. Indeed, among the numbers less than 200, 199 has the maximum sum of squares of the digits, and this sum is 163, which is less than 200. Thus, for all other numbers less than 200, the sum of squares of the digits is also less than 200.

The second assertion is also easy to see (it is sufficient to test several big numbers to satisfy yourself that it is true). Below, we give a rigorous proof of this fact.

Assume that a certain number $N = a_n \ldots a_1 a_2 a_0$ is less than the sum of squares of its digits:

$$a_n \cdot 10^n + \ldots + a_2 \cdot 10^2 + a_1 \cdot 10 + a_0 < a_n^2 + \ldots + a_0^2.$$

Then,

$$a_n(10^n - a_n) + \ldots + a_2(10^2 - a_2) + a_1(10 - a_1) + a_0(1 - a_0)\ 10^n - a < 0$$

In this sum, all terms, except for the last one, are nonnegative, and this last term is greater or equal than $9(1 - 9) = -72$. If at least one digit a_n, for $n \geq 2$, is nonzero $(1 \leq a_n \leq 9)$, then

$$a_n(10^n - a_n) < 10^n - 9 \geq 91,$$

and the total sum is positive. Therefore, $n \leq 1$ and $N < 100$.

It follows from Propositions 1 and 2 and the Periodicity Theorem that the operation of taking the sum of squares of the digits of a number necessarily leads to a periodic sequence.

It remains for us to find which periodic sequences can be obtained. It turns out that whatever the initial number may be, this sequence will be either 145, 42, 20, 4, 16, 37, 58, 89 or the "singleton" sequence 1. Verify this fact for yourself. It is sufficient to take numbers less than 200 as the initial numbers (why?) or even the numbers less than 161.

What if we consider the sum of the cubes of the digits rather than the sum of their squares? Try to solve this problem yourself.

In conclusion, let us consider an unsolved problem. Take a number. If it is even, divide it by 2. Otherwise, multiply it by 3 and add 1. Again, if the number obtained, N, is even, take $N/2$, and otherwise take $3N + 1$, and so on. For the number 34, this process is illustrated in fig. 9. It gives a periodic sequence 4, 2, 1, 4, 2, 1, ... with period 3.

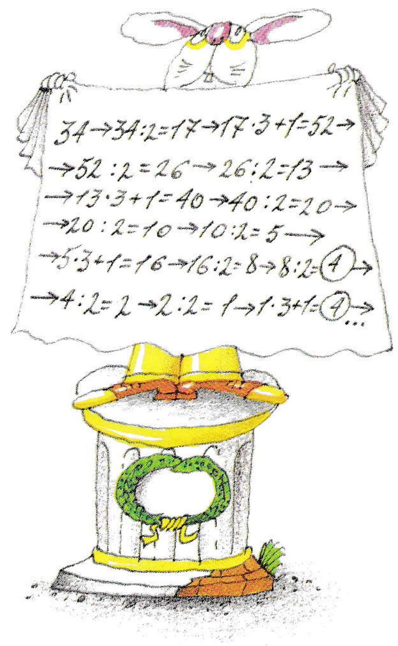


Figure 9

Various initial numbers have been tried, and all of them lead to the same sequence. However, this fact has not been proved so far.

We don't hold much hope that any of our readers will solve this problem. However, try to solve the following similar problem: *if the number is even, divide it by* 2, *and if it is odd, add* 101 *to it.* What sequences can be obtained? ◘

CONTINUED FROM PAGE 13

Now it turns out that only individualized local information vanishes, but the Universe remembers everything. The universal "coherent" memory is not only preserved but is continuously being enriched as long as history goes on. The transition from past to future looks more like the formation of a hologram than some permutations of persistent objects.

However, no analogy can be strictly precise. Having managed to solve some old physical puzzles, we have obtained new ones! This is the beauty of science, isn't it? ◘

Quantum on thermodynamic laws and quantum mechanics:

A. Buzdin, "Keeping Cool and Staying Put," May/June 1993, pp. 17–20.

S. R. Filonovich, "The Power of Likeness," September/October 1991, pp. 23–27.

Kaleidoscope: "What a Commotion," May 1990, pp. 32–33.

A. Savin, "Mathematics in Perpetual Motion," July/August 1994, pp. 5–8.

A. Stasenko, "Not All Is Revealed," November/December 1996, pp. 42–44.

New CONTEST FORMAT NAMES

NEW CONTEST, NEW FORMAT, NEW NAMES

The American Mathematics Competitions is pleased to announce a new contest as well as new names and a modified format for two of the current exams. The AJHSME is now the American Mathematics Contest ⇨8 (AMC ⇨8) and the AHSME is now the American Mathematics Contest ⇨12 (AMC ⇨12). The new contest is the American Mathematics Contest ⇨10 (AMC ⇨10), for students in grades 10 and below. This new contest will give more young students a chance to successfully participate in a significant mathematical problem solving experience.

Why should my school sign up? Because the AMC ⇨10 and AMC ⇨12 provide an excellent opportunity to challenge your students' mathematical abilities. It is but a means for furthering mathematical interest and development.

The AMC ⇨10 and AMC ⇨12 will each be 75 minutes long and will consist of 25 questions each. Each correct answer is worth 6 points and a blank is worth 2 points. The AMC ⇨10 and AMC ⇨12 will have several questions in common and will be given at the same time, on the Tuesday before the third Monday in February (the current AHSME date). The students should choose between AMC ⇨10 and AMC ⇨12. Students in 10th grade and under may take either the AMC ⇨10 or AMC ⇨12, but 11th and 12th grade students may not take the AMC ⇨10. The school team score will be determined from the AMC ⇨12. To qualify for the AIME a student must score at least 100 points on the AMC ⇨12 or be in the top 1% of the AMC ⇨10 participants.

The registration fee for one or both contests is $30.00. One bundle of ten AMC ⇨12 is $12.00 and one bundle of ten AMC ⇨10 is $10.00. The first bundle of the AMC ⇨10 will be free for the year 2000 only.

<u>2000 AMC exam dates:</u>

AMC⇨8 - TUESDAY, November 14, 2000
AMC⇨10 - TUESDAY, February 15, 2000
AMC⇨12 - TUESDAY, February 15, 2000
AIME - TUESDAY, March 28, 2000
USAMO - TUESDAY, May 2, 2000

www: http://www.unl.edu/amc

AMC

A Program of the
Mathematical Association of America

Titu Andreescu, Director
American Mathematics Competitions
University of Nebraska, Lincoln
P.O. Box 81606
Lincoln, Nebraska 68501-1606
phone: 800-527-3690
fax: 402-472-6087
e-mail: titu@amc.unl.edu

A question of complexity

by Arthur Eisenkraft and Larry D. Kirkpatrick

> *Man is an overcomplicated organism. If he is doomed to extinction he will die out for want of simplicity.*
> *—Ezra Pound*

WHAT WOULD HAPPEN IF a volcano erupted under a glacier? Can we solve such a problem? Can we do it with elementary physics? At the 1998 International Physics Olympiad (IPhO) in Iceland, the competitors were asked to solve this problem using data from an eruption that occurred in Iceland a year or two before the Olympiad took place.

Many problems at the IPhO are presented in the context of problems that physicists solve to understand the physical world. Geophysicists can generate rather elaborate three-dimensional computer codes to analyze the volcanic event, but they can get a "feeling" for what happens by simplifying the problem. For instance, they can simplify the geometry by assuming that the hot lava melts a conical cavity in the ice. They also assume that thermal conduction away from the cavity is small and that the water does not flow away. And this is what was presented to the competitors. Solving the simplified problem did not require any knowledge of physics beyond what is typically learned in the first-year university physics sequence.

What happens when you are given a problem that is either too complex to solve or one where there are crucial ingredients missing? We have observed three types of students. The first mentions that there is a piece of missing information (for example, we don't know the angle between the particle and the field) and refuses to solve the problem. The second mentions that we don't know the angle but goes on to assume that it is 90° and solves the problem. The third assumes that the angle is 90°, solves that problem, and then describes how the solution changes if the angle were not 90°. It's the third student who has the flair of a physicist.

Many problems in the real world are very complex and cannot be solved if all the complexity is included. If, however, we are able to simplify the problem without losing the key elements, we can often reduce a complex problem to a problem (or a series of problems) that we already know how to solve.

At other times a problem may appear to be very complex because of the context in which it is presented. When we were academic directors of the US Physics Team, students would comment that the difference between the more difficult problems in an introductory physics text and problems given at the International Physics Olympiad is that you only realized that Olympiad problems are easy after you have solved them!

This is nicely illustrated by the first problem on the theoretical exam given at the IPhO that was held in Padua, Italy, in July. (See *Happenings* for a report on the success of the US Physics Team at the IPhO and a description of the other problems.) The text of this problem runs for an entire typewritten page, but can be summarized as follows: A vertical cylinder filled with gas and capped by a moveable glass plate is illuminated for a finite time by a laser. As the gas absorbs the light, the glass plate is observed to move upward. The competitors were then asked a series of quantitative questions about the situation.

What possible assumptions about the physics could provide a pathway from complexity to simplicity? Let's begin with the friction between the glass plate and the cylinder walls. If we know nothing about the friction, we assume that the friction between the glass plate and the cylinder walls is sufficient to damp any oscillations that occur but not large enough to produce any significant loss of energy relative to the other energies involved in the problem.

What assumptions do we usually make when solving gas problems? We first assume that the gas is in thermal equilibrium. Unless otherwise stated, we usually assume that the gas is ideal and that the amount of gas is constant. In this case we assume that the cylinder does not leak gas around the piston. This allows us to use the ideal gas law

Art by Tomas Bunk


LAOCOON 2000
TOM BUNK
OIL

$$PV = nRT,$$

where P is the pressure, V is the volume, n is the number of moles of gas, $R = 8.31$ J/K · mol is the universal gas constant, and T is the temperature in kelvin.

Then we need to decide whether the system is thermally isolated, that is, can thermal energy enter or leave the cylinder? We have previously solved problems of both types, so if this is not explicitly stated, we will need to infer this from the context of the problem.

If we are not given any information about the coefficient of thermal conductivity between the cylinder walls and the glass plate, we simplify the problem by assuming a very low thermal conductivity and/or a very short time so that thermal losses can be neglected. In this problem, the competitors were told that the cylinder walls and the glass plate had very low thermal conductivities.

The detailed text told the competitors that light from a constant power laser was shined through the glass plate into the cylinder for a specified time interval. The radiation passed through the air and the glass plate without being absorbed but was completely absorbed by the gas in the cylinder. The molecules absorbing the radiation were excited to higher energy states and then quickly cascaded back to their ground states by emitting infrared radiation. This infrared radiation was reflected by the cylinder walls and the glass plate and absorbed by other molecules.

What was this telling the competitors? Independent of the details, the gas was being heated at a constant rate, and this energy increased the average kinetic energy associated with the chaotic motion of the molecules.

The data often provides other hints about how to simplify. In this problem, competitors were given the following values:

Atmospheric pressure:
$P_0 = 101.3$ kPa
Room temperature:
$T_0 = 20.0$°C
Inner diameter of the cylinder:
$2r = 100$ mm
Mass of the glass plate:
$m = 800$g
Quantity of gas:
$n = 0.100$ mol
Molar specific heat at constant volume:
$c_V = 20.8$ J/(mol · K)
Wavelength of the laser:
$\lambda = 514$ nm
Irradiation time:
$\Delta t = 10.0$ s
Displacement of glass plate:
$\Delta s = 30.0$ mm

Even though the initial temperature of the gas is not given in the data table, the problem stated that the gas was initially in equilibrium with its surroundings. Therefore, the initial temperature of the gas is the same as the room temperature.

The pressure of the gas is not given. However, we've done piston problems before. Because the glass plate is in equilibrium, the force on the lower surface must exceed that on its upper surface by the weight mg of the glass plate. Moreover, this must be true after the heating as well as before. That is, the initial and final pressures are the same.

We are left with another problem that we have solved in other simpler contexts. What is the increase in temperature of a gas kept at constant pressure when the volume increases by a specified amount? The wrinkle in this problem is that we need to calculate the initial height of the piston first.

Notice that the power of the laser is not given. Therefore, we need a relationship between the properties of the gas and the energy added to the gas. This is provided by the first law of thermodynamics,

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

where U is the internal energy of the gas, Q is the heat added to the gas, and W is the work done by the gas. To use this law to obtain the heat Q, we need to know the other two quantities. Calculation of the work is straightforward, but where are we to find the change in internal energy of the system? Looking at the data gives us a hint. Because we are given the molar specific heat at constant volume for the gas, we are prompted to recall that

$$\Delta U = nc_V(T_f - T_0).$$

When we are asked about the number of photons emitted by the laser per second, we must remember that the laser beam consists of photons with energy hf and hope that the beam has a single frequency. According to the data it does.

In this problem absorption of optical energy produces a change in the gravitational potential energy of the glass plate. How efficient is this process? The word "efficiency" brings to mind its definition:

$$\eta = \frac{W}{Q},$$

where W is the part of the work done to increase the gravitational potential energy of the glass plate and Q is the energy received from the laser.

Many assumptions are required to illuminate our understanding of a gas absorbing light. Once we solve the simpler problem, we can then begin to allow the complexity to creep back in to refine our understanding.

As our contest problem for this month we present the quantitative questions from the IPhO problem that we've been discussing.

A. What are the temperature and pressure of the gas after the irradiation?

B. How much mechanical work does the gas perform? (Hint: don't forget the external pressure.)

C. How much radiant energy was absorbed by the gas?

D. What was the power of the laser and the number of photons emitted per second?

E. What was the efficiency of converting optical energy into gravitational potential energy of the plate?

F. If, after the irradiation, the cylinder is slowly rotated by 90° so that its axis is horizontal, do the temperature and pressure of the gas change? If so, what are their new values? (What simplifications do you need to make? Is the process adiabatic?)

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

The eyes have it

Our contest problem on vision and optics in the May/June issue of *Quantum* inspired correct solutions from two consistent teacher contributors, H. Scott Wiley of Wescaco HS in Texas and Art Hovey of Amity Regional HS in Connecticut.

A. If the human eye is modeled as a simple lens, we can use the lens equation

$$\frac{1}{f} = \frac{1}{D_o} + \frac{1}{D_i}$$

to find the focal lengths required for near and distant objects. Plugging in the values $D_i = 2.50$ cm and $D_o = 20$ cm yields $f = 2.22$ cm. Likewise, using the values $D_i = 2.50$ cm and $D_o = 2000$ cm yields $f = 2.50$ cm.

However, the eye is not a simple lens, as the rays that enter the cornea/lens from air do not return to air. Let's use the equation derived in the article for a single refracting surface:

$$\frac{n_1}{D_o} + \frac{n_2}{D_i} = \frac{n_2 - n_1}{R}.$$

Since $f = D_i$ when $D_o \to \infty$, the right-hand side of the equation must be equal to n_2/f. Therefore,

$$\frac{n_1}{D_o} + \frac{n_2}{D_i} = \frac{n_2}{f}$$

or

$$\frac{1}{f} = \frac{n_1/n_2}{D_o} + \frac{1}{D_i}.$$

Given that $n_2 = 1.376$ and $n_1 = 1.000$, plugging in the values $D_i = 2.50$ cm and $D_o = 20$ cm yields $f = 2.29$ cm. Using $D_i = 2.50$ cm and $D_o = 2000$ cm we get $f = 2.50$ cm.

B. If the human eye accommodated for distance by moving the lens and changing the image distance, we would get the following

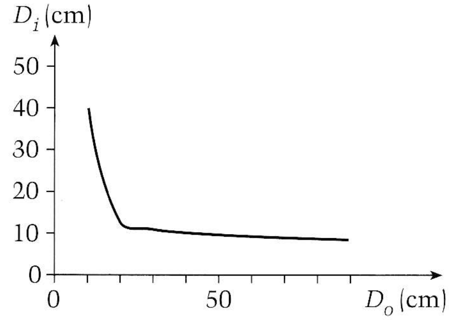


Figure 1

results given a constant focal length of 2.50 cm and using the simple lens equation: $D_o = 20$ cm yields $D_i = 2.86$ cm and $D_o = 20$ cm yields $D_i = 2.50$ cm. The range of lens movement would be 0.36 cm.

C. The only difference in the image after removal of half the lens will be in its lowered intensity due to some of the light not converging on the image. Research conducted with many people suggests that a misconception arises whereby these people think of the two or three rays drawn in ray diagrams as the only rays which contribute to the image. This is perhaps why people do not appreciate that removal of half the lens will not produce half an image. Similarly, the size of the object does not have an impact on the completeness of the image. Once again, students may assume that if the ray we choose to draw from the top of the object parallel to the principal axis is not able to hit the lens, part of the image may disappear. It is worthwhile to conduct experiments to see how changes in the lens shape or object size affect images.

D. Given the data in the problem and the simple lens equation, you can find the focal length for each pair of measurements and find an average of 7.99 cm.

The lens equation can be derived from a graph of D_i versus D_o shown in figure 1. We recognize this as a hyperbola that has been translated by $y = f$ and $x = f$. We also notice that in the general equation for a hyperbola ($xy = C$), the value of C appears to be f^2. Assuming this to be true, we obtain the Newtonian form of the lens equation:

$$(D_o - f)(D_i - f) = f^2.$$

This can be rearranged to obtain

$$\frac{D_o}{D_i} - f(D_i + D_o) = 0$$

$$\frac{1}{f} = \frac{1}{D_o} + \frac{1}{D_i}.$$

E. To find the conditions for a thick lens to have no chromatic aberration for two different colors we require the lens to have the same focal length for the corresponding two indices of refraction.

$$(n_1 - 1)\left(\frac{1}{R_1} - \frac{1}{R_2} - \frac{(n_1 - 1)}{n_1}\frac{d}{R_1 R_2}\right)$$
$$= (n_2 - 1)\left(\frac{1}{R_1} - \frac{1}{R_2} - \frac{(n_2 - 1)}{n_2}\frac{d}{R_1 R_2}\right)$$

Multiplying both sides of the equation by $R_1 R_2 n_1 n_2$ and redistributing and canceling terms yields

$$n_1(n_1 n_2 R_2 - n_1 n_2 R_1 - n_1 n_2 d + d)$$
$$= n_2(n_1 n_2 R_2 - n_1 n_2 R_1 - n_1 n_2 d + d).$$

Since $n_1 \neq n_2$,

$$n_1 n_2 R_2 - n_1 n_2 R_1 - n_1 n_2 d + d = 0$$

and

$$d = \frac{n_1 n_2 (R_2 - R_1)}{n_1 n_2 - 1}.$$

Notice that the dispersion $n_1 - n_2$ is not part of the solution—only the product of the indices appears. Since the indices of refraction are greater than 1 and the thickness of the lens is positive, we conclude that

$$R_2 - R_1 > 0.$$

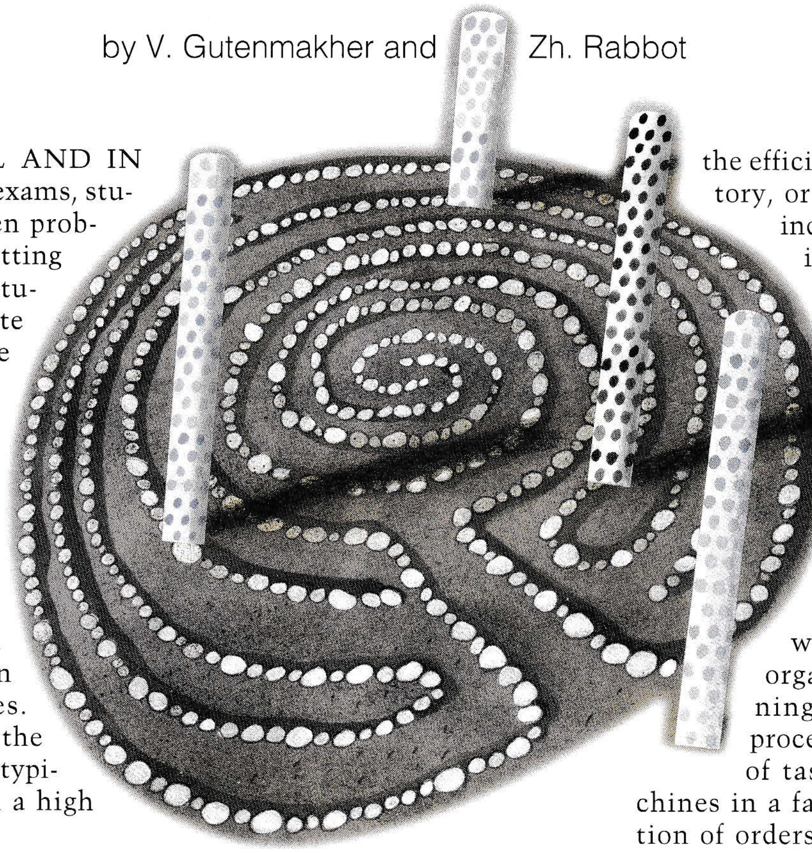
The first result is that the lens cannot be plano-convex or plano-concave since those lenses require an infinite radius of curvature and would require an infinite thickness.

The second result is that a double concave (diverging) lens is possible if R_1 is negative and R_2 is positive.

The third result is that a converging lens is possible with either $R_2 > R_1$ or $R_2 < -R_1$. The converging lens cannot be symmetric ($R_2 \neq R_1$).

Selecting the best alternative

by V. Gutenmakher and Zh. Rabbot



IN HIGH SCHOOL AND IN university entrance exams, students are often given problems that require setting up an equation. The student must first translate the conditions of the problem from natural language into the language of mathematics and then solve the equations and inequalities obtained. For the situation described in the problem the student is to find certain quantities given some other quantities. For example, here are the first and last phrases of typical problems found in a high school problem book:

Two typists have a manuscript to type... How many hours does it take each typist to type the entire manuscript?

A motor launch heads downstream from a river depot... How much time did it take the cyclist to go from the town to the tourist center?

A copper-zinc alloy containing 5 kg *of zinc is fused with* 15 kg *of zinc... What was the initial mass of the alloy?*

It is important to know how to solve such problems, which come up often in industry and economics, where one needs to calculate and combine various indicators, analyze the operation of a company, etc. Such an analysis results in a better understanding of the current situation. The next natural step is to plan future activities. Here, of course, there are numerous alternatives available, and one wants to choose the best of them.

The statement and solution of such problems are the subject of *mathematical programming.* The Russian mathematician L. V. Kantorovich (1912–1986) was one of the first to use mathematics for solving practical problems of this kind. In 1939, he published a book called *Mathematical Methods for Organization and Planning of Production.* In the introduction to this book he wrote:

"There are two ways to increase the efficiency of a shop, a factory, or an entire branch of industry. One way is to improve the technology, i.e., to provide new capabilities of the individual machines, to modify the technological processes, or to find new and better kinds of raw materials. The other, still underutilized way is to improve the organization and planning of the production process. The distribution of tasks among the machines in a factory, the distribution of orders between factories, and the distribution of various kinds of raw materials, fuel, etc., fall into this category."

Many years have passed since the publication of that book, and mathematical programming has developed into a large branch of mathematics based on economic-mathematical methods and the extensive use of computers.

In real problems of planning and management one deals with a very large number of variables simultaneously. In this article, however, we consider only simple examples with a small number of variables; in these examples, the solution can be obtained by using methods familiar even to first-year algebra students, such as proportions, properties of

Art by Vasily Vlasov

linear functions of a single variable, exhaustive testing of a small number of possibilities, and common sense.

Choosing the site of a bathhouse

The village of Soap has 100 *inhabitants and the village of Towel has* 50. *Choose a site for a bathhouse on the road connecting these villages such that the total distance traveled to the bathhouse by all* 150 *inhabitants of the two villages is minimal.*

Let us reformulate the condition of the problem in the language of mathematics. Let the distance between the villages be a km, and let the bathhouse be a distance of x km from the village of Soap. Thus, $0 \le x \le a$. The 100 inhabitants of Soap have to walk a total of $100x$ km to the bathhouse, and the inhabitants of Towel have to walk $50(a - x)$ km. Thus, the total distance traveled by all inhabitants of both villages is $S = 100x + 50(a - x)$ km.

We have obtained the following mathematical problem: *find the minimum value of the quantity* $S = 50x + 50a$ *under the condition* $0 \le x \le a$, *where* a *is a fixed number.*

This problem is very easy. To solve it, just notice that S decreases as x decreases; therefore, S takes the minimum possible value when x takes the minimum admissible value; that is, $x = 0$, and *the bathhouse must be built at Soap.*

Let us discuss this result. Almost everyone we asked this question answered that the bathhouse should be built at a place that is twice as far from Towel as from Soap. It is likely that they used the following physical model: a simple see-saw with two people at the ends of it, one of them twice as heavy as the other.

If the bathhouse is built at the distance of $a/3$ from Soap, its inhabitants altogether will walk the same distance in total as the inhabitants of Towel: $100 \cdot a/3 = 50 \cdot 2a/3$. This no doubt seems like the fairer solution if instead of considering all the inhabitants we set one village in opposition to the other and try to find an "equilibrium point." However, this is the solution of a quite different mathematical problem: *what is the value of* x $(0 \le x \le a)$ *for which the quantity* $f = |100x - 50(a - x)|$ *takes the minimum value?*

It is clear that the phrase "best alternative" can be interpreted in different ways. In order to assign a specific meaning to this phrase, we define an *objective function.* In our problem it was the function $S = 50x + 50a$, and in the other problem it was the function $f = |150x - 50a|$. Finding the value of x for which the objective function takes the minimum value gives the best result (from a given standpoint).

It is clear that the inhabitants of Towel may disagree with either of these solutions and can possibly come up with strong arguments. For example, it may well happen that there are more senior citizens and children in Towel than in Soap, that there are fewer cars in Towel, or that there is better water supply in Towel, and so on. If we take into account all these arguments, we obtain another mathematical problem or, in other words, a different *mathematical model.*

We hope that the somewhat whimsical subject of our bathhouse problem does not give the reader the wrong impression. Here are some similar problems of a serious nature: *choose a place for a lunchroom on the grounds of a big plant with several shops. What is the best way to schedule the serving of the workers in the lunchroom?*

In this case the choice of the objective function is rather clear (to minimize the serving time). However, the optimal plan depends not only on the number of workers in the different shops but on specifics of the production process and other considerations.

The best way to reach a railway station

Mr. Smith must get to the railway station as quickly as possible. He may call a taxi, which will take 24 *minutes to arrive, and then go by taxi at a speed of* 30 km/h, *or he can walk at a speed of* 6 km/h. *Which method is better if the distance to the railway station is (a)* 2 km; *(b)* 3 km; *(c)* 5 km*?*

In order to better compare the motion of the pedestrian and the car, let us introduce one more person, Mr. Smith's wife.

Suppose that Mr. Smith started walking to the railway station, and as soon as he went out, his wife noticed that he had left his ticket. She immediately called a taxi, waited for it, and started after her husband.

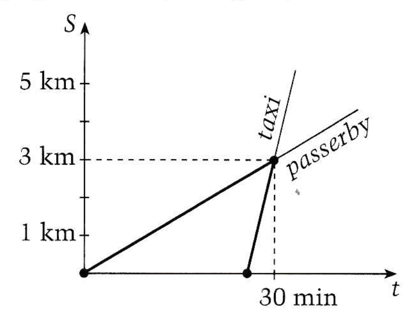
Let us calculate the time needed to catch up with him. At time t after leaving home, Mr. Smith will be at a distance of $6t$ km from his home. If $t > 24$ min $= 2/5$ h, the taxi will travel $30(t - 2/5)$ km.

The taxi can catch up with Mr. Smith when

$$6t = 30(t - 2/5);$$

that is, at $t = 1/2$ h. If the walk to the railway station takes less than 1/2 h, the wife will not be able to catch up with her husband; otherwise, she will catch up with him and take him to the railway station. In 1/2 hour, Mr. Smith can walk 3 km. As a result, we arrive at the following conclusion: if the distance to the railway station is less than 3 km (as in case (a)), it is better for Mr. Smith to walk there; if the distance is 3 km, walking and going by taxi are equivalent, and in case (c), it is better to go by taxi.

The solution can be understood more easily if it is represented in graphic form (see figure).



When solving this problem, we implicitly made certain assumptions: the pedestrian and taxi move uniformly, the call and departure of a taxi are instantaneous (in practice, this is not so), etc. In addition, we are implicitly assuming that the most valuable resource for Mr. Smith was time. However, in case (b), we have two alternatives that are equivalent from this point of view. To choose one of them, we must involve another consideration: either the cost of the trip (in which case Mr. Smith will probably prefer walking) or convenience (he'll go by taxi).

One bicycle for two

Two brothers, Tom and Dick, want to visit their grandmother who lives 40 km *away. They have a bicycle, which they have loaded with their things. Tom can walk as fast as* 6 km/h *and ride a bicycle at* 20 km/h*; Dick walks at* 4 km/h *and rides at* 30 km/h. *The bicycle may be left on the road unattended. What is the quickest way to reach their grandmother's?*

Let us suppose that the brothers arrive at their grandmother's simultaneously, and calculate the time required. Then we'll prove that one of them cannot reach their grandmother's in less time; that is, if one of them spends less time on the road, the other will inevitably arrive later than in the case when they arrive simultaneously.

Assume that Tom went x km by bicycle and walked the other $(40 - x)$ km, and Dick, conversely, walked x km and rode $(40 - x)$ km. Then Tom spent $(x/20 + (40 - x)/6)$ hours on the road and Dick spent $(x/4 + (40 - x)/30)$ hours. If they arrived simultaneously, then

$$\frac{x}{20}+\frac{40-x}{6}=\frac{x}{4}+\frac{40-x}{30}.$$

From this equation, we find that $x = 16$ km, and it took the two of them $4^4/_5$ hours to reach their grandmother's.

This solution might be implemented in the following way. Tom starts off from home on the bicycle, rides 16 km, and then leaves the bicycle on the road and starts walking. Dick starts off from home on foot, walks to where Tom left the bicycle for him, and then goes by bicycle.

Let us verify that one of them cannot reach their grandmother's house in less time. Indeed, if Tom rides less than 16 km, he must walk the extra distance and will thus spend more time than in the first solution. If Tom rides more than 16 km, then Dick will have to walk this extra distance and so he will arrive later than before.

In this problem, the simplest practical considerations led us to the optimal solution: the brothers must bring the bicycle to their grandmother's and arrive simultaneously. It is clear that if the brothers stop to rest, it will only add to their time of arrival.

In this problem we again have several options. In fact, there exists an infinite number of solutions that provide the minimum travel time: each brother may leave the bicycle for the other several times, provided that Tom walks 16 km and Dick walks 24 km in total. You may judge for yourself which solution is best.

Minimize copper

There are three alloys in a laboratory. The first one contains 40% *copper and* 60% *nickel; the second one contains* 60% *copper and* 40% *cobalt; and the third contains* 60% *cobalt and* 40% *nickel. An experiment requires one kg of a new alloy containing* 40% *cobalt and as little copper as possible (see the table below). How can it be made?*

	Cu	Ni	Co
I	40%	60%	–
II	60%	–	40%
III	–	40%	60%

Let us construct a mathematical model of the problem. Take x kg of the first alloy, y kg of the second, and z kg of the third.

The problem requires that $x + y + z = 1$. The new alloy will contain $0.4y + 0.6z$ of cobalt. So we have $0.4y + 0.6z = 0.4$. The new alloy will also contain $0.4x + 0.6y$ of copper.

Thus the following mathematical model of the problem is constructed: *find nonnegative numbers x, y, and z that satisfy the following system of equations:*

$$\begin{cases} x+y+z=1, \\ 0.4y+0.6z=0.4 \end{cases}$$

and are such that the quantity $m = 0.4x + 0.6y$ takes on the minimum possible value.

Using this system of equations, we express x and z in terms of y and substitute them in the expression for m; then we find the minimum of the resulting function of y (taking into account that all variables x, y, and z must be nonnegative).

From the second equation, we have:

$$z=\frac{2}{3}-\frac{2}{3}y.$$

Substituting this value of z in the first equation, we obtain

$$x=\frac{1}{3}-\frac{1}{3}y.$$

Therefore,

$$m=\frac{2}{15}+\frac{7}{15}y.$$

We see that the smaller the value of y, the smaller the corresponding value of m. However, the minimum possible value of y is 0. In this case, $m = 2/15$ and $x = 1/3$, $z = 2/3$. Thus we must take 1/3 kg of the first alloy and 2/3 kg of the third. The second alloy is not used at all.

Notice that in this case we again have a linear function:

$$m=\frac{2}{15}+\frac{7}{15}y,$$

which takes the minimum value at the minimum possible value of y (here $y = 0$). If we expressed m in terms of x or z rather than in terms of y, it would be more difficult to find an interval in which to seek its minimum value. The problem of choosing a variable in terms of which the object function should be expressed is very important.

We could do without these manipulations if we noticed the following circumstance. In order for the new alloy to contain 40% cobalt and as little copper as possible, it must contain as much nickel as possible. Therefore, the second alloy, which might seem appropriate because it already contains 40% cobalt, is actually better left out, since it contains no nickel at all (see the table above). Thus we must fuse the first and the third alloys. If we take a kg of the first alloy, we must take $(1 - a)$ kg of the third. The new alloy will then contain $0.6(1 - a)$ kg of cobalt. Then $0.6(1 - a) = 0.4$. Therefore, $a = 1/3$, which gives the desired result.

Christmas problem

There is a budget of \$100 *dollars to buy Christmas tree ornaments. Ornaments are sold in sets. A set containing* 20 *items costs* \$4, *a set containing* 35 *items costs* \$6, *and a set containing* 50 *items costs* \$9. *Which sets should be chosen in order to buy the maximum possible number of ornaments?*

Each item in the first set costs 1/5 dollar, each item in the second set costs 6/35 dollars, and each item in the third set costs 9/50 dollars. Let us list these numbers in ascending order: $6/35 < 9/50 < 1/5$. We see that the second set contains the cheapest ornaments, and the first set contains the most expensive ones.

I	II	III
20	35	50
\$4	\$6	\$9

I	II	III
1	1	1
\$1/5	\$6/35	\$9/50

In order to buy as many ornaments as possible for \$100, we certainly must buy as many cheap ornaments as possible. Thus, we can buy at most 16 sets of the second type (6 dollars each), which cost a total of \$96. For the remaining \$4, we can buy only the first set, which gives us $16 \cdot 35 + 20 = 580$ decorations.

This seems to be the best choice. To make sure, let us check one other possibility. If we buy fifteen \$6 sets, we'll have \$10 dollars left, which we may spend on one \$9 set or two \$4 sets. In either case the total number of ornaments turns out to be less than 580.

Thus the best solution is to buy sixteen \$6 sets and one \$4 set. This gives 580 ornaments in total.

We obtained this solution using the simple consideration: the cheaper the ornaments, the more of them can be bought.

Our reasoning seems very sound. But how do we know, for instance, that things won't turn out better if we buy 14 of the cheapest sets? Let's now give a more rigorous solution.

Let x be the number of the sets of the type I, y the number of type II, and z the number of type III. We must find nonnegative numbers x, y, and z such that $4x + 6y + 9z \le 100$ and the quantity $S = 20x + 35y + 50z$ is as large as possible.

Since

$$4x + 6y + 9z = \frac{6}{35}S + \frac{4}{7}x + \frac{3}{7}z \ge \frac{6}{35}S,$$

we find that

$$\frac{6}{35}S \le 100,$$

which implies that

$$S \le 583\frac{1}{3}.$$

Since S is an integer divisible by 5, we have $S \le 580$. For $x = 1$, $y = 16$, and $z = 0$, all conditions of the problem are satisfied and $S = 580$.

This problem pertains to *integer programming*, which is one of the most complex divisions of mathematical programming.

Problems

1. Three shop assistants, George, Jacob, and Leo, must be assigned to three sections of a department store: radio, photo, and musical instruments. The director asked a psychologist to help him make the assignments in the best way. The psychologist tested the knowledge and inclinations of the shop assistants for each of the fields and assessed them in points as shown in the table.

	radio	*photo*	*music*
George	5	4	7
Jacob	6	7	3
Leo	8	11	2

How must the director make the assignments to obtain the maximum possible number of points? (For example, if George is assigned to the radio department, Jacob is assigned to the music department, and Leo is assigned to the photo department, the total sum of points is $5 + 3 + 11 = 19$.)

2. A frying pan can hold two lambchops. They can be fried in 10 minutes on both sides. What is the quickest way to fry 3 lambchops on this pan?

3. Two points, A and B, lie on the same side of a line. How must a segment MK of length a be placed on this line so that the polygonal line $AMKB$ will have the shortest possible length?

4. Three brothers bought a bicycle. They must get it home, 30 km away from the shop. Each of the brothers walks at a speed of 4 km/h and rides the bicycle at 20 km/h. What is the minimum time in which they can reach their home? (The bicycle may be left on the road unattended.)

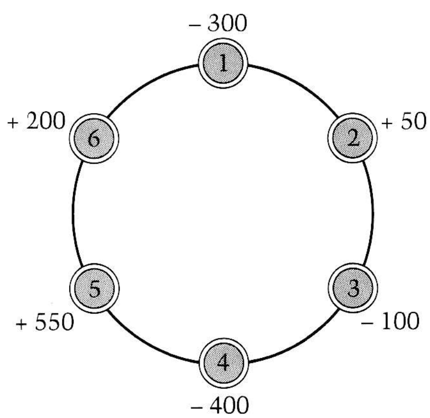
5. Nils can eat a cake in 10 minutes, eat a jar of jam in 8 minutes, and drink a carton of milk in 4 minutes. Karlsson can do the same things twice as fast. What is the smallest amount of time it will take them together to consume a breakfast consisting of a cake, a jar of jam, and a carton of milk?

6. Each of four vessels contains 1 liter of a mixture of acid with water. The percentage of acid in them is 10%, 30%, 60%, and 80%, respectively. A laboratory assistant has to prepare a 50% mixture of acid with water. What is the maximum quantity of such a mixture that can be obtained by mixing the mixtures available?

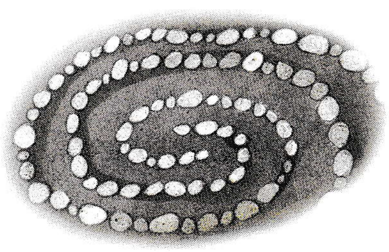
7. It is required to prepare an alloy containing 40% tin. There are three alloys available containing 60%, 10%, and 40% tin. The price of 1 kg of these alloys is \$4.30, \$5.80, and \$5.50, respectively. Which alloys and what proportions should be used to make the new alloy as cheaply as possible?

8. Three types of apartment blocks can be constructed from building components of two types. To build a 12-apartment block requires 70 components of the first type and 100 components of the second type, a 16-apartment block requires 110 and 150 components of the first and second types, respectively, and a 21-apartment block requires 150 and 200 components, respectively. There are 900 components of the first type and 1300 components of the second type available. How many apartment blocks of each type must be built in order to obtain the maximum possible number of apartments?

9. There are three warehouses and three shops on a circular road: the distance between two neighboring points is 1 km. The following figure



shows the availability of goods at the warehouses (indicated with a plus sign) and the demand of the shops (indicated with a minus sign) in tons. You must draft the most efficient plan of delivery to transport all the goods from the warehouses to the shops in such a way that the total sum of ton-kilometers transported is minimal. (Give the best plan of delivery and present your reasoning.)



Heating water from the top

by V. Pentegov

PEOPLE HEAT WATER EVERY day, and in almost every case they place the heater under the container. This is quite understandable: convection occurs throughout the entire volume of water, thereby uniformly heating all the water to the desired temperature.

What will happen if we place the heater at the top of water instead? This is not an idle question. Recall that 70% of the surface of the Earth is covered with water, which is heated from the top by the Sun. Thus heating from top to bottom occurs on a large scale in Nature. Many processes involved in this way of heating are so complicated that there is no universal agreement on the mechanisms.

For example, not long ago it was found that temperature in the depths of the ocean does not change gradually but by abrupt jumps. These jumps occur in a very narrow region between layers of constant temperature. Besides the temperature, the density and the salinity (concentration of salts) are also constant within each layer. With time the jumps are smoothed out (temperature much more rapidly than salinity), but the boundaries between the layers remain unshifted and persist for nearly the entire time.

It is natural to wonder why this occurs. Before attempting to explain it, let's turn to some simple experiments and observations.

Two quite different liquids: water and... water

Take a 3-liter jar of water and an electric immersion heater. Immerse the heater not very deeply into the water (figure 1), turn it on, and wait until the water in the upper part of the jar boils or is near the boiling point (at this moment large bubbles will break from the heater). Now tint the water with ink. You will see that two zones are formed in the water: the upper zone is tinted, while the lower zone remains clear. These zones are separated by a distinct boundary, through which the molecules of the dye do not penetrate (figure 2, *a*). Temperature measurements in these zones yield the following data: the temperature in the upper layer is the same everywhere and is approximately equal to the boiling point (95 °C), while in the lower layer the temperature doesn't rise higher than 40–45 °C. In other words, there is a temperature jump of 50–60 °C at the boundary between the layers, and the water in the upper and lower parts behaves like two different, immiscible liquids.

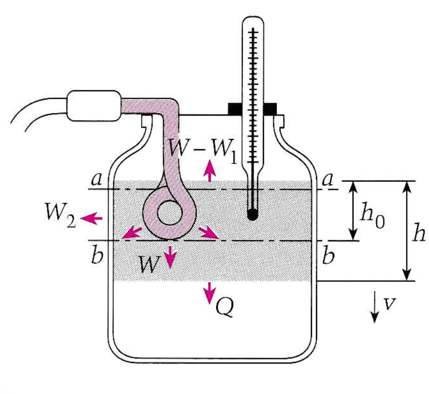


Figure 1. *Experimental setup.*

Waves under water

Transverse waves can arise at the boundary surface. (By the way, the interlayer boundary can be observed even without tinting the water, because the waves at the boundary surface reflect and scatter light.) This means that there is a surface tension at the boundary. A clear illustration of this fact is the reflection of vortex rings from the boundary.

Turn the heater off (to reduce the movement of water in the hot layer) and drop some ink from a dropper into the jar from a height of 1–2 cm above the water level. A vortex ring, moving downward, will form in the water. If the energy of the ring is not too high, it will be reflected elastically from the boundary (the latter will bend a little) and then disintegrate. By contrast, if the ring's energy is high enough, it can pass through the boundary, only to disintegrate immediately below it.

Moving boundary

After watching the interlayer boundary for some time, you will notice that it moves. Experimental plots of the upper layer thickness h versus the time t and of the speed v of the boundary versus the upper layer thickness or, strictly speak-

When the author wrote this paper, he was a 10th-grade student in Kiev (Ukraine).

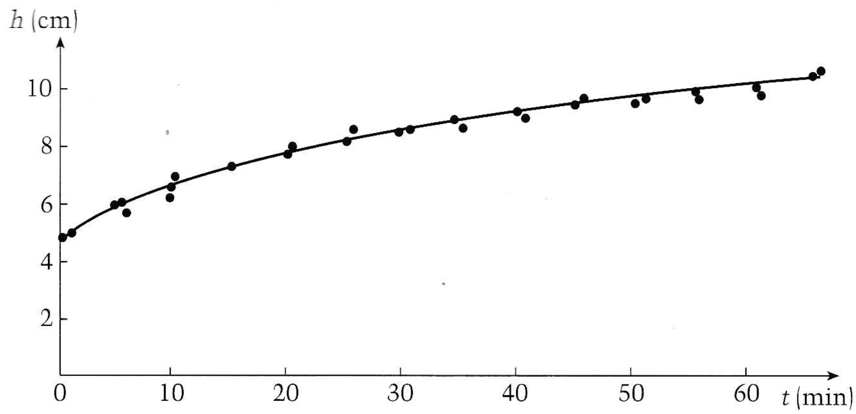


Figure 3. *Thickness* h *of the upper layer versus time* t *(data of two experiments).*

ing, the difference of the levels $h - h_0$ (figure 1) are shown in figures 3 and 4. The measurements were obtained for a 3-liter container, using a heater with a power of approximately 120 W.

A significant shift of the boundary only occurs when the temperature of the upper layer is near or at the boiling point: when the heater is switched off, motion of the boundary stops. It can be supposed that the motion of the boundary results from intensive mixing in the hot layer, which occurs specifically at high temperatures. To test this hypothesis, let's take the next step. We turn the heater off (the boundary will stop moving) and cautiously begin to mix the water in the upper layer. As expected, the boundary starts moving again.

Based on the above facts, one may propose the following mechanism to explain the boundary motion. Consider a small domain at the boundary surface separating the cold and hot water layers (figure 5, *a*: the red and blue points show molecules of hot and cold water, respectively). As a result of diffusion and collisions between the molecules, the "hot" molecules penetrate below the interlayer boundary. A thin layer of water becomes hot, and after a while it is carried away by convective flows of hot water (figures 5, *b* and *c*). The result of this molecular process is a sinking of the boundary.

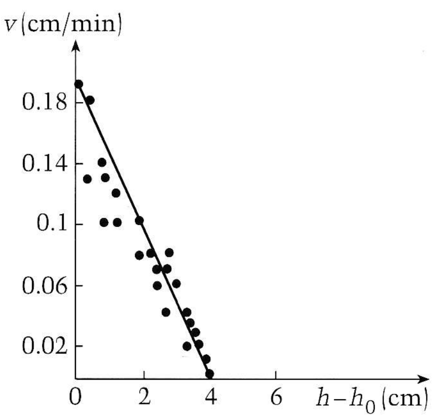


Figure 4. *Velocity* v *of the boundary versus the depth* $h - h_0$ *(data of two experiments).*

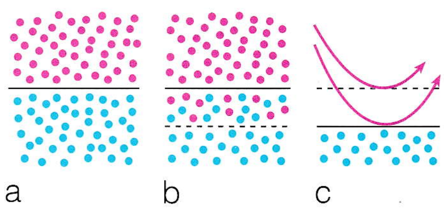


Figure 5. *Illustration of the mechanism by which the boundary shifts.*

Thus when water is heated from the top, heat is transferred not only by thermal conduction (which is quite natural) but also by the motion of the boundary due to convection of the water in the hot layer. This convection brings about a practically uniform temperature in all parts of the upper layer.

Now let's try to answer yet another question raised by these experiments: does the motion of the interlayer boundary depend on the power of the heater? Clearly the thickness of the upper layer cannot be smaller than the depth of the level *a-a* (figure 1), because at the higher levels convection occurs all the time. On the other hand, the initial thickness cannot be greater than the depth of level *b-b*, because there is not yet any convection below it (if the heater is pointlike or flat, the two levels *a-a* and *b-b* coincide). Hence, a certain initial thickness of the upper layer should exist which is determined mainly by the depth of immersion of the heater. Motion of the boundary begins with the start of vigorous mixing of water in this layer. As we have said above, it occurs at a temperature near the boiling point. Thus the heater should be capable of heating a liquid layer of the initial thickness virtually to the boiling point under real experimental conditions, where some heat is dissipated in the surroundings. This condition determines the minimum heating power needed for motion of the boundary.

The experiments showed that when the heating power is greater than the minimum, the speed of the boundary motion will change linearly with this power (see the Appendix for details).

Experimental results

We may draw a number of conclusions from our experiments:

1. The boundary between warm and cold water can exist for a long time and maintain its integrity only in the presence of convective flows of water near the boundary: these flows counterbalance the diffusion and maintain the boundary.

2. The boundary moves with a speed determined by the temperature difference. As the temperature difference decreases, the speed of the boundary decreases until it stops completely, but the boundary itself is maintained as long as there is a flow of water near it.

3. In our experiments the large temperature jumps (and high temperature of the heated water) were needed only for producing intense convective flows in the hot water. If the flows of water on the two sides of the boundary are maintained in some other way (say, by a mechanical device), the temperature jump could be small.

Guesses and hypotheses

Now let's return to the problem of the heating of oceanic water by the Sun, and try to answer some questions. Why do the boundaries between different layers in the ocean remain in place even though a gradual leveling of the temperature and salinity occurs? And why does the leveling of the salinity take so much longer than the establishment of temperature equilibrium?

Again this can be explained by convection, which in the ocean occurs in every layer (in contrast to our experiments, where it went on only in the hot layer).

Consider two neighboring layers. Diffusion and molecular collisions produce "hot" and "cold" molecules on both sides of the boundary. They are carried away from the boundary by convective flows. If the convection is equally strong on both sides of the boundary, the boundary doesn't move, although energy transfer takes place, resulting in a gradual leveling of the temperature.

Salinity equilibrium is established much more slowly. This is because of surface tension at the boundary. It is known that the coefficient of surface tension of a saline solution is greater than that of pure water. Like every physical system, the boundary "tries" to minimize its energy. Therefore, the ions produced by the decomposition of salt molecules are predominantly located far from the boundary, and a marked diffusion of them begins only after the surface tension at the boundary has fallen significantly.

The formation of a boundary between cold and warm water could probably explain not only the laminated structure of oceanic water but may also underlie the distinct and stable boundaries of oceanic currents, the recently discovered giant oceanic vortices, and many other wonders.

Appendix

We denote by W the power of the heater (it should be larger than the minimum power). Although some part of the heater's energy is spent evaporating water, this loss is stable at constant temperature. Therefore, for convenience, let's introduce the net power W_1, which is equal to the difference between the power of the heater and the power expended on evaporation. We assume the net power to be constant.

It is natural to suppose that the power W_2 dissipated to the surroundings through the lateral surface of the container is proportional to the thickness h of the hot layer: $W_2 = kh$. The proportionality factor depends on the experimental conditions and doesn't vary during an experiment.

Let the boundary shift by a distance dh over a time dt. This means that a water layer of thickness dh has been heated from the temperature of the cold layer to that of the hot layer, that is, by ΔT degrees. This process requires a thermal energy $dQ = c\rho S\Delta T dh = k_1 dh$ (here c is the specific heat of water, ρ is its density, and S is the horizontal cross-sectional area of the container). According to energy conservation,

$$W_1 dt = dQ + W_2 dt,$$

or

$$W_1 dt = k_1 dh + khdt.$$

First, we rearrange the equation:

$$\frac{dh}{W_1 - kh} = \frac{dt}{k_1},$$

and then we integrate it to obtain the dependence of the thickness h of the hot layer on time t:

$$h = a - be^{-\alpha t},$$

where $a = W_1/k$, $b = W_1/k - h_0$, and $\alpha = k/k_1$ are constant coefficients, and h_0 is the initial depth where the boundary is formed.

Now let's find the speed v of the boundary motion:

$$v = \frac{dh}{dt} = \frac{W_1 - kh}{k_1}.$$

It is seen that the speed of the boundary depends linearly on the layer thickness, and at a certain thickness (which is equal to W_1/k) the speed must be zero. This is because the energy losses grow with thickness, and there will come a time when all the energy from the heater will leak to the surroundings. In the region where the boundary comes to rest, the adjacent convective flows become so weak that the movement of liquid comes practically to a stop, and the boundary will gradually disperse on account of diffusion and thermal conductivity.

Quantum on the heating of water:

V. Mayer and E. Mamayeva, "Two Physics Tricks," March/April 1991, p. 35.

A. Abrikosov, "The Story of a Dewdrop," September/October 1992, pp. 34–38.

A. Buzdin and V. Sorokin, "Double, Double Toil and Trouble," May/June 1992, pp. 52–53.

T. Polyakova, V. Zabolotsky, and O. Tsyganenko, "Stirring up Bubbles," March/April 1997, pp. 52–55.

I. Mazin, "Physics in the Kitchen," September/October 1997, pp. 54–56.

Obtaining symmetric inequalities

by S. Dvoryaninov and E. Yasinovyi

TEXTBOOKS AND MATHEMATICAL COMPetitions often include problems in which one must prove an inequality in several variables. Numerous papers and books are dedicated to general theorems and various methods of proving such inequalities.

Here we examine Muirhead's theorem, which concerns inequalities between certain types of symmetric polynomials. This theorem was proved in 1903. It is remarkable not only for its generality, but also because combinatorial concepts related to this theorem (Young diagrams and their majorization) also occur in various fields of pure and applied mathematics.

We will deal with inequalities involving *homogeneous* polynomials. Some examples of such inequalities are:

$$x^2 + y^2 \ge 2xy, \tag{1}$$

$$x^5 + y^5 \ge x^3y^2 + y^3x^2, \tag{2}$$

$$x^2 + y^2 + z^2 \ge xy + yz + zx, \tag{3}$$

$$x^3 + y^3 + z^3 \ge 3xyz, \tag{4}$$

$$x^2y^2 + y^2z^2 + z^2x^2 \ge x^2yz + y^2xz + z^2xy, \tag{5}$$

$$x^4 + y^4 + z^4 + w^4 \ge 4xyzw. \tag{6}$$

These inequalities hold for all nonnegative values of the variables. For the case of two variables, they can be easily proved by grouping the terms and factoring. For example, let us prove inequality (2) for all nonnegative x and y. For this purpose, consider the following difference:

$$\begin{aligned} x^5 + y^5 - x^3y^2 - y^3x^2 &= x^3(x^2 - y^2) - y^3(x^2 - y^2) \\ &= (x^3 - y^3)(x^2 - y^2). \end{aligned}$$

We can see that the last product is nonnegative, because both factors are either nonnegative (for $x \ge y \ge 0$) or nonpositive (for $y \ge x \ge 0$).

It is more difficult to devise a symmetric proof for inequalities in three or more variables. For the time being, we demonstrate the general idea by way of example (5). In this case, it will be convenient to move all terms to one side of the inequality, multiply them by 2, and arrange them in three groups:

$$x^2(y^2 - 2yz + z^2) + y^2(x^2 - 2xz + z^2) + z^2(x^2 - 2xy + y^2)$$
$$= x^2(y - z)^2 + y^2(x - z)^2 + z^2(x - y)^2 \geq 0.$$

Exercise 1. Prove inequalities (1), (3), and (6).

To learn how inequalities of this type can be proved, and to formulate a general theorem, it is necessary to become familiar with some new concepts described in the next section.

Symmetrization of a monomial

Suppose we are given several nonnegative variables—for example, the three variables x, y, and z, which take on nonnegative values. Suppose also that we are given a set of the same number of nonnegative integers: $\alpha = (k, j, i)$, where $k \geq j \geq i$, which we will call *exponents*. Let us draw a table consisting of three squares and insert the corresponding exponent in the upper right corner of each square. We insert the three variables into the squares to obtain the monomial $x^k y^j z^i$. Now, we insert these variables into the table in a different order to obtain another monomial, for example, $y^k x^j z^i$ (fig. 1). We can easily count the number of different monomials that can be obtained in this way. Each of the three vari-

Art by Sergey Ivanov

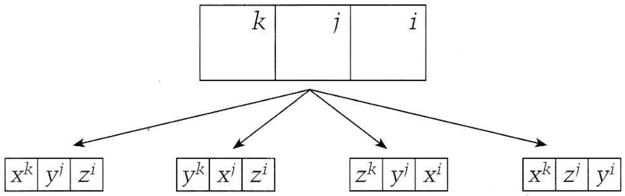
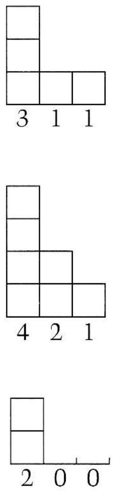


Figure 1



$$T_{(3,\,1,\,1)}(x, y, z) = 2(x^3yz + y^3zx + z^3xy)$$

$$T_{(4,\,2,\,1)}(x, y, z) = x^4y^2z + y^4z^2x + z^4y^2x + x^4z^2y + y^4x^2z + z^4x^2y$$

$$T_{(2,\,0,\,0)}(x, y, z) = 2(x^2 + y^2 + z^2)$$

Figure 2

ables x, y, and z can be placed into the first square, which gives three possibilities. Then, any of the two remaining variables can be placed in the second cell, which gives $3 \cdot 2 = 6$ possibilities. (The placement of two variables determines which variable we place in the third cell, so these are the only six possibilities for filling cells with variables.) Then we add all the monomials we've just written down to obtain a polynomial in three variables, x, y, and z, which we will denote by $T_{(k,\,j,\,i)}(x, y, z)$, or $T_\alpha(x, y, z)$, or just by T_α (T stands for the word *Table*). Polynomials of this kind are called symmetric, because they do not change if the variables are permuted. The *degree* of each monomial is $s = k + j + i$. Here are some examples:

$$T_{(2,\,1,\,0)}(x, y, z) = x^2y + y^2x + z^2x + x^2z + y^2z + z^2y,$$

$$T_{(3,\,1,\,1)}(x, y, z) = x^3yz + y^3xz + z^3xy + x^3zy + y^3zx + z^3yx = 2(x^3yz + y^3zx + z^3xy),$$

$$T_{(2,\,2,\,2)}(x, y, z) = 6x^2y^2z^2.$$

The last two examples show that if there are equal numbers among the exponents, we can collect like terms in T_α and present it in a shorter form.

If the set of exponents $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ consists of n numbers, the table must contain n squares, and we must use n variables $x_1, x_2, \ldots, x_n$. Then, the polynomial $T_\alpha(x_1, \ldots, x_n)$ involves $n! = 1 \cdot 2 \cdot \ldots \cdot n$ terms (before like terms are collected).

Thus, a polynomial T_α is assigned to every set of integers $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq 0$. This polynomial is called the *symmetrization* of the monomial with exponents $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Any such set α can be represented by a "staircase" consisting of n steps, the height of each step being equal to the corresponding exponent and the width of each step being 1. Such a staircase can be drawn on a sheet of graph paper: the total number of squares will be equal to the power of the polynomial T_α, $s = \alpha_1 + \alpha_2 + \ldots + \alpha_n$. Figure 2 shows several staircases that correspond to the polynomials occurring in the inequalities mentioned in this article. These staircases have a scientific name: they are called Young diagrams, and they prove to be useful in various problems involving combinatorics, algebra, and calculus.

Exercise 2. Write the polynomials T_α and draw the corresponding staircases for the following sets α: (3, 2); (3, 2, 1); (3, 3, 0, 0); (4, 1, 1, 0); (5, 0, 0, 0, 0); (1, 1, 1, 1, 1).

Comparing Young diagrams

Figure 3 shows the pairs of staircases that correspond to the inequalities (1), (2), and (4)–(6). We see that the steeper staircase corresponds to the greater of the two polynomials. The more sloping staircase can be obtained from the steeper one by moving several bricks to the right and downward (fig. 4). We now formulate a more precise definition of the word *steeper*. We do this for staircases consisting of three steps.

Let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and $\beta = (\beta_1, \beta_2, \beta_3)$ be two sets of in-

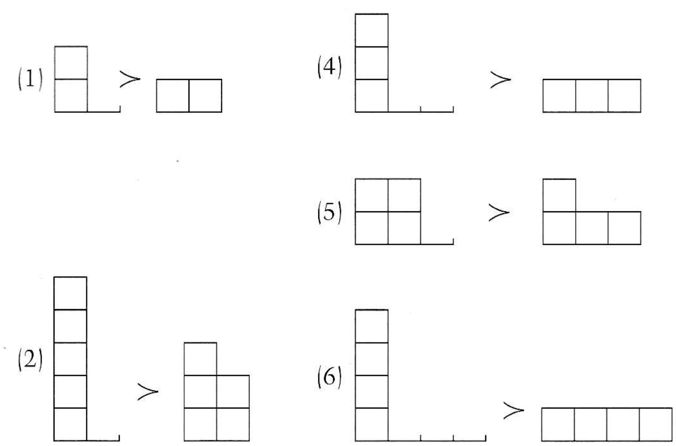


Figure 3

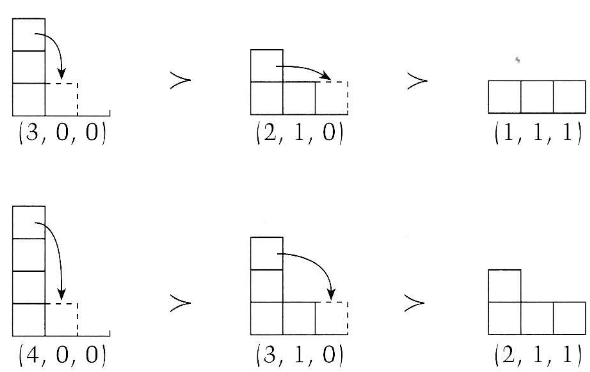


Figure 4

tegers such that $s = \alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3$, $\alpha_1 \ge \alpha_2 \ge \alpha_3$, and $\beta_1 \ge \beta_2 \ge \beta_3$. We say that $\alpha > \beta$ (α *majorizes* β) if the following condition is met: β can be obtained from α by performing the following operation several times (or only once or not at all):

$$\begin{array}{ccc} & (k, j, i) & \\ \swarrow & \downarrow & \searrow \\ (k-1, j+1, i) & (k-1, j, i+1) & (k, j-1, i+1) \end{array} \qquad (*)$$

This condition can also be written in another, equivalent, form: $\alpha > \beta$ if

$$\begin{cases} \alpha_1 \ge \beta_1, \\ \alpha_1 + \alpha_2 \ge \beta_1 + \beta_2, \\ \alpha_1 + \alpha_2 + \alpha_3 \ge \beta_1 + \beta_2 + \beta_3. \end{cases} \qquad (**)$$

Similarly, for nonincreasing sets of nonnegative integers $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ we write $\alpha > \beta$ if

$$\begin{cases} \alpha_1 \ge \beta_1, \\ \alpha_1 + \alpha_2 \ge \beta_1 + \beta_2, \\ \quad \ldots \\ \alpha_1 + \alpha_2 + \cdots + \alpha_{n-1} \ge \beta_1 + \beta_2 + \cdots + \beta_{n-1}, \\ \alpha_1 + \alpha_2 + \cdots + \alpha_n = \beta_1 + \beta_2 + \cdots + \beta_n. \end{cases}$$

For example, $(4, 2, 1) > (3, 2, 2)$, since $4 \ge 3$, $4 + 2 \ge 3 + 2$, and $4 + 2 + 1 = 3 + 2 + 2$.

The relationship $>$ between two sets is similar to the order relationship $\ge$ between two numbers: $\alpha > \alpha$, and if $\alpha > \beta$ and $\beta > \gamma$, then $\alpha > \gamma$. However, this order relation is only a partial one: it may happen that two sets with equal sums are incomparable (see exercise 6).

Exercises

3. Verify that conditions $(*)$ and $(**)$ are met for the pairs of Young diagrams shown in figure 3.

4. Prove that conditions $(*)$ and $(**)$ are equivalent; that is, if inequalities $(**)$ hold, the set β can be obtained from α by moving the bricks to the right and downward.

5. Draw all staircases consisting of $s = 4$ bricks in decreasing order, beginning with the steepest one, (4, 0, 0, 0), and ending with the most sloping, (1, 1, 1, 1). Do the same for the staircases of 5 bricks.

6. (a) Verify that Young diagrams (4, 1, 1) and (3, 3, 0) are incomparable—neither of them majorizes the other. Do other incomparable sets with the sum 6 exist?

(b) Find all incomparable pairs of sets for $s = 7$.

Muirhead's theorem

Now we are ready to formulate the basic theorem.

Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ *and* $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ *be two sets of exponents with the same sum. If* $\alpha > \beta$, *then* $T_\alpha(x_1, x_2, \ldots, x_n) > T_\beta(x_1, x_2, \ldots, x_n)$ *for all nonnegative*

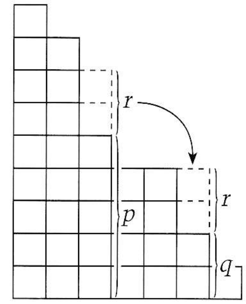


Figure 5

$x_1, x_2, \ldots, x_n$. *Conversely, if this inequality holds, then* $\alpha > \beta$.

We will not formally prove this theorem for the general case (this proof can be found in references 1 and 2 at the end of the article, and reference 2 is completely dedicated to variations of the majorization relationship, its applications, and generalizations). Instead, we describe the basic ideas underlying the proof and demonstrate them by way of several examples.

The proof of the second part of the theorem's assertion—the necessity of the assumption $\alpha > \beta$—follows from the simple fact that of two polynomials in one variable t (with positive leading coefficients), the polynomial with the greater power is greater for large t (see exercise 8).

The proof of the sufficiency of the assumption $\alpha > \beta$ is based on two ideas. The first is the idea of moving bricks to the right and downward. Two sets $\alpha > \beta$ can be connected by a chain of sets so as to make two neighboring sets in the chain differ at only two places. We can pass from each set of the chain to the next by moving a brick from one step to the next on the corresponding staircase. The second idea is that of *symmetrical grouping:* the difference between the polynomials corresponding to neighboring sets can be represented as the sum (over all sets of variables x and y) of identical groups of the form

$$(x^{p+r}y^q + y^{p+r}x^q - x^p y^{q+r} - y^p x^{q+r})Z$$

(see fig. 5), where Z is the product of other variables corresponding to identical exponents in the neighboring sets. The last expression can be factored, and it can be easily verified that it is nonnegative for $p > q$ and $r \ge 0$.

If the reader analyzes several examples by carefully writing out all manipulations, everything will become clear. Here, we give two examples.

Inequalities involving means

The inequality between the arithmetic and geometric means of three nonnegative numbers is as follows:

$$\frac{a_1 + a_2 + a_3}{3} \ge \sqrt[3]{a_1 a_2 a_3}.$$

Introducing the notation $a_1 = x^3$, $a_2 = y^3$, and $a_3 = z^3$, we arrive at inequality (4). We prove it by following

Muirhead's method. Looking at figure 4a, we see that we must prove the two following inequalities:

$$T_{(3,0,0)}(x,y,z) \ge T_{(2,1,0)}(x,y,z) \ge T_{(1,1,1)}(x,y,z).$$

Let's consider the first of them and take the difference

$$\begin{aligned} R_1 &= T_{(3,0,0)}(x,y,z) - T_{(2,1,0)}(x,y,z) \\ &= 2(x^3+y^3+z^3) - x^2y - y^2z - z^2x - x^2z - y^2x - z^2y. \end{aligned}$$

We rearrange the terms in groups of four according to the following principle: for each pair of variables, we include in a group all terms in which the exponents of these variables change from (3, 0) to (2, 1) (and the common exponent of the third variable is equal to zero). Thus, we have:

$$\begin{aligned} R_1 &= (y^3+z^3-y^2z-z^2y) + (x^3+y^3-x^2y-y^2x) \\ &\quad + (z^3+x^3-z^2x-x^2z) \\ &= (y^2-z^2)(y-z) + (x^2-y^2)(x-y) + (z^2-x^2)(z-x) \ge 0 \end{aligned}$$

for all nonnegative x, y, and z.

We now prove the second inequality.

$$\begin{aligned} R_2 &= T_{(2,1,0)}(x,y,z) - T_{(1,1,1)}(x,y,z) \\ &= x^2y + y^2z + z^2x + x^2z + y^2x + z^2y - 6xyz. \end{aligned}$$

Here, the common exponent of the variable y is equal to 1, and the exponents (2, 0) of the variables x and z change to (1, 1). We have

$$\begin{aligned} R_2 &= x(y^2+z^2-2yz) + y(z^2+x^2-2zx) + z(x^2+y^2-2xy) \\ &= x(y-z)^2 + y(z-x)^2 + z(x-y)^2 \ge 0, \end{aligned}$$

which proves inequality (4).

Another inequality

Now we look at the inequality

$$\begin{aligned} a+b+c &\le \frac{a^2+b^2}{2c} + \frac{b^2+c^2}{2a} + \frac{c^2+a^2}{2b} \\ &\le \frac{a^3}{bc} + \frac{b^3}{ca} + \frac{c^3}{ab}. \end{aligned}$$

One possible way of proving this inequality is Muirhead's method. Multiplying both sides of this inequality by $2abc$, we can reduce it to the following "Muirhead" type:

$$2(a^4+b^4+c^4) \ge a^3b + b^3a + b^3c + c^3b + c^3a + a^3c, \qquad (7)$$

$$a^3b + b^3a + b^3c + c^3b + c^3a + a^3c \ge 2(a^2bc + b^2ac + c^2ab). \quad (8)$$

The corresponding Young's diagrams are shown in figure 4b. The difference between the left-hand and right-hand side of each inequality can be written, as before, as a sum of three groups consisting of four terms each. To save space, we write down only one group (the others can be obtained by replacing variables a, b, and c by b, c, a and c, a, b, respectively.

Proof of (7):

$$(a^4+b^4-a^3b-b^3a) + \ldots = (a^3-b^3)(a-b) + \ldots$$

Proof of (8):

$$(a^3b + c^3b - a^2cb - c^2ab) + \ldots = b(a^2-c^2)(a-c) + \ldots$$

Thus, the initial inequality is proved.

More Exercises and Problems

7. Write down all the Muirhead inequalities for polynomials of degree 4.

8. Let $T_{(\alpha_1,\alpha_2,\alpha_3)}(x,y,z) \ge T_{(\beta_1,\beta_2,\beta_3)}(x,y,z)$. Prove that inequalities (**) hold. (Consider several cases: $x=y=z=t$, $x=y=t$, $z=1$, $x=t$, $y=z=1$, and compare the powers of the polynomials obtained.)

9. Prove the following inequalities (for nonnegative x, y, z, v, and w):

(a) $x^4y^2z + y^4x^2z + z^4yx^2 + x^4z^2y + z^4x^2y$
$\ge 2\left(x^3y^2z^2 + y^3z^2x^2 + z^3x^2y^2\right)$,

(b) $x^5 + y^5 + z^5 \ge x^2y^2z + y^2z^2x + z^2x^2y$,

(c) $x^3 + y^3 + z^3 + v^3 \ge xyz + xyv + xzv + yzv$.

10. Derive the inequality for the arithmetic and geometric means of n nonnegative numbers from Muirhead's theorem. How many bricks must be moved for the set $(n, 0, 0, \ldots, 0)$ to be transferred to the set $(1, 1, \ldots, 1)$?

11. Formulate and prove Muirhead's theorem for all nonnegative exponents (not necessarily integers).

12. For certain sets of exponents (with an even sum s), the inequality in Muirhead's theorem holds for all values of the variables (and not only for nonnegative ones). Try to describe all such cases.

References

1. Hardy, G. H., Littlewood, J. E., and Polya, G., *Inequalities*, Cambridge Univ. Press, 1952.

2. Marshall, A. W. and Olkin, I., *Inequalities: Theory of Majorization and Its Applications*, New York: Academic Press, 1979.

COMMENTARY

The toy that drove the universe

by Jef Raskin

RECENTLY, I PURCHASED a toy: a hollow, transparent sphere about 10 cm across, into which I placed sixty 5-mm-diameter white beads. Each bead was imprinted with a number from 1 to 60. The sphere can be rotated by a motor and a simple circuit arranged so that one button makes a curved arm stir the beads and another button causes the arm to move in the opposite direction and pick up one of the beads for viewing.

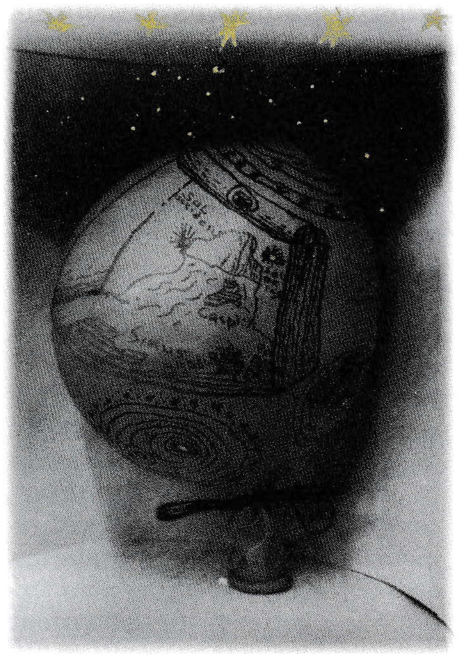
The beads are nearly identical—in any case, their differences do not make one more likely to be chosen than the other. The device was made for games of chance, and it can be considered a good randomizer. The other morning, my friend Julie ran the machine 100 times, each time with a full load of beads. Here is the output:

06, 31, 42, 29, 33, 05, 26, 01, 05, 28, 22, 32, 59, 59, 09, 57, 16, 46, 12, 13, 16, 25, 45, 14, 12, 38, 37, 51, 10, 34, 21, 10, 09, 35, 21, 23, 60, 09, 04, 33,12, 32, 32, 13, 28, 11, 54, 46, 58, 33, 25, 07, 09, 02, 19, 60, 52, 23, 29, 48, 52, 35, 18, 13, 57, 45, 15, 24, 28, 24, 05, 59, 03, 03, 45, 22, 48, 53, 27, 18, 49, 01, 59, 37, 17, 51, 36, 33, 09, 41, 04, 43, 06, 39, 31, 60, 32, 06, 17, 41.

The probability p of getting this sequence is $1/(60^{100})$, or approximately

$$p = 1.5 \cdot 10^{-178}.$$

Each turn of the machine takes about one second; producing another sequence of this length would therefore take about 100 seconds. To have a fifty-fifty chance of repeating



this sequence, you'd have to turn out approximately 10^{178} more sequences, which would take about 10^{180} seconds. The age of the Universe, 12 billion years (one current estimate), is "only" about $4 \cdot 10^{17}$ seconds. So even if the machine generated sets of 100 numbers for the next few billion years, the sequence shown above would not likely turn up again.

If the force of gravity was ever so slightly different, if the coefficient of friction of the surface of the balls changed just a touch, if the charge on the moving electrons that drove the motor were a shade off, or if any of a number of other physical laws or constants had changed, the sequence would have come out differently. In other words, the result of this experiment depended on the nature of the Universe, the interplay of its laws, the value of its constants, and its unique history. Can we therefore conclude that "all the particular laws and regularities in nature are united in a single principle law: Somewhere in the universe this machine must create this particular sequence"?

If this argument seems peculiar and unjustified—that because of a highly improbable result we can conclude that behind the laws of the Universe lies a need to evoke this result—then your thinking cap is on straight. If the phrase at the end of the previous paragraph sounds familiar, perhaps that's because it is a paraphrase of the last sentence in A. Kuzin's article, "The Anthropic Principle," in the January/February 1999 issue of *Quantum*.

The anthropic principle is the teleological belief that the Universe was "tuned" to make the creation of life—and in particular, human, conscious life—inevitable. Teleology is the notion that causality works backwards, with a later event influencing a prior event—that somehow a system "knows" to head toward a "desired" outcome.

Kuzin begins his article by saying, "The discovery of the vastness of the Universe has led to a fundamental problem: Does a human being mean anything in this immense Universe?" As a scientist, I cannot justify the term "immense." Sizes, of course, are relative. We can say that the Sun is far larger than a rabbit. But is the Sun large? It's small relative to our gal-

Art by Yury Vashchenko

axy. Is the rabbit small? It's large relative to a proton or a protozoa.

The impulse to make the Universe revolve around us has ancient roots. One of humankind's great achievements was to discover that Earth is not the center of the Universe, not even the center of our solar system. More recently, we have come to understand that humans are not the end product of evolution; we are not a goal to which other life forms are only stepping-stones. We should not make a similar mistake and base our judgment of absolute sizes on the size of humans. In other words, there is no inherent measure of length, and the universe is neither immense nor tiny.

But that's merely a quibble. Kuzin's opening sentence also poses the seemingly deeper question of whether a human being "means" anything. What are the implications of a human being having a "meaning"? In fact, what is the "meaning" of any object? For an object *x* to have a meaning implies that *x* encodes, is a symbol for, or represents something else—say, *y*. The red dot labeled "You Are Here" on a map in the park symbolizes your location on the map. The red dot and its location are an object *x*, and the location of the map itself in the park is *y*. For the red dot to have meaning to us, we must realize that *y* is represented by *x*, and we must be able to interpret the map. In Kuzin's article we must ask, what entity would interpret the "meaning" of a human being?

Much more can be said about the question of meaning (a library's worth of books have been written on the subject), but even this elementary view reveals the problems in Kuzin's approach. Meanings are separate from the objects themselves. When a male coyote marks its territory with urine, the urine becomes a symbol for "I live here; keep away" to other male coyotes. What is a human being a symbol of, and to whom is the signal directed? Kuzin does not specify, and without such specification his words are empty.

Kuzin says that "the anthropic principle is the child of a mental experiment. In this experiment we assume some change in the natural laws and then see whether or not a human could exist in the modified world." Unfortunately for his thesis, my experiment with the ball-choosing machine and the sequence that it picked is just as valid. There is no difference between choosing humans and my machine's sequence as the "goal" of the "evolution of the Universe."

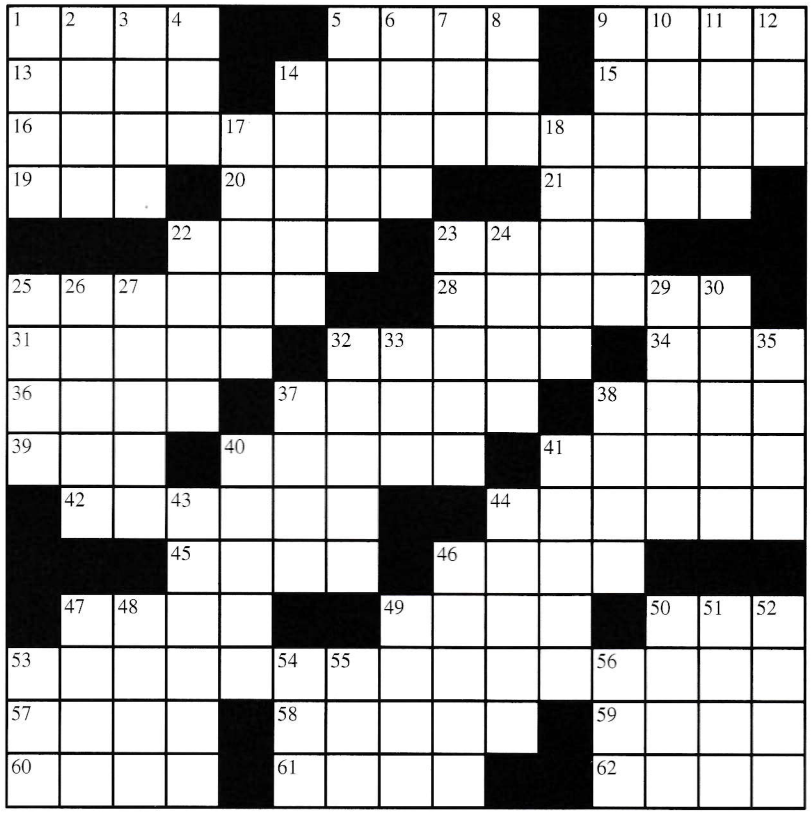
We also have to be very careful when we use the word "evolution" with physical systems. In biology, the term "evolution" describes change in species through natural selection. Evolution increases the fit between an organism and its environment. Changes in systems which do not have a selective mechanism, such as our galaxy, are not evolution in the same sense—they are merely changes over time. In common parlance, we find a third usage of "evolution": an improvement or advancement. Because of these varied meanings, we must be very careful in using the term in scientific discourse. In particular, we should remember that in biology, evolution does not imply increasing complexity, but a better fit to the current environment. For example, many cave-dwelling species are simpler than their ancestors (numerous cave-dwelling organisms no longer have eyes).

A large part of Kuzin's article is devoted to what is known about the sequence of events that make up the history of the Universe (what he calls the "evolution" of the Universe). The point of this extended discussion is that "due to a long chain of 'coincidences'... much carbon is produced [by stars], which is so important for life in the Universe," as if what seems to be a low-probability event can be explained only by its eventually being "needed" to create humans. Kuzin also points to the limited temperature range required for organic chemical reactions as signifying their unlikelihood. However, as my opening example shows, unlikelihood does not justify teleology. Once something has happened, the probability that it happened is 1. What Kuzin sees as essential for intelligent life reveals his anthropocentric bias and perhaps a lack of imagination regarding the possibilities for other forms of sentient beings. It's narrow-minded to assume that any sentient being would have to be similar to ourselves. The Universe has always had a great capacity to overturn our prejudices and presumptions, and I see no reason for it not to happen with our ideas on possible life forms.

The last section of Kuzin's argument observes that we have more equations representing physical laws than universal constants. This fact might point, as he believes, to potential unification of these laws. He does not stop to consider other alternatives though, such as other essential constants; perhaps whole classes of phenomena are as yet undiscovered. His observation leads him to venture that "the very structure of the natural laws hides some extremely important principle. At present we don't know how to describe this in mathematical language... Everything we presently know is just the consequence of this main principle." But he has presented no basis for thinking there is such a principle. Then he makes another leap beyond logic, and continues, "All the particular laws and regularities in nature are united in a single principle law: Somewhere in the Universe a human being must appear." As I have already noted, the last phrase can, with equal justification (or, rather, equal lack of justification) be replaced with, "All the particular laws and regularities in nature are united in a single principle law: Somewhere in the Universe a little plastic machine must create this particular sequence." Kuzin's version of the anthropic principle is neither a principle nor a part of science, but a quasi-religious belief disguised as a scientific argument.

Criss × cross science

by David R. Martin



Across

1 Surgeons William James and Charles Horace ___
5 Neurologist Antonio ___ Moniz (1874–1955)
9 Health clubs
13 Certain constellation: abbr.
14 Polycyclic compound
15 Plant geneticist Edward ___ (1879–1938)
16 Great flexibility
19 ___ number (of motor oils)
20 Scandinavian gods
21 Not any
22 Mine entrance
23 Type of parity
25 Shadow regions
28 1952 chem. Nobelist Archer ___
31 French composer
32 Mentally retarded person
34 Twelve grams of Carbon-12
36 Singer ___ James
37 909,038 (in base 16)
38 ___-particle duality
39 100 square meters
40 Direction finder
41 Astronomer Carl ___
42 ___ velocity
44 Strain's partner
45 Eskers
46 Elemental particle
47 Chess, e.g.
49 Turkish flag
50 Star Wars Program (abbr.)
53 Humidity above 100%
57 Publisher Conde ___ (1874–1942)
58 Name
59 Down: pref.
60 Mine products
61 Malayan boat
62 Anthropologist ___ Hrdlicka (1869–1943)

Down

1 Measure of inertia
2 Greenish-blue
3 Cry of fright
4 Spanish cheer
5 Fanfare
6 Equipment
7 Solution: abbr.
8 Collection
9 Trigonometric function
10 ___ production
11 Bryozoan colony
12 Pigpen
14 Orbital point
17 Plane detector
18 Like a noble gas
22 Width times length
23 Moslem ruler
24 1982 Physiol. Nobelist John ___
25 Carbamide
26 Unit of length (Brit. sp.)
27 Data units
29 Real or virtual follower
30 Exploding stars
32 Totaler
33 ___ culpa
35 Focusing device
37 Dihydroxyphenylalanine
38 Heat up
40 Monochromatic radiation source
41 Leaf pore
43 Small planets
44 One cubic meter
46 Bird wing part
47 Cluster bean
48 Church part
49 10^{-18}: pref.
50 Granitic layer
51 Bestow love
52 Indium arsenide
53 Stannous oxide
54 1 atm. and 0 °C: abbr.
55 It's mostly Nitrogen
56 ___ cycle (Kreb's cycle)

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE SEPTEMBER/OCTOBER PUZZLE

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

Three golds and two silvers in Italy

by Mary Mogge and Leaf Turner

THREE MEMBERS OF THE 1999 United States Physics Team, Peter Onyisi, Andrew Lin, and Benjamin Mathews, won gold medals at the XXX International Physics Olympiad held in Padua, Italy, July 18 through July 27. Natalia Toro and Jason Oh received silver medals. This was the second-best US finish in the 14 years the US has been competing. While the competition is among individuals, unofficial rankings placed the US third after Russia and Iran. Russia also had the top student, Konstantin Kravtsov, who scored 49.8 out of 50 points. Iran's team was the only team with five gold medals. Overall 291 students from 62 countries competed.

Peter Onyisi of Arlington, Virginia, was the top US competitor, placing 10th. He had previously won a bronze medal at the 1998 Olympiad in Iceland. Peter graduated from Phillips Exeter Academy where his physics teacher was Cynthia Beals. This fall he is attending the University of Chicago.

Two US representatives, Andrew Lin and Benjamin Mathews, tied for 14th place. Andrew Lin of Cheshire, Connecticut, was a silver medalist in last year's competition. He graduated from Choate Rosemary Hall where he studied physics with Jonathan Gadoua. He is currently attending Yale University. Benjamin Mathews of Dallas, Texas, graduated from St. Mark's School of Texas. His physics teacher was Stephen Balog. This fall Benjamin is a student at Cal Tech.

Silver medalist Natalia Toro of Boulder, Colorado, graduated from Fairview High School and was nominated by her counselor, Karen Petterson. Natalia won the top award of the Intel Science Talent Search this year. At 14 years of age, she is the youngest person to have ever won that award. Natalia is enrolled at the Massachusetts Institute of Technology. Jason Oh is a senior at the Gilman School in Baltimore, Maryland, this fall. His physics teacher is Edwin Lewis.

Selection and Training

The selection of the 1999 U.S. Physics Team began in January when high school teachers throughout the country nominated almost 1200 students. The first round of examinations in late January produced approximately 175 semi-finalists who were given a second screening examination in March. Using the results of the second examination, transcripts, and letters of recommendation, the 24 members of the team were selected.

The team members met at the University of Maryland for an eight-day extensive training camp in early June. Their activities at the camp included tutorials, laboratories, problem sets, examinations, and guest lectures on current research topics. At the end of the training camp, five team members were selected to represent the US at the Olympiad.

Medal winners (left to right): Peter Onyisi, Benjamin Mathews, Andrew Lin, Natalia Toro, Jason Oh.

They and alternate Nilah Monnier reconvened at the University of Maryland July 12–15 for a mini-camp devoted to enhancing their laboratory skills. On July 15, the five team representatives flew to Italy accompanied by coaches Mary Mogge and Leaf Turner and Director Bernard Khoury.

Padua—The City of Galileo

The travelers arrived in Italy two days before the start of the competition to reset their internal clocks six hours ahead to Central European Time. The first day in Italy was spent wandering with the pigeons and other travelers around Venice's Piazza San Marco under an incredibly clear blue sky. Then it was time to investigate gelato (the delicious, rich Italian ice cream) as an antidote to jet-lag.

The Olympiad was held at the University of Padua. Founded in 1222, it is the second-oldest university in Italy. Copernicus studied medicine there. Galileo taught at Padua from 1592 until 1610 and discovered the four satellites of Jupiter that bear his name. The rostrum Galileo's students constructed for him is still on display at the university. Both the Physics Department and a street in Padua are named for Galileo.

The Exams

The five-hour experimental examination on July 20 was an investigation of the properties of a torsion pendulum. The pendulum consisted of an outer cylinder and an inner threaded rod that was not removable but could be screwed in or out to adjust its length and rotational inertia. The students used their measurements of the pendulum's center of mass, equilibrium angle, and period to determine its torsion elastic constant, moment of inertia, and the length of the threaded rod.

The five-hour theoretical exam on July 22 consisted of three problems. The first dealt with a gas in a cylindrical container capped with a movable glass plate. Laser light entered and was absorbed by the gas inside. The students considered two theoretical models of the magnetic field due to a V-shaped current carrying wire in the second problem. The third problem modeled a method frequently used to accelerate space probes—the slingshot effect.

United States Physics Team and coaching staff.

Giotto and Gelato

When not challenged by interesting physics problems, the students toured the attractions of Padua and northeastern Italy. They saw the Giotto frescos that cover the walls and ceilings of the Scrovegni Chapel. Completed in 1305, the frescos marked a turning point in art. While their papers were graded and their scores debated, the students toured the Dolomites, an area of the Alps noted for rough, jagged pinnacles and spectacular scenery. They also visited Venice, 40 km away, rode a gondola, and returned laden with souvenirs for their family and friends.

Meals were eaten in the student canteens. Food was very good and abundant. A typical meal featured a "first plate" of pasta, a "second plate" of meat or fish with vegetables, a salad, and fruit for dessert. The travelers also had a chance to sample Italian pizza, which has much thinner crust, sauce, and cheese than the American version. Fortunately competition sites were spread out and everyone got a lot of exercise. Water fountains were nonexistent. Drinking water came in bottles and was "gassed" or "ungassed." The US travelers developed many ingenious ways to ungas a bottle of gassed water.

Prato della Valle is a huge 88,620-m^2 elliptical plaza containing a statue-lined canal. Paduans use it as an open air market during the day and in-line skating track during the evening. It was hot and humid the evening coaches and students met to discuss the just-finished experimental exam. Everyone decided to stay outside and be treated to gelato. Discussing physics and eating gelato while sitting on a canal bank in the middle of a giant roller rink was truly a most memorable experience.

The 1999 United States Physics Team

The other members of the US Physics Team (with their teachers and high schools) are: Owen Baker (Michael Morrill, Columbia HS, Maplewood, NJ), Raymond Cassella (Dominick Capozzi, Baldwin Senior HS, Baldwin, NY), Tanner Fahl (Carey Inouye, Iolani School, Honolulu, HI), Nicholas Guise (Penny Valentini, Centerville HS, Centerville, OH), Devon Haskell (Robert Shurtz, Hawken School, Gates Mills, OH), Steven Hassani

(Gregory Matthes, Robert E Lee HS, Springfield, VA), Charvak Karpe (Pratima Karpe, home schooled, Stillwater, OK), Abraham Kunin (Deborah Ormond, Virgil I Grissom HS, Huntsville, AL), Nilah Monnier (Caroline Evans, Brookline HS, Brookline, MA), Anthony Nannini (Alan Kersey, Waubonsie Valley HS, Aurora, IL), Paul Oreto (Caroline Evans, Brookline HS, Brookline, MA), Marc Popkin-Paine (Mark Kinsey, St. John's School, Houston TX), Tomokazu Sato (Jeff Levy, Horace Mann School, New York, NY), Alexander Schwartz (Mary Quinlan, Radnor HS, Radnor, PA), Katherine Scott (Virginia Baner, Montgomery HS, Skillman, NJ), Dmytro Taranovsky (Robert Siskind, Long Reach HS, Columbia MD), Ryan Timmons (Leonard Klein, Wylie E Groves HS, Beverly Hills, MI), Kevin Wang (Robert Shurtz, Hawken School, Gates Mills, OH), Joseph Yu (Glenn Malin, University HS, Irvine, CA).

Assisting the authors were Warren Turner—coach (Brunswick School, Greenwich, CT), Boris Zbarsky—junior coach (MIT undergraduate, member of the 1996 and 1997 teams, and gold medallist in 1997), Jennifer Catelli—senior lab assistant and Ryan McAllister—lab assistant, (both University of Maryland graduate students). The support staff is headed by Maria Elena Khoury and Patrick Knox at the American Association of Physics Teachers. Major financial support is provided by AAPT, the American Institute of Physics, and its member societies.

The XXXI International Physics Olympiad will be held in Leicester, England from July 8 to 16, 2000. If you are interested in applying or nominating a student and do not receive an application by early December, please contact Maria Elena Khoury at AAPT [telephone: (301) 209-3344 or email: mkhoury@aapt.acp.org].

Mary Mogge *(professor of Physics at California State Polytechnic University–Pomona) has been a coach of the US Physics Team since 1995 and is currently academic director.*

Leaf Turner *(physicist in the Theoretical Division of Los Alamos National Laboratory) served as senior coach this year and has been a member of the coaching staff since 1997.*

Bulletin Board

The milk bottle of human kindness

Relying on the kindness of your neighbors was the key to this month's CyberTeaser, which involved trading in empty milk bottles for full ones. To learn how to make the most out of your limited resources, turn to inside back cover. This month's winners are

Christian Grothaus (Bielefeld, Germany)
Steven Buczkoski (Malden, Massachussets)
Bruno Konder (Rio de Janeiro, Brazil)
Jorge G. Moya (Culiacan, Mexico)
Michael Marfil (Quezon City, Philippines)
Yiming Yao (New Westminster, British Columbia, Canada)
Nick Fonarov (Staten Island, New York)
Igor Astapov (Kingston, Ontario, Canada)
Jerold Lewandowski (Troy, New York)

Congratulations to our winners, who will receive a *Quantum* button and a copy of this issue.

Everyone who submitted a correct answer was eligible to win a copy of our brainteaser collection *Quantum Quandaries*. Visit http://www.nsta.org/quantum to find out who won the book, and while you're there, try your hand at the new CyberTeaser!

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Publication title: Quantum. (2) Publication No. 008-544. (3) Filing Date: 10/99. (4) Issue Frequency: Bi-Monthly. (5) No. of Issues Published Annually: 6. (6) Annual Subscription Price: $47.00. (7) Complete Mailing Address of Known Office of Publication: 175 Fifth Avenue, New York, NY 10010-7858. Contact Person: Joe Kozak, Telephone 212-460-1500 EXT 303. (8) Complete Mailing Address of Headquarters or General Business Office of Publisher: 175 Fifth Avenue, New York, NY 10010-7858. (9) Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor: Publisher: Gerald F. Wheeler, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000 in cooperation with Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010-7858. Editors: Larry D. Kirkpatrick, Mark E. Saul, Albert L. Stasenko, Igor F. Sharygin, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. Managing Editor: Kenneth L. Roberts, National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. (10) Owner: National Science Teachers Association, 1840 Wilson Boulevard, Arlington, VA 22201-3000. (11) Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities: None. (12)Tax Status. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 Months. (13) Publication Name: Quantum. (14) Issue Date for Circulation Data Below: September/October 1999. (15) Extent and Nature of Circulation: (a.) Total No. Copies (Net Press Run): Average No. Copies Each Issue During Preceding 12 Months, 7483; No. Copies of Single Issue Published Nearest to Filing Date, 7200. (b.) Paid and/or Requested Circulation: (1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541: Average No. Copies Each Issue During Preceding 12 Months, 4661; No. Copies of Single Issue Published Nearest to Filing Date, 4529. (2) Paid In-County Subscriptions: Average No. Copies Each Issue During Preceding 12 Months, 0; No. Copies of Single Issue Published Nearest to Filing Date, 0. (3) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales, and Other Non-USPS Paid Distribution: Average No. Copies Each Issue During Preceding 12 Months, 1080; No. Copies of Single Issue Published Nearest to Filing Date, 1060. (4) Other Classes Mailed Through the USPS: Average No. Copies Each Issue During Preceding 12 Months, 0; No. Copies of Single Issue Published Nearest to Filing Date, 0. (c.) Total Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 5741; No. Copies of Single Issue Published Nearest to Filing Date, 5589. (d.) Free Distribution by Mail: (1) Outside-County as Stated on Form 3541: Average No. Copies Each Issue During Preceding 12 Months, 136; No. Copies of Single Issue Published Nearest to Filing Date, 127. (2) In-County as Stated on Form 3541: Average No. Copies Each Issue During Preceding 12 Months, 0; No. Copies of Single Issue Published Nearest to Filing Date, 0. (3) Other Classes Mailed Through the USPS: Average No. Copies Each Issue During Preceding 12 Months, 0; No. Copies of Single Issue Published Nearest to Filing Date, 0. (e.) Free Distribution Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months, 772; No. Copies of Single Issue Published Nearest to Filing Date, 830. (f.) Total Free Distribution: Average No. Copies Each Issue During Preceding 12 Months, 908; No. Copies of Single Issue Published Nearest to Filing Date, 957. (g.) Total Distribution: Average No. Copies Each Issue During Preceding 12 Months, 6649; No. Copies of Single Issue Published Nearest to Filing Date, 6546. (h.) Copies Not Distributed: Average No. Copies Each Issue During Preceding 12 Months, 834; No. Copies of Single Issue Published Nearest to Filing Date, 654. (i.) (Sum of 15g. and h.): Average No. Copies Each Issue During Preceding 12 Months, 7483; No. Copies of Single Issue Published Nearest to Filing Date, 7200. (j.) Percent Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 86.34%; No. Copies of Single Issue Published Nearest to Filing Date, 85.38%. (16) Publication of Statement of Ownership will be printed in the Nov/Dec 1999 issue of this publication. I certify that all information furnished on this form is true and complete. (17) Signature and Title of Editor, Publisher, Business Manager, or Owner:

Dennis Looney
Senior Vice President and Chief Financial Officer
Date: 9/21/99

ANSWERS, HINTS & SOLUTIONS

Math

M276

Let all edges of the pyramid $ABCD$, except for AD, be of length 1. Let O be the center of the sphere circumscribed about this pyramid. All edges of pyramids $ABCO$ and $BCDO$ are of length 1; that is, each

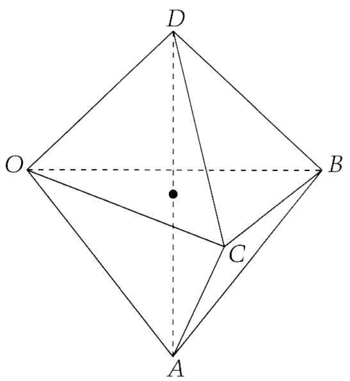


Figure 1

of them is a regular tetrahedron with unit edge. Edge AD equals twice the altitude of the unit regular tetrahedron and thus, its length is $2\sqrt{6}/3$.

M277

Let each of the fractions be equal to t. We have $x - 1 = t(xy - 3)$, $y - 2 = t(xy - 4)$, and $3 - x - y = t(7 - x^2 - y^2)$. Add these equations to obtain $0 = t(x - y)^2$. Thus either $t = 0$ or $x = y$. Then the solution proceeds in an obvious way. The system has two solutions: $(1, 2)$ and $(-1, -1)$.

M278

First, consider the case when $|\tan x| < 1$. Since the left side is $\pm 2 \tan x$ or $\pm 2 \cos 3x$, it is less than 2 in absolute value in this case, whereas the right side is greater than 2. Therefore, $|\tan x| \geq 1$, and the left side is $\pm 2 \tan x$. Thus the solutions to our equation satisfy one of the equations $\tan^2 x - 3 = \pm 2 \tan x$. Therefore, $\tan x = \pm 1$ or $\tan x = \pm 3$. It is easy to see that all solutions to these equations are also solutions to the initial equation.

Answer:

$$\frac{\pi}{4} + \frac{\pi}{2}k$$

and

$$\pm\arctan 3 + \pi k.$$

M279

Multiply both numbers by $2^{49}5^{21}$. We obtain $29^{200}2^{200}5^{21}$ and $3^{300}5^{300}2^{49}$. It is easily seen that $5^3 < 2^7$ (since $125 < 128$) and $(29 \cdot 2)^2 < (3 \cdot 5)^3$ (since $3364 < 3375$). Therefore, $5^{21} < 2^{49}$ and $29^{200}2^{200} < 3^{300}5^{300}$. Multiplying the first of these inequalities by the second, we obtain $29^{200}2^{200}5^{21} < 3^{300}5^{300}2^{49}$. Thus, $29^{200}2^{151} < 5^{279}3^{300}$.

M280

Point Q is the center of the circle inscribed in the given triangle. We assume that Q lies inside triangle ABM (fig. 2). Denote by P, L, and E the points of tangency of this circle with the sides of the triangle (as shown in the figure). Triangles APQ, MLQ, and AEQ are congruent right triangles (they are congruent by hypotenuse-leg). Therefore, $\angle QAP = \angle QML$. Quadrilateral $ACMQ$ can be inscribed in a circle, since $\angle QAP + \angle QMC = 180°$. If $\angle AQM = \gamma$, then $\angle ACM = 180° - \gamma$. Now, $\angle AQM$ has the minimum possible value if $\angle ACM$ has its maximum. In addition, $BA = BE + EA = BL + LM = BM$. Therefore, $BC = 2BA$. Thus we obtain the following problem: find the maximum value of angle BCA if $BC = 2BA$. If points B and C are fixed, then point A must belong to the circle centered at B with a radius equal to half of BC (fig. 3). The maximum possible value of angle BCA is attained at a point A_0 such that CA_0 is tangent to this circle ($\angle BA_0C = 90°$). For this triangle, $\angle BCA_0 = 30°$. Therefore, the minimum value of angle AQM is 150°.

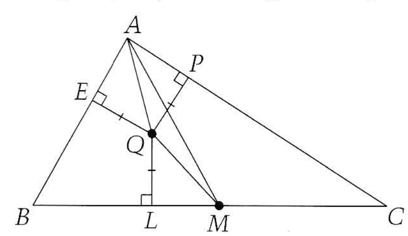


Figure 2

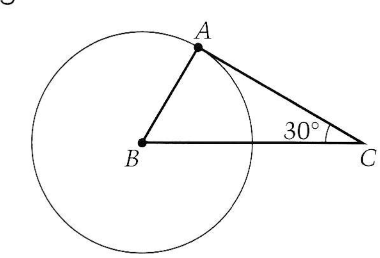


Figure 3

Physics

P276

Let's find the "minimal" orbit of a satellite. It should be tangent to the Earth's surface at point A, which is the nearest to the displaced center of mass (figure 4). Acceleration of the satellite at this point is perpendicular to the velocity $\mathbf{v}_1$ and is determined by the gravitational attraction of the "Earth":

$$a = \frac{GM}{(R-d)^2} = \frac{v_1^2}{R}.$$

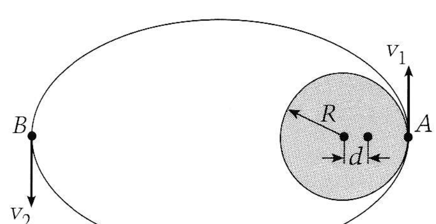


Figure 4

Note that we used R for the radius in the last term because the radius of curvature of the orbit cannot be less than the Earth's radius R. From this equation we can obtain the smallest possible velocity at point A:

$$v_1 = \frac{\sqrt{GMR}}{R-d}.$$

Now let's consider the farthest orbital point B. If we denote the altitude of the satellite above the Earth's surface by x, then the distance from the satellite to the center of mass for this point will be $R + d + x$. To determine the relationship between the speeds at the nearest and most distant points of the trajectory, we use angular momentum conservation (Kepler's second law):

$$v_2(R+d+x) = v_1(R-d),$$

and energy conservation:

$$-\frac{GMm}{R-d} + \frac{mv_1^2}{2} = -\frac{GMm}{R+d+x} + \frac{mv_2^2}{2}.$$

Note that the gravitational potential energy for the interaction of the satellite and the Earth is negative and there is nothing "dummy" about that!

Plugging in the value v_1 taken from the previous equation and eliminating v_2, we find the altitude x:

$$x = \frac{2d^2}{R-2d} \cong 3200 \text{ m}.$$

This is a very small altitude; therefore, the length of the semimajor axis of the orbital ellipse is practically equal to the radius of the Earth.

Thus the period of revolution T_1 is almost equal to $T_0 = 2\pi\sqrt{R/g} \cong$ 5060 s (the period of an orbit around the Earth along a circular trajectory of radius R). The ratio of these periods can be found with the help of Kepler's third law:

$$\frac{T_1}{T_0} = \left(\frac{R+x/2}{R}\right)^{3/2} \cong 1.0004.$$

Note: Figure 4 shows a very elongated ellipse, but our calculations showed that the trajectory should be almost circular.

P277

In addition to atmospheric pressure, the nitrogen inside the bubble is compressed by the pressure due to the surface tension $\Delta p = 8\sigma/d$, where d is the diameter of the bubble. This formula can be obtained most simply by "cutting" the bubble into two equal parts with a plane passing through the center. Let's consider the equilibrium conditions for these hemispheres. If the additional pressure inside the bubble is Δp, the halves are driven away from each other by a force $\Delta p\pi d^2/4$. On the other hand, they are attracted by the surface tension of the soap film, which acts along the circular perimeter πd. This force is equal to $2\sigma\pi d$ (the coefficient 2 appears because the soap film has two surfaces, the inner surface and the outer surface). Setting these forces equal, we get the value of the additional pressure in the bubble.

The bubble will float when the buoyancy (which is equal to the weight of the air displaced by the bubble at atmospheric pressure p_0) becomes larger than the weight of nitrogen in the bubble, which is compressed by the pressure $p_0 + \Delta p$. According to the ideal gas equation,

$$\frac{M_a p_0 \pi d^3}{6RT} \geq \frac{M_N(p_0 + \Delta p)\pi d^3}{6RT},$$

which gives

$$d \geq \frac{8\sigma M_N}{p_0(M_a - M_N)} \cong 10^{-4} \text{ m}.$$

P278

The combined capacitor is equivalent to two capacitors connected in series. We denote the electric fields in them by E_1 and E_2, the surface charge densities by σ_1 and σ_2, and the voltages across them by V_1 and V_2, respectively. The charges on the two sides of the middle plate have opposite signs, so the total charge on the middle plate is

$$q = (\sigma_2 - \sigma_2)S.$$

When the charge distribution is steady, we can use the formulas of electrostatics.

Therefore,

$$E_1 = \frac{\sigma_1}{\varepsilon_0\kappa_1},\ E_2 = \frac{\sigma_2}{\varepsilon_0\kappa_2},$$

$$V_1 = E_1 d_1,\ V_2 = E_2 d_2,\ V_1 + V_2 = V.$$

On the other hand, since electric current flows in the circuit, the capacitors can be considered as conductors with the corresponding resistances $R_1 = \rho_1 d_1/S$ and $R_2 = \rho_2 d_2/S$. According to Ohm's law,

$$V_1 = I\frac{\rho_1 d_1}{S},\ V_2 = I\frac{\rho_2 d_2}{S}.$$

Solving the system of eight equations simultaneously, we get

$$q = \frac{\kappa_2\rho_2 - \kappa_1\rho_1}{d_2\rho_2 + d_1\rho_1}\varepsilon_0 VS.$$

P279

In the first case only the coil connected to the voltage source generates a magnetic field. The magnetic flux in this coil is

$$\Phi_1 = LI_1,$$

where L is inductance of the coil and I_1 is the current in it. Clearly the magnetic flux in the third coil is also proportional to current I_1, that is

$$\Phi_3 = MI_1,$$

where M is the mutual inductance of the coils. The voltage ratio for these coils is

$$\frac{V_3}{V_1} = \frac{\Phi_3'}{\Phi_1'} = \frac{\Phi_3}{\Phi_1} = \frac{M}{L} = \frac{1}{2}.$$

In the second case electric current flows in two coils: in the primary coil (connected to the voltage source) and in the short-circuited coil. Neglecting the resistance of the circuit, we get:

$$\dot{\Phi}_1 = L\dot{I}_1 + M\dot{I}_2,$$

$$\dot{\Phi}_3 = M\dot{I}_1 + M\dot{I}_2,$$

$$\dot{\Phi}_2 = M\dot{I}_1 + L\dot{I}_2 = 0,$$

and so

$$\dot{I}_2 = -\dot{I}_1\frac{M}{L}$$

and

$$\frac{V_3^*}{V_1^*} = \frac{\Phi_3^*}{\Phi_1^*} = \frac{MI_1^* + MI_2^*}{LI_1^* + MI_2^*} = \frac{M - M^2/L}{L - M^2/L}$$

$$= \frac{M}{L}\,\frac{1 - M/L}{1 - M^2/L^2} = \frac{M}{L}\,\frac{1}{1 + M/L} = \frac{1}{3}.$$

Therefore, in this case the voltmeter will read one-third of the source voltage.

P280

The object to be observed is situated below the boundary of two media, the glass and air. It would be wrong to consider the paper as an object immersed in glass. If that were the case, the paper could be observed for all values of the index of refraction, because the rays would radiate in the glass in all directions, and some fraction of them would leave the prism through the transparent face without being trapped by total internal reflection.

Of course, the rays radiated by the object under the glass also travel in all directions, but they cross the air-glass boundary and are refracted. The rays only enter the prism within a cone with an apex angle given by $2\alpha_0 = 2 \arcsin (1/n)$. Some of the rays may not reach the face BC (see fig. 5).

If $\alpha_0 > \pi/4$, that is, $n < 1/\sin (\pi/4) = \sqrt{2}$, the entire text will be seen through BC. Even the points near vertex A will send rays to BC, and their angles of incidence will be smaller than $\pi/4$ (that is, they will leave the prism after refraction).

At higher n (and, accordingly, smaller α_0) all the rays of the cones emitted by some points near the vertex A will hit the frosted face AB. To find the invisible part AD, note that $\angle ABD = \pi/4 - \alpha_0$, and recall the law of sines for triangle ABD, from which we can determine the visible fraction of the text $k = DC/AC$:

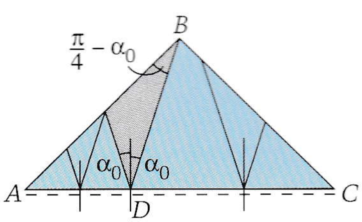


Figure 5

$$\frac{1-k}{\sin(\pi/4 - \alpha_0)} = \frac{1\ \sqrt{2}}{\sin(\pi/2 + \alpha_0)},$$

$$k = \frac{1}{2}\left(1 + \tan \alpha_0\right) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{n^2 - 1}}\right).$$

However, in order for at least some of the rays that hit face BC to be able to cross it, the smallest angle of incidence on this face, $\pi/4 - \alpha_0$, must be less than α_0, that is, $\alpha_0 > \pi/8$, and $n < 1/\sin (\pi/8) \cong 2.61$. For the limiting value n,

$$k_{\min} = \frac{1}{2}\left(1 + \tan \frac{\pi}{8}\right) \cong 0.7.$$

Thus the entire text is seen when $n < \sqrt{2}$, while at $\sqrt{2} < n < 1/\sin (\pi/8)$ only a fraction

$$k = \frac{1}{2}\left(1 + \frac{1}{\sqrt{n^2 - 1}}\right)$$

of the text is visible. When $n > 1/\sin (\pi/8)$, the text cannot be seen at all.

Brainteasers

B276

It is not hard to see that the family can drink 6 + 1 = 7 bottles of milk, and then it will have three empty bottles left. Then, the family can borrow one empty bottle, exchange for one more bottle of milk, drink it, and then return the bottle borrowed. Thus, the family can drink eight bottles of milk.

B277

Yes, it is possible. For example, a regular hexagon can be covered by six squares as shown in figure 6 (the side of each square equals the side of the hexagon).

B278

See figure 7.

B279

Yes, it is possible. For example, an isosceles triangle with a base of length 8000 and the altitude to the base of length 0.5 has an area of 2000. Readers are invited to verify that each altitude is less than 1.

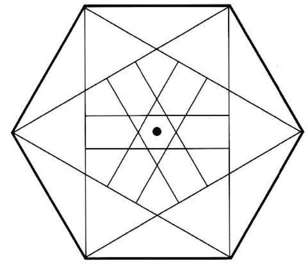


Figure 6

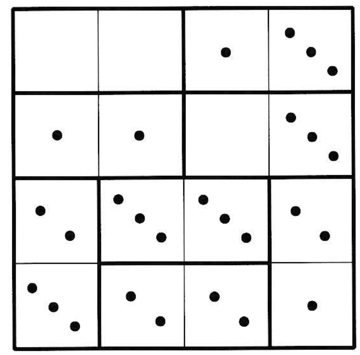


Figure 7

B280

The apparent size of an object depends not only on its inherent size but also on the distance from the observer. Indeed, you can "cover up" the moon with a match head. Everything you see from the train moves relative to you the same distance in a unit of time. However, the farther the object is away, the smaller this distance appears to be. Thus the apparent velocity decreases with distance from the observer.

Investigating Air

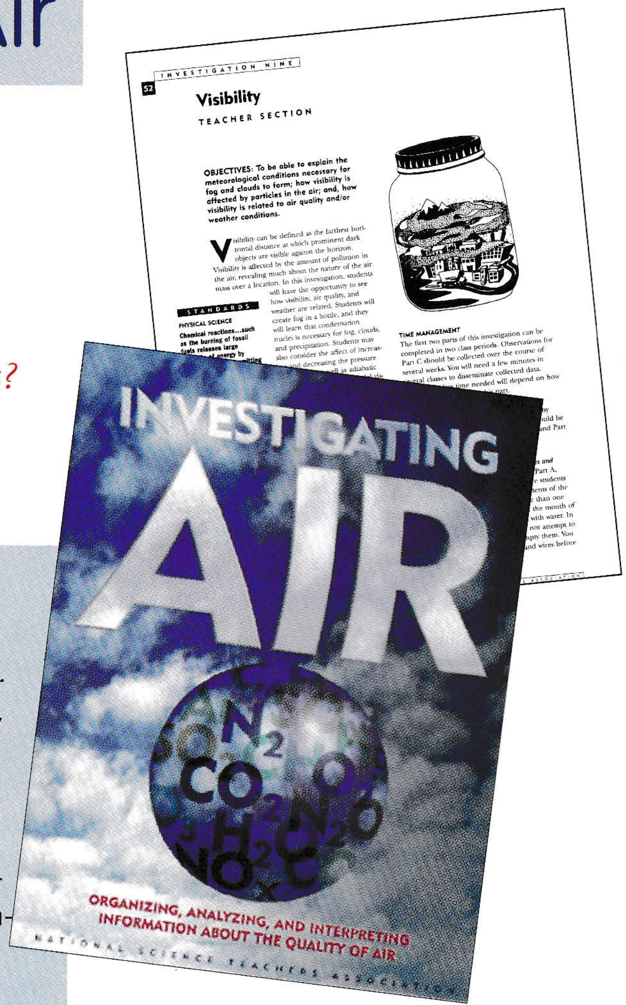
What is air made of?

How does air work?

Do humans impact air quality?

How does air quality impact humans?

Do we have a right to clean air?



Investigating Air provides the investigations and information that lead students to the answers to these and other questions. Students use data generated by researchers, and generate their own data, in situations analogous to real air quality issues—increases in atmospheric CO_2, acid deposition, ozone issues, and visibility. This resource culminates in a decision-making investigation that builds on the science and equity knowledge students have developed.
(grades 9–12, 1998, 78 pp.)

Produced by the National Science Teachers Association with support from the US Environmental Protection Agency.

#PB148X **$15.95**

Table of Contents

- Introduction
- Investigations *(Each investigation has both a teacher and student section.)*
  1) School Energy Consumption
  2) Home and Personal Energy Consumption
  3) Calculating Individual Air Intake and Expenditure & Measuring Indoor Air Volume and Quality
  4) Measuring Air's CO_2 Content
  5) Energy Conversion & the Internal Combustion Engine
  6) Air Quality and the Weather
  7) Air Quality, Acid Deposition, and pH
  8) Effects of Acid Deposition
  9) Visibility
  10) Deciding on Public Regulatory Policies
- Bibliography and Resources

P999Q2

To order, call 1-800-722-NSTA

NSTA Pub Sales, 1840 Wilson Blvd.
Arlington, VA 22201-3000
www.nsta.org/scistore