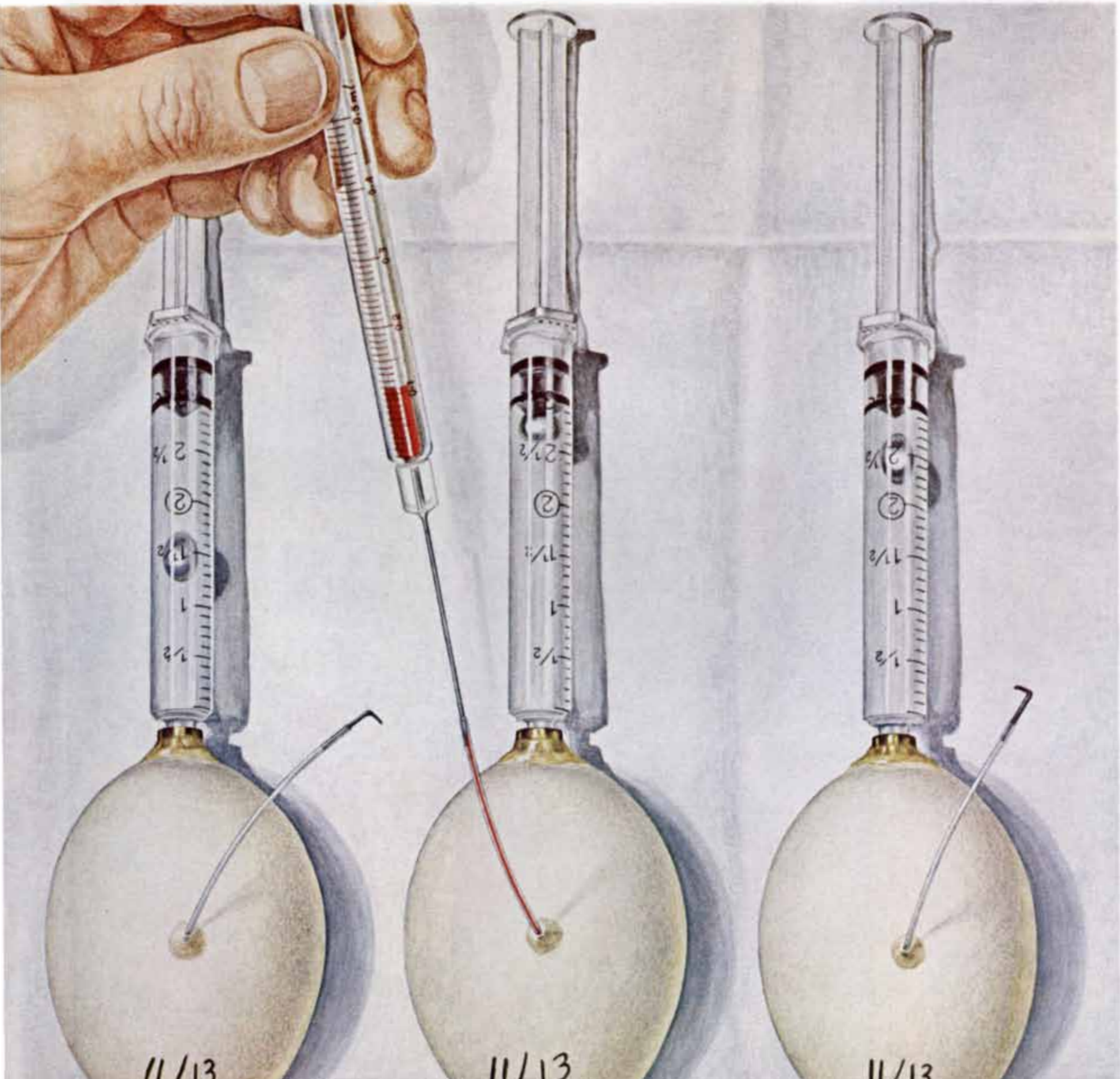
SCIENTIFIC AMERICAN



HOW EGGS BREATHE

$1.50

February 1979

With nearly everything going up, it's good to know that one thing is still a bargain.

That bargain is New York Life insurance.

The fact is, a dollar's worth of New York Life insurance costs about the same today as it did ten years ago – remarkable in a time of continuing inflation.

Of course, your income and standard of living have risen over the years. That means you probably need *more* protection. But life insurance remains the best way – usually the only way – to give your family real financial security.

Your New York Life Agent has a variety of policies to meet your needs, your family's needs, your business needs – at a cost you can afford. Call today.

New York Life. For all of your life.

New York Life Insurance Company, 51 Madison Avenue, New York, New York 10010. Life, Group and Health Insurance, Annuities, Pension Plans.

THE ONLY THING IT HAS IN COMMON WITH OTHER 7-INCH TAPE DECKS IS THE SIZE OF ITS REELS.

Pioneer's new RT-707 has a lot more in common with today's most sophisticated 10-inch tape decks than it does with most 7-inch tape decks.

Because unlike other 7-inch tape decks, the RT-707 isn't filled with 15 year old ideas.

Take the drive system of the RT-707.

Instead of the old fashioned belt-drive system, the RT-707 is driven by a far more accurate and efficient AC Servo direct-drive motor. This motor generates its own frequency to help correct even the slightest variation in tape speed. Which all but eliminates wow and flutter. And because it doesn't generate heat like the belt-driven "dinosaurs" it doesn't need a fan. So all you'll hear is music with a clarity and crispness not possible on any 7-inch, or many 10-inch tape decks.

Our direct-drive system also makes pitch control possible. To help you regulate the speed of the tape and give you greater control over your recordings.

With technology like this it shouldn't surprise you that our super-sensitive heads will deliver frequencies from 20 to 28,000 Hertz. And our pre-amp section is built to handle 30 decibels more than any other 7-inch tape deck without distorting.

But great sound isn't everything.

As you can see, the RT-707 is smaller and more compact than other tape decks. It's also rack-mountable. And unlike *any* other tape deck, it's stackable. So it'll fit right in with the rest of your components.

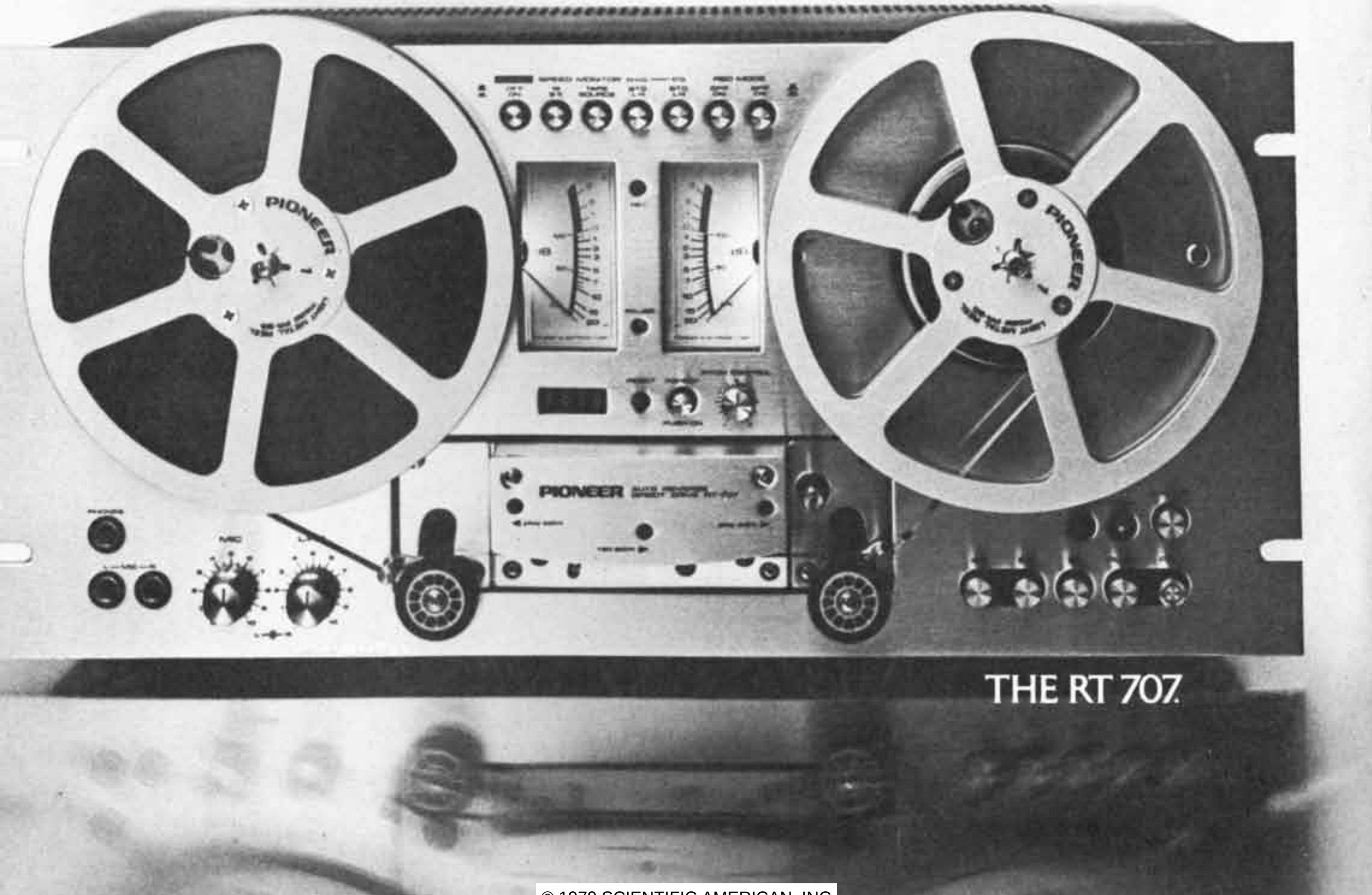
But frankly, all the revolutionary thinking that went into the RT-707 wouldn't mean much if it weren't also built to fit comfortably into your budget. It is.

See your Pioneer dealer for a closer look at this extraordinary 7-inch tape deck.

We think you'll find the only things that the RT-707 has in common with other 7-inch tape decks is the size of the reels.

And the size of the price.

High Fidelity Components
PIONEER
WE BRING IT BACK ALIVE.



Established 1845 SCIENTIFIC AMERICAN February 1979 Volume 240 Number 2

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES

Language Translator

Communicate in any language without saying a word, or learn a new language with the world's first computerized language translator.

Learning languages is not easy. It takes books, classes, cassettes, and hard work.

Now, however, you have a choice. You can communicate in a foreign country without speaking the language, or you can learn the language more easily thanks to a new electronic miracle.

Craig Corporation has just introduced the Translator–a pocket-sized personal computer that stores up to 7,000 words and translates them from one language to another at the touch of a button.

You enter a word in English on the alphabetical keyboard, press the translate button, and the word is shown on the display in the foreign language. You have a choice of languages such as Japanese, French, German, Italian, and Spanish. And you can translate from any one language into another.

You program the unit with four small memory capsules which contain approximately 1500 words each. By swapping capsules, you can change languages, so the unit will hold a total of 7,000 words in its four memory capsules.

WORLD COMMUNICATION

The Craig Translator is truly a pocket-sized translation computer that will help you communicate in practically every major nation in the world–even if you don't want to learn the language. There are a series of numbers listed on the back of the computer which correspond to 50 of the most commonly used and most practical phrases. 25 of the phrases are complete statements such as: "How much does this cost?" The remaining 25 statements require one or two additonal words like: "May I please have the..." You simply press the number for the phrase you want and then enter the missing word on the keyboard which has keys with both letters and numbers.

When you press the translate button, your phrase is flashed on the display in the foreign language. If the phrase is longer than the 16 character width of the screen, it rolls to the left–just like on a movie marquee.

Armed with the 50 phrases and up to a 7,000 word vocabulary, you can take your small information retrieval system throughout the world and literally carry on complete conversations without saying a word. Just enter the translated phrase or word and show it to the person with whom you are communicating. Have that person enter the answer, and before long you'll be carrying on complete conversations. You can negotiate prices at an outdoor market, order from menus, clarify statements, read newspapers, and street signs.

Phrase books are handy for travelers because all the most common words and phrases are indexed by category. The Craig Translator is also indexed, but for instantaneous retrieval. Just enter the subject, press a button, and you can recall a subject and its related words in alphabetical order. You can index travel, shopping, business, or medical terms and scan them faster than turning the pages of a book.

FOR THOSE WHO WANT TO LEARN A LANGUAGE

Learning a language is not easy, and the Craig Translator is not a substitute for practice and hard work. The Translator is, however, a very valuable teaching aid–one that can accelerate the learning process and teach you more vocabulary faster.

The most rewarding aspect of learning a language with the Craig system is "memory retention." When you need to know a word and can look it up at that very moment, your chances of remembering it are several times greater than in a classroom environment.

The Craig Translator not only gives you the answer, it does so immediately–at that moment–so it provides a very useful learning function. And it's fun to use and carry with you all day long.

The Craig Translator uses four systems of reinforcement. Reinforcement is a teaching concept in which the teacher praises you when you're right and gives you the correct answer when you're wrong.

FOUR TEACHING SYSTEMS

The Craig's four reinforcement systems are:

Frequency You can sort out words by how frequently they come up in normal conversation and then learn these words. Each one will be flashed on the display for a few seconds, and you say aloud the correct translation.

Spelling A word is flashed in English and you guess the spelling of the foreign word which you then enter on the keyboard. If you enter it wrong, the unit tells you and then flashes the correct answer on the display.

Category You can learn by category. Select a subject and access the associated words alphabetically. If you are preparing for a trip, you can learn all the words that relate to travel.

Alphabet You can learn by the alphabet. Start at any letter and the unit will display all the words that start with that letter.

ALL THE ANSWERS

There are other features that make the Craig Translator an outstanding teacher:

Spelling Program If you enter a misspelled word, the display will flash a question mark. By pressing a key, you instruct the unit to find the correct spelling. On the display will flash a series of words it thinks are the correct ones, and you simply select the right one.

Double Meanings When you enter an English word that has two meanings in the foreign language, the unit will ask you which meaning by listing the possible choices. Again you select the right one by pressing a key. For example, "watch" would be two different things depending on whether it was meant as a verb or a noun.

MEMORY CAPSULE

The languages are contained in small memory capsules four of which hold up to 256,000 bits of information or 7,000 words. The Craig Translator will accept other data so that eventually you will be able to store complete dictionaries, recipes, calorie equivalents, useful statistics, and other learning programs and then display this data on your pocket-sized information retrieval system. Language capsules cost only $24.95 each–about the cost of a few language textbooks, so the unit can inexpensively grow to fit many different applications.

You don't even have to read the instructions to operate the unit. A free starter Memory Capsule that comes with the unit will tell you how in six different languages. The capsule also contains the programs necessary to use your unit as a calculator or for currency conversions. It contains the metric conversion tables and 20 words or phrases in six languages—words such as hello, goodbye, thank you, etc.

ADVANCED TECHNOLOGY

There was no existing display that could be used in the Translator, so the display had to be developed exclusively for Craig. The letters appear in bright fluorescent blue/green and are very easy to read.

The unit, its memory, its Memory Capsule system—all represent major breakthroughs in technology. But with all its sophisticated electronics, it was built simple enough to be easy and fun to use.

MANY PERSONAL USES

If you are a language student or a frequent traveler, the Craig Translator would be the perfect companion. That might appear obvious. But what about the shop owner who has to communicate with people from other countries? Or the grandparents looking for the perfect gift for a young high school student learning a new language?

And don't overlook the young grammar school child. With the Craig Translator and a little tutoring, you'll be amazed at how quickly he or she picks up a language.

You'll really appreciate the Translator the first time you use it. It's like a friend—always ready to give you an answer when you need one. It's your interpreter in a foreign country, your resident genius with its huge data bank, and your personal advisor with its indexing system.

TRY ONE FIRST

We would like to offer a suggestion. Order the Translator on a special 30-day trial. Use it to begin learning a language, or see how easy it is to communicate with somebody who doesn't speak your language. Take it on a business trip to Europe or Japan. Take it shopping with you. Use it at a restaurant, or negotiate a business deal with it. See how much fun it is to learn a language and how much faster your vocabulary increases.

After you've really discovered how valuable a friend your Craig Translator can be, then decide if you want to keep it. If not, return it anytime during our 30-day trial period for a prompt and courteous refund including your $2.50 postage and handling. There is no risk.

JS&A offers you the opportunity to own, use, and experience the world's first pocket-sized language translator. But don't wait. The demand for the Craig Translator may be great, and since the announcement in this magazine is one of the first ever made on this product, we urge you to place your order promptly. Deliveries will start in January, 1979 and then only on a limited basis.

The Craig Translator is manufactured by Craig Corporation—a substantial public company. Both JS&A and Craig Corporation have been doing business together since 1971 when JS&A introduced the world's first pocket calculator—the Craig 4501. Back in 1971, we said that the technology represented in the first pocket calcultor was "...the most important electronic breakthrough since the transistor." The Craig Translator represents another quantum leap in technology, and we are proud to be associated with Craig Corporation in its introduction.

Craig Corporation has a complete service-by-mail facility should service ever be required. Just slip your unit in its handy mailer, and it will be repaired and shipped back to you promptly. You shouldn't have a single problem with your unit, but it's reassuring to know that even service is an important consideration in this program. Your unit is also backed by a one-year parts and labor limited warranty.

JS&A is America's largest single source of space-age products—further assurance that your modest investment is well protected.

To order your Craig Translator, send your check for **$199.95** plus $2.50 for postage and handling. Credit card buyers may call our toll-free number below. With each unit you will receive a free starter cartridge plus the English language cartridge, an AC adapter, and carrying case. Four rechargeable "AA" batteries cost only $12.40. Or you can use any Four "AA" cell alkaline batteries.

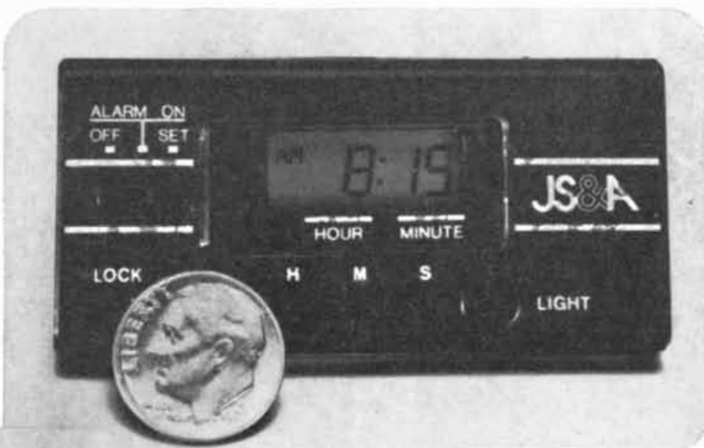
You may also order the other languages at **$24.95** each. Please specify French (order number **5121**), German (**5131**), Italian (**5141**), Spanish (**5151**), and Japanese (**5161**). Other languages will be available later.

Remember, the unit holds four cartridges at the same time, and you will receive from us a list of cartridges that will be available in the future. Only one cartridge in each language is available now, although later, more advanced vocabulary capsules will be available.

For the first time in history, Americans can carry with them the information contained in volumes of books and communicate with this information faster and more efficiently than was ever thought possible. Join the era of the real personal computer. Order a Craig Translator at no obligation, today.

JS&A ® PRODUCTS THAT THINK®
Dept.SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-7000
Call TOLL-FREE 800 323-6400
In Illinois Call .... (312) 564-7000

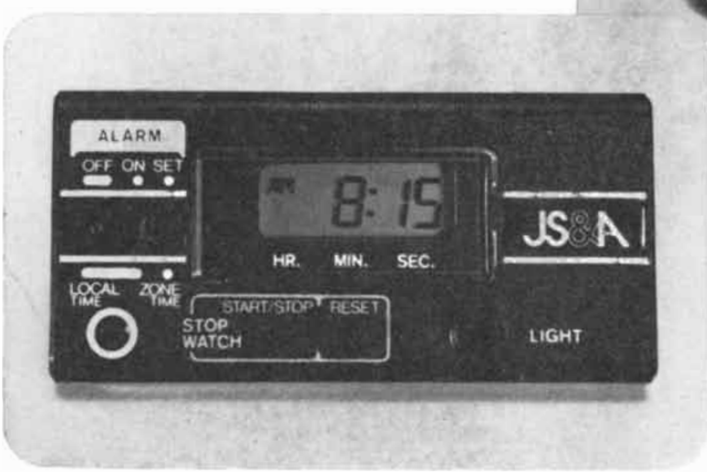
Mini Travel Alarm



It's small. And because it's small, it fits anywhere. In your briefcase or in your pocket.

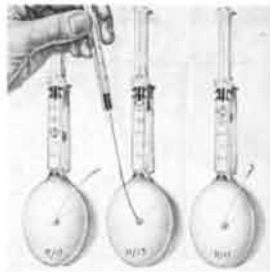
The new JS&A Mini Travel Alarm measures only 3/8" x 1¼" x 2½" and has a small easel support on the back. Just set the alarm, and the electronic beep will wake you up. The clock movement is totally solid-state, and a built-in night light lets you view the time in the dark.

But the JS&A Mini Travel Alarm does more. First, it makes a great pocket watch. The small imitation black leatherette carrying case that comes with the unit has a window so you can view the time even when the unit is in its case. Secondly, it tells accurate time—within fifteen seconds accuracy per month. And finally, it's inexpensive—only **$29.95** complete with carrying case and two readily-available hearing aid batteries. It makes a perfect gift for everyone on your gift list.



There is also a deluxe version for **$39.95** with a built-in timer and dual time zone capability. You can now display one time while keeping the second time in memory.

Each model has a one-year warranty. Order by sending your check to the address shown above or call our toll-free number above. Please add $2.50 for postage and handling and Ill. residents add 5% sales tax.



THE COVER

The painting on the cover depicts a technique for measuring the pressures of oxygen and carbon dioxide in the chicken egg a few days prior to hatching (see "How Bird Eggs Breathe," by Hermann Rahn, Amos Ar and Charles V. Paganelli, page 46). Eighteen-day-old chicken eggs, each specially fitted with a metal connector and a fine plastic catheter, are removed temporarily from an incubator for the experiment. A plastic syringe is attached to the connector in order to sample the gas in the air cell at the blunt end of the egg. (After the sample has been taken the syringe is tilted so that the drop of mercury inside seals off the opening, preventing the gas sample from being contaminated with atmospheric air.) At the same time the smaller glass syringe is connected to the plastic catheter and a sample of oxygenated blood is removed from a blood vessel in the chorioallantois, a placenta-like membrane within the egg that mediates the exchange of gases between the atmosphere and the developing chick embryo. After samples of blood and gas have been taken from each egg the concentrations of oxygen and carbon dioxide in the samples are determined. The experiment was designed by H. Tazawa of Yamagata University in Japan.

THE ILLUSTRATIONS

Cover painting by Enid Kotschnig

Page	Source
16–41	Ilil Arbel
43	Itek Corporation
44	Ilil Arbel
47	J. H. Becking, Institute for Atomic Sciences in Agriculture, Wageningen, the Netherlands
48	Dennis R. Atkinson, State University of New York at Buffalo
49–50	Patricia J. Wynne
51	Ewald Weibel
52–55	Patricia J. Wynne
57	Aimé Bocquet, University of Grenoble
58	Alan D. Iselin
59	Aimé Bocquet, University of Grenoble (*top*); Alan D. Iselin (*bottom*)
60	Alan D. Iselin
61–62	Aimé Bocquet, University of Grenoble
63	Alan D. Iselin
64	Aimé Bocquet, University of Grenoble
77–78	Richard Pardo, American Forestry Association (*top*); Michael Wotton, Weyerhaeuser Company (*bottom*)
79	Ilil Arbel
80–90	Tom Prentiss
93–110	George V. Kelvin
113	Claudio Rebbi and Laurence Jacobs, Brookhaven National Laboratory
114–116	George V. Kelvin
119	Charles R. Carrigan and David Gubbins
121–126	Allen Beechel
135	Robert W. Hardy (*top left*) and Kiyoteru T. Tokuyasu (*top right and bottom*), University of California at San Diego
136–142	Bunji Tagawa
149	Ralph Morse, courtesy of Warren Anatomical Library, Harvard Medical School
150–153	Ralph Morse, courtesy of Francis A. Countway Library of Medicine
154	Ilil Arbel
155	Courtesy of Massachusetts General Hospital
160	*The Louisville Times*
162–165	Michael Goodman

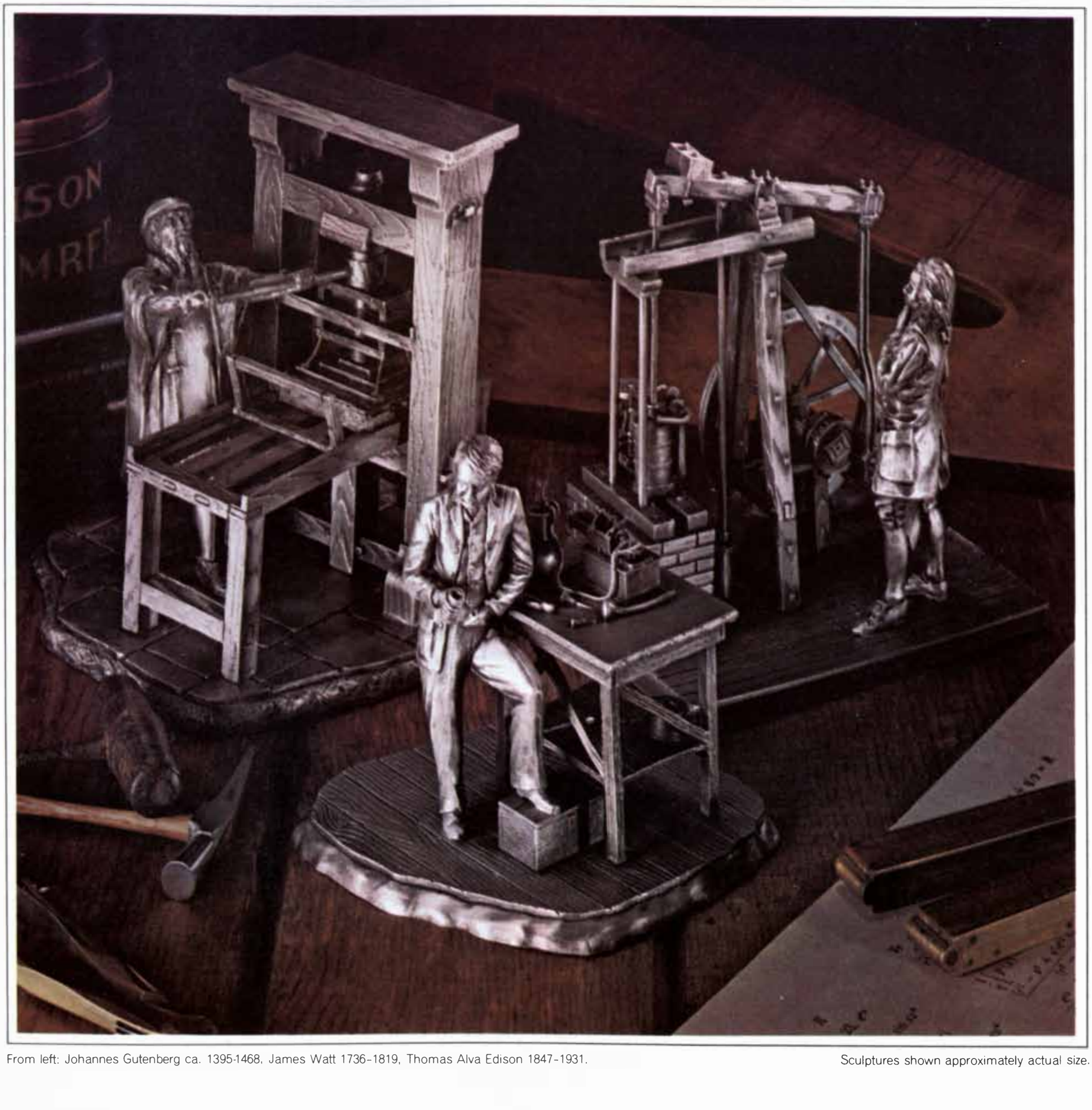
One of the world's best engineered cars...plus.

This is Eldorado at its best—with an unprecedented combination of driving features found on no other U.S. car. **FRONT-WHEEL DRIVE** for impressive traction in rain, snow or sleet. And since the floor is flat, there's added roominess, too. **FOUR-WHEEL INDEPENDENT SUSPENSION** as each wheel reacts independently to road variations for a smooth ride. **FOUR-WHEEL DISC BRAKES** that automatically self-adjust with every application. And an advanced level of sophistication in automotive electronics that combines **ELECTRONIC LEVEL CONTROL** . . . **ILLUMINATED ENTRY SYSTEM** . . . **TWILIGHT SENTINEL** . . . an **ELECTRONIC-FUEL-INJECTED ENGINE** with an on-board analog computer and **ELECTRONIC SPARK SELECTION**. All this standard. There are even side-window defoggers. And a Diesel V8 engine is available. (Eldorados are equipped with GM-built engines produced by various GM Divisions. See your Cadillac dealer for details.) Plus . . . Cadillac craftsmanship, beauty, luxury and prestige. Eldorado. You may choose it just because it's so beautiful.

To Commemorate Its 75th Anniversary, *The Deutsches Museum* presents

The Inventions that Shaped the Modern World

An official collection of ten superbly crafted pewter sculptures... honoring the genius of mankind



From left: Johannes Gutenberg ca. 1395-1468, James Watt 1736-1819, Thomas Alva Edison 1847-1931.

Sculptures shown approximately actual size.

A single, limited edition.

Available solely by subscription application.

Limit: One collection per subscriber.

Subscription deadline: February 28, 1979.

This year marks the 75th Anniversary of The Deutsches Museum—the largest museum of science and technology in the world. To observe the occasion in an appropriately unique and memorable way, the Directors of the Museum have authorized the creation of an unprecedented commemorative issue—honoring:

"THE INVENTIONS THAT SHAPED THE MODERN WORLD"

This official commemoration will comprise ten superbly crafted pewter sculptures—each re-creating an historic "moment of invention." In all, they will pay homage to ten of the most important landmarks in the history of modern science—depicting the inventions themselves and the men of genius credited with their creation. They include:

THE PRINTING PRESS
Johannes Gutenberg

THE TELESCOPE
Galileo Galilei

THE STEAM ENGINE
James Watt

THE CAMERA
Louis Daguerre

THE X-RAY
Wilhelm Roentgen

THE PHONOGRAPH
Thomas Alva Edison

THE WIRELESS TELEGRAPH
Guglielmo Marconi

THE TELEPHONE
Alexander Graham Bell

THE AUTOMOBILE
Gottlieb Daimler

THE AIRPLANE
The Wright Brothers

Each of the sculptures crafted for the collection will be a wholly *original work of art*—created exclusively for this 75th Anniversary celebration. Moreover, the collection will reflect the most scrupulous standards of *scholarship*, for each sculpture will be fully authenticated for scientific accuracy by the Museum. Indeed, the beauty and detail that distinguish each of the ten sculptures—the costumes . . . the facial expressions . . . and especially the detail lavished upon the actual inventions themselves—pay tribute to the highest criteria of science and art.

So vividly detailed, in fact, are these ten extraordinary sculptures that you will actually be able to see the first crude type form developed by Gutenberg for his printing press . . . the innovative separate condenser that proved so vital to the success of James Watt's steam engine . . . virtually every pipe and bolt of the historic internal combustion engine that was destined to establish Gottlieb Daimler as the father of the automobile. . . .

The world's foremost creator of commemorative issues

To ensure the creation of a commemorative issue that will retain its power and significance for generations to come, The Deutsches Museum has enlisted the skills and talents of an organization whose name has become synonymous with many of today's most prestigious commemoratives: The Franklin Mint. Hence "The Inventions that Shaped the Modern World" will be the direct creation of the very artists who are universally recognized as the contemporary world's foremost inheritors of the great Renaissance tradition of medallic sculpture.

And when the sculptors of The Franklin Mint have completed their commission, each of these exquisitely crafted pewter sculptures will be individually hand-finished—to bring out every feature and detail . . . the subtlest nuances of the artist's work . . . as well as the rich, gleaming luster of the sculptured metal itself.

The assurance of lasting rarity

Furthermore, "The Inventions that Shaped the Modern World" will be created in only one, extremely limited edition—to be issued solely in honor of the 75th Anniversary of The Deutsches Museum, and made available only by subscription application.

In addition, there will be a further limit of just one collection per subscriber. Consequently, the entire edition will exactly equal the precise number of valid applications entered before the subscription rolls close.

To certify the official nature of the collection, the base of each sculpture will bear the emblem of The Deutsches Museum and identify the creator of the invention depicted. Moreover, a Certificate of Authenticity bearing the signature of The Director-General of the Museum—attesting to the limits of the edition—will accompany the sculptures. Subscribers will also receive authoritative background literature on the evolution of each invention and its enduring impact upon mankind.

An heirloom collection of universal significance

"The Inventions that Shaped the Modern World" has been expressly conceived as a commemorative of truly *international* significance — fully reflective of the overwhelming universal importance of the subject it honors.

For collectors who recognize and appreciate the beauty of sculptured art . . . for parents who seek to imbue their children with a keener understanding of the world they inhabit, and a deeper awareness of man's potential for greatness . . . indeed, for everyone who is enthralled by the story of civilization itself . . . there can be few acquisitions more inspiring, more enlightening, more lastingly rewarding.

Subscription deadline: February 28, 1979

Following the absolute closing date of this single edition, "The Inventions that Shaped the Modern World" will never be offered again in the United States. Please note that you need send *no money now* to enter your subscription for the collection. However, your application can be accepted only if it is postmarked no later than the final subscription deadline of *February 28th.*

SUBSCRIPTION APPLICATION

The Inventions that Shaped the Modern World

Valid only if postmarked by February 28, 1979.

Limit: One collection per subscriber.

The Franklin Mint
Franklin Center, Pa. 19091

Please enter my subscription for "The Inventions that Shaped the Modern World"—an official collection of ten superbly crafted pewter sculptures to be issued by The Deutsches Museum on the occasion of its 75th Anniversary.

The sculptures will be issued at the rate of one every three months, and the issue price of $90.* each will be billed to me in three equal monthly installments of $30.* I need send no money now, and will be billed for the initial installment prior to shipment of my first sculpture.

**Plus my state sales tax*

Signature ______________________

Name ______________________
PLEASE PRINT CLEARLY

Address ______________________

City ______________________

State __________ Zip __________

ALL APPLICATIONS ARE SUBJECT TO ACCEPTANCE. 3


On the way up, the work may not get easier,
but the rewards get better.
Johnnie Walker
Black Label Scotch
YEARS 12 OLD
12 YEAR OLD BLENDED SCOTCH WHISKY, 86.8 PROOF. BOTTLED IN SCOTLAND. IMPORTED BY SOMERSET IMPORTERS, LTD., N.Y.

LETTERS

Sirs:

It was with great interest that I read Jearl Walker's article on whispering galleries ["The Amateur Scientist," SCIENTIFIC AMERICAN, October, 1978].

I thought you might be interested to know about the whispering wall of the Temple of Heaven in Peking. The circular wall surrounding the temple, known as the Huang Ch'iung-yü, is a whispering wall of the type of St. Paul's Cathedral. There is, of course, no dome here, so that the Chinese can never have misattributed the whispering effect to such a structure. What is most interesting, however, is that the same building has yet another acoustic effect. Three stones of the processional path leading up to the Huang Ch'iung-yü are known as the Three Sound Stones. If you stand on the first stone and clap your hands, the sound returns like an ordinary echo. If you do the same on the second stone, the echo returns twice, and on the third stone it returns three times.

I hope your readers will find this information of interest. The Temple of Heaven is well known, and descriptions of it can be found in many guidebooks.

JAN FONTEIN

Director
Museum of Fine Arts
Boston

Sirs:

In Harold M. Edwards' article "Fermat's Last Theorem" [SCIENTIFIC AMERICAN, October, 1978] he neglected to mention the following possibility for the resolution of the problem. Perhaps there is no proof of either Fermat's last theorem or its negation (in one of the standard axiomatic frameworks for mathematics, say the Peano axioms). The existence of a statement that is neither provable nor disprovable was proved by Kurt Gödel in 1930. Gödel's statement is somewhat contrived, but recently logicians have been able to show that even simple and mathematically interesting statements may be undecidable. In 1970 the Russian mathematician Yuri Matyasevich showed that there is a Diophantine equation (Fermat's last theorem is itself equivalent to a single Diophantine equation) that has no solutions in whole numbers but is such that this fact cannot be proved from the Peano axioms. More recently Jeffrey B. Paris and Leo Harrington have found a very simply stated combinatorial assertion about whole numbers that cannot be decided on the basis of the Peano axioms.

Thus at this point it is entirely conceivable that neither Fermat's last theorem nor its negation has a proof in any of the standard axiom systems. A proof of this fact, however, would not leave matters as undecided as it might seem. If one could "prove" that Fermat's last theorem has no proof, it would follow by well-known theorems of logic that it is true in the standard model of arithmetic, i.e., the set consisting of the numbers 1, 2, 3, 4, ...

STANLEY WAGON

Department of Mathematics
Clark Science Center
Smith College
Northampton, Mass.

Sirs:

As Dr. Wagon correctly observes, in the case of ordinary integers the only way to show that Fermat's last theorem cannot be disproved is to show that it is true. How, then, could this demonstration be a "possibility for the resolution of the problem"?

HAROLD M. EDWARDS

Courant Institute of Mathematical Sciences
New York University
New York

Scientific American, February, 1979; Vol. 240, No. 2. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, vice-president and treasurer; C. John Kirby, vice-president; Arlene Wright, assistant treasurer.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Manuscripts are submitted at the author's risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 40 cents.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. For change of address, notify us at least four weeks in advance. Send both old and new addresses and enclose an address imprint from a recent issue. (Date of last issue on your subscription is shown at upper right-hand corner of each month's mailing label.)

Name ______________________

New Address ______________________

____

____

Old Address ______________________

____

____

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

FEBRUARY, 1929: "Where is the solar system located within our galaxy? The first real idea of how far away from the center we are came when Harlow Shapley applied a new method of determining stellar distances: the photometric method. The story has often been rehearsed of how the true brightness of the Cepheid variables was found to depend on their period of pulsation. In addition the globular star clusters form a gigantic assemblage with its center about 50,000 light-years from the sun in the direction of the constellation Sagittarius. One region in particular, about eight degrees square, on the borders of Scorpio and Ophiuchus, has been found to contain no fewer than 450 variable stars. Twenty per cent of them lie at distances ranging from 21,000 to 30,000 light-years. The rest are fainter and range from 35,000 to 55,000 light-years, with a sharp concentration at about the middle of this interval. These stars lie within a few degrees of the direction of the system of globular clusters and at substantially the same distance. It appears, then, that there is actually an enormous mass of stars concentrated near this point that forms the central nucleus of our galaxy."

"It has been shown by Rutherford that the atomic nucleus deflects an alpha particle as if the force between them were one of repulsion between two electric charges. Thus by counting the number of collisions occurring when a group of alpha particles passes through a known number of atoms we can determine the charge on the nucleus. Measurements of this kind have shown that the nucleus of the hydrogen atom has a positive charge equal to that of one electron, helium that of two electrons, lithium three and so on down the list of chemical elements to uranium. This suggests that the nucleus of the atom may be built up of units carrying a positive charge equal to the negative charge of the electron. Such a unit we find in the nucleus of the hydrogen atom. It is perhaps surprising that the positive unit of electric charge should be associated with a mass almost 2,000 times greater than that associated with the negative unit. Rutherford has, however, performed a series of experiments that gives us good reason to believe our guess is correct. These experiments consist in shooting alpha rays from radium through various substances. It is found that particles having the same charge and mass as the hydrogen nucleus can be knocked out of some of the lighter elements. An event of this kind is shown in a remarkable photograph taken by Mr. Blackett. The evidence seems very strong that the nuclei of the various atoms are indeed built up of an aggregate of hydrogen nuclei, which are now called protons."

SCIENTIFIC AMERICAN

FEBRUARY, 1879: "If Mr. Edison wishes public faith in that electric light of his to remain steadfast, he will have to give an early demonstration of the truth of his claim that it is a practical success. When he first announced that he had solved the problem of dividing the light and of adapting it to domestic uses, there was a very general inclination to accept the story with absolute confidence, because Mr. Edison had proved by his previous inventions that he could achieve some things that had been regarded by other men as impossible. But, after all, the proof of the pudding is in the eating, and the world, after waiting patiently for the public display of an invention that sent gas stocks down as soon as it was heralded, will be disposed, unless Mr. Edison shows his hand, to suspect that the Edison electric light and the Keely motor will have to be ranked together as enterprises containing much more of promise than of performance."

"Professor Loomis of Washington appears to be still enthusiastically carrying on his experiments in aerial telegraphy in West Virginia. Aerial telegraphy is based on the theory that at certain elevations there is a natural electric current, by taking advantage of which wires may be wholly dispensed with. It is said that Professor Loomis has telegraphed as far as 11 miles by means of kites flown with copper wire. When the kites reached the same altitude or got into the same current, communication by means of an instrument similar to the Morse instrument was easy and perfect but ceased as soon as one of the kites was lowered. Professor Loomis has built towers on two hills about 20 miles apart, and from the tops of them has run up steel rod into the region of the electric current. He announces that he has recently discovered that the telephone can be used for this method of communication as well as telegraphic instruments, and that of late he has done all his talking with his assistant, 20 miles away, by telephone, the connection being aerial only. He claims he can telegraph across the sea without wires other than those necessary to reach the elevation of the current."

"The Bell Telephone Company is turning out 1,500 telephones a month, and orders are so numerous that many are more or less delayed. There are now 17,500 instruments out and 15,000 actually rented. Instruments are supplied principally to telephonic exchanges, which are being rapidly introduced into all the larger cities. In Albany and Troy there are 350 instruments in circuit, in Buffalo there are 250 subscribers, in Detroit about 150 instruments in circuit, in Chicago 550, in Indianapolis 150, in St. Louis 325, in Cincinnati 200, in Philadelphia 500 subscribers and 250 instruments in circuit, in Columbus 200 subscribers and about 50 instruments in circuit, in Baltimore 300 subscribers and 100 instruments in circuit, and in Washington, New Orleans, Louisville and Nashville exchanges are being started. In Boston there are 500 subscribers and about 150 instruments in circuit, in Lowell 200 instruments in circuit, in New Haven 350, in Bridgeport 175, and in Springfield, Hartford and Providence exchanges are being started. An exchange has just been started in New York, where there are about 750 subscribers and 250 instruments in circuit."

"A new style of road vehicle, designed to be propelled by mechanical power, has made its appearance in London. The carriage closely resembles an ordinary dog cart; the shafts are very short and inclined together, meeting two feet in front of the dashboard; between them there is a third wheel, working upon an upright shaft, which could be turned by a handle placed the same as that of a bicycle. This handle is worked by reins in the hands of the driver. The motive power is obtained by the combustion of benzoline, a small jet of which is admitted into a burner. The burner, of the size of an ordinary chimney-pot hat, is lined by coils of a copper tube containing water. The steam generated in the tube passes at one end into the cylinders of a small torpedo engine, which rotates a horizontal shaft. It then passes into a cooler, where it is condensed by a current of cold air driven against the outside of the vessel by a revolving fan, and the water so produced is forced back into the other end of the tubular boiler by a pump. The engine shaft works the driving shaft not directly but by the medium of two cones placed side by side, their bases being reversed in position. A figure-of-eight band connects the two, and as it is moved toward the base of the one it nears the apex of the other and thus increases or diminishes the speed of the driving shaft, which is connected with the driving wheel, or off wheel, by an endless band."

RCA announces SelectaVision 400.

The video cassette recorder that turns on and off and changes channels for a whole week ...all by itself.

Think of the four shows you want to put on video tape this week. The game on Monday, the special on Wednesday, perhaps the Friday movie, or something educational for the kids.

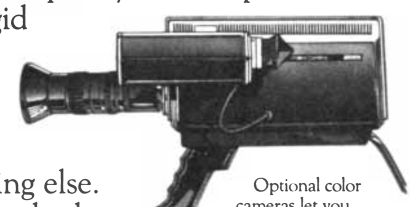
Now, simply by touching a few buttons, you program your selections into the timer of the incredible new SelectaVision 400. The rest is automatic: The 400 will turn itself on at kickoff time, silently record the game, then turn itself off. When it's time for your second selection, the 400 turns itself to the proper channel and starts recording again—automatically. The entire schedule is preset by you up to a whole week in advance—as many as four different shows or even the same program for seven straight days. And you've got up to four hours before changing cassettes.

The 400's programmable timer turns the recorder on and off and changes channels—automatically. Set it up to seven days in advance!

The 400 has more going for it. Like the maximum time available on a single cassette—up to 4 hours. Plus new electronic channel selection. Remote pause control. Direct-drive motor. Special circuitry that automatically compensates for changing signal strength. And quality video tape made to our own rigid specifications.

It's all there in the new 400.

So's something else. Our new optional color cameras. With a Canon 6:1 zoom lens. An electronic viewfinder. And a price that puts home color productions well within your budget.



Optional color cameras let you create your own home SelectaVision productions. Model CC002 features a Canon 6:1 zoom lens.

The new RCA SelectaVision 400. The 4-hour video cassette recorder with the 7-day memory.

Let RCA turn your television into

SelectaVision.™

RCA

CAUTION: The unauthorized recording of television programs and other materials may infringe the rights of others.

THE AUTHORS

LES ASPIN ("The Verification of the SALT II Agreement") is a member of the U.S. House of Representatives. He was born in Milwaukee, Wis., where he attended grade school and high school. In 1960 he was graduated from Yale College summa cum laude and was awarded a Rhodes scholarship to the University of Oxford. He obtained his master's degree at Oxford in 1962 with a combined major in economics, politics and philosophy. He then did doctoral work in economics at the Massachusetts Institute of Technology, earning his Ph.D. in 1965. Commissioned in the army under the Reserve Officers' Training Corps in 1966, Aspin was made economic adviser in the office of Secretary of Defense Robert S. McNamara. He left the army with the rank of captain in 1968 and worked for the next two years as assistant professor of economics at Marquette University. In 1970 he was elected to Congress from the First Congressional District in southeastern Wisconsin. Aspin currently serves on the House Armed Services committee and the Government Operations committee; he is chairman of the Oversight subcommittee of the House Select Committee on Intelligence.

HERMANN RAHN, AMOS AR and CHARLES V. PAGANELLI ("How Bird Eggs Breathe") are respiratory physiologists. Rahn and Paganelli are professors of physiology at the State University of New York at Buffalo School of Medicine, and Ar is senior lecturer in the department of zoology at Tel-Aviv University in Israel. They write: "Our interest in bird eggs started 10 years ago when O. D. Wangensteen, now associate professor at the University of Minnesota, joined our respiration group as a postdoctoral fellow. One day he asked, 'How do eggs breathe?' Since none of us knew, he proceeded to demonstrate for the first time how Fick's law of diffusion would explain the gas exchange across the eggshell of the chicken. We have since studied respiratory processes in the eggs of many bird species, both in the laboratory and on field expeditions to various parts of the world."

AIMÉ BOCQUET ("Lake-Bottom Archaeology") is lecturer in prehistory at the University of Grenoble and director and founder of the Center for the Documentation of Alpine Prehistory, a 20-year-old private association dedicated to the diffusion of regional prehistorical information among scientists, educators and the general public. An oral surgeon by profession, Bocquet organized and has directed since 1972 the experimental workshop and the lake-bottom excavations at Lake Paladru in the French province of Dauphiné, not far from the Swiss border.

STEPHEN H. SPURR ("Silviculture") is professor of botany and public affairs at the University of Texas at Austin. He did his undergraduate work in botany at the University of Florida and his graduate work at the Yale University School of Forestry and Environmental Studies, which awarded him his Master of Forestry degree in 1940. He then began a research career in forest ecology at the Harvard Forest, serving for a time as its acting director. After obtaining his Ph.D. in forest ecology from Yale in 1950, he joined the faculty of the School of Forestry of the University of Minnesota. He moved in 1952 to the University of Michigan, where over the next 19 years he served in several key administrative positions, including vice-president and dean of the graduate school. In 1971 he was appointed president of the University of Texas, a position he held until 1974. His forestry research has involved work with aerial photography, the measurement of vegetation and a wide range of studies in silviculture and forest ecology. Spurr also has a strong interest in the social aspects of resource utilization.

CLAUDIO REBBI ("Solitons") is a theoretical physicist on the staff of the Brookhaven National Laboratory. Born in Trieste, he studied physics at the University of Turin, from which he received his doctorate in 1967. After a postdoctoral fellowship at the California Institute of Technology he taught for two years at the University of Trieste. He then spent two years as a research associate at the European Organization for Nuclear Research (CERN) in Geneva, returning to the U.S. in 1974 as visiting associate professor of physics at the Massachusetts Institute of Technology. He joined the Brookhaven staff in 1977.

CHARLES R. CARRIGAN and DAVID GUBBINS ("The Source of the Earth's Magnetic Field") are geophysicists at the University of Cambridge. Carrigan is currently a NATO Postdoctoral Fellow in the department of geodesy and geophysics. A native of southern California, he did his undergraduate work in astronomy and physics at the University of California at Los Angeles and went on to do graduate work in geophysics there under Friedrich H. Busse. After receiving his Ph.D. from U.C.L.A. in 1977 for simulation experiments on the earth's core, Carrigan went to Cambridge on a fellowship to work with Dan P. McKenzie on "experimental problems related to thermal convection in the earth's upper mantle." Gubbins is professor of geodesy and geophysics at Cambridge, where he obtained his Ph.D. in 1972 in the laboratory of Sir Edward Bullard. He then worked at the University of Colorado, the Massachusetts Institute of Technology and finally U.C.L.A., where he met Carrigan, who was then a graduate student. Gubbins returned to Cambridge in 1976 and joined the faculty the following year.

JAMES F. CROW ("Genes That Violate Mendel's Rules") is John Bascom Professor of genetics at the University of Wisconsin. He was born in Pennsylvania, grew up in Kansas and was graduated from Friends University in Wichita. "In college my first intention was to be a musician," he writes, "but I soon discovered I lacked sufficient talent and switched to science. I chose genetics instead of biochemistry for graduate work because I could get a fellowship that way." After obtaining his Ph.D. in genetics from the University of Texas in 1941, he joined the faculty of Dartmouth College. Seven years later he moved to Wisconsin. Crow's research has been partly theoretical and partly experimental, and it has involved problems in population genetics and evolutionary theory. He is scientific adviser to the Japanese-American joint study of the atomic-bomb survivors and their children, and he is chairman of the Risk Panel of the National Academy of Sciences Committee on Nuclear and Alternative Energy Systems. In his spare time Crow plays the viola in the Madison Symphony and in amateur chamber-music groups.

STANLEY JOEL REISER ("The Medical Influence of the Stethoscope") is director of the history of medicine program at the Harvard Medical School and clinical associate in medicine at the Massachusetts General Hospital. A graduate of Columbia University, he received his M.D. degree at the Downstate College of Medicine of the State University of New York in 1963. After a medical internship at Long Island College Hospital he went to Massachusetts General as a clinical fellow and simultaneously pursued studies of science and public policy at Harvard University's John F. Kennedy School of Government. He obtained his master's degree in public administration at Harvard in 1966 and worked for the next two years as a research associate in biomedical sciences in the university's Program on Technology and Society. In 1969 he began teaching the history of medicine at the Harvard Medical School and the following year he received his Ph.D. Reiser helped to found the history of medicine program at Harvard and is also codirector of the Interfaculty Program in Medical Ethics.

MATHEMATICAL GAMES

About rectangling rectangles, parodying Poe and many another pleasing problem

by Martin Gardner

This month, as is my custom on occasion, I offer a set of unrelated elementary problems. I shall give the answers to them next month, along with references to relevant sources. Any feedback of special interest from readers will be discussed in a later column. I open with a delightful new combinatorial problem of unknown origin. It was passed on to me by Robert Tappay of Toronto, who believes it comes from the U.S.S.R.

1. The Rotating Table. Imagine a square table that rotates about its center. At each corner is a deep well, and at the bottom of each well is a drinking glass that is either upright or inverted. You cannot see into the wells, but you can reach into them and feel whether a glass is turned up or down.

A move is defined as follows: Spin the table, and when it stops, reach each hand into a different well. You may adjust the orientation of the glasses any way you like, that is, you may leave them as they are or turn one glass or both.

Now, spin the table again and repeat the same procedure for your second move. When the table stops spinning, there is no way to distinguish its corners, and so you have only two choices: you may reach into any diagonal pair of wells or into any adjacent pair. The object is to get all four glasses turned in the same way, either all up or all down. When this task is accomplished, a bell rings.

At the start the glasses in the four wells are turned up or down at random. If they all happen to be turned in the same direction at this point, the bell will ring at once and the task will have been accomplished before any moves were made. Therefore it should be assumed that at the start the glasses are not all turned the same way.

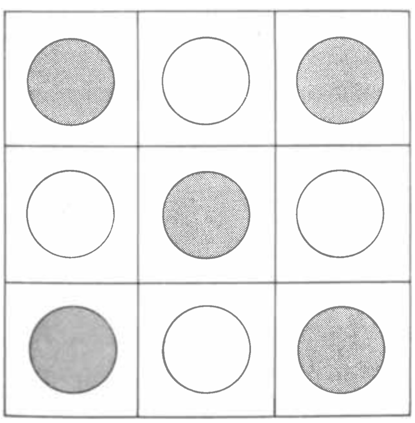
Is there a procedure guaranteed to make the bell ring in a finite number of moves? Many people, after thinking briefly about this problem, conclude that there is no such procedure. It is a question of probability, they reason. With bad luck one might continue to make moves indefinitely. That is not the case, however. After no more than n correct moves one can be certain of ringing the bell. What is the minimum value of n, and what procedure is sure to make the bell ring in n or fewer moves?

Consider a table with only two corners and hence only two wells. In this case one move obviously suffices to make the bell ring. If there are three wells (at the corners of a triangular table), the following two moves suffice.

(1) Reach into any pair of wells. If both glasses are turned the same way, invert both of them, and the bell will ring. If they are turned in different directions, invert the glass that is facing down. If the bell does not ring:

(2) Spin the table and reach into any pair of wells. If both glasses are turned up, invert both, and the bell will ring. If they are turned in different directions, invert the glass turned down, and the bell will ring.

Although the problem can be solved in a finite number of moves when there are four wells and four glasses, it turns out that if there are five or more glasses (at the corners of tables with five or more sides), there are no procedures guaranteed to complete the task in n moves. Next month I shall give one solution for the problem with four glasses and discuss some generalizations of the problem developed by Ronald L. Graham and Persi Diaconis.



The game of turnablock

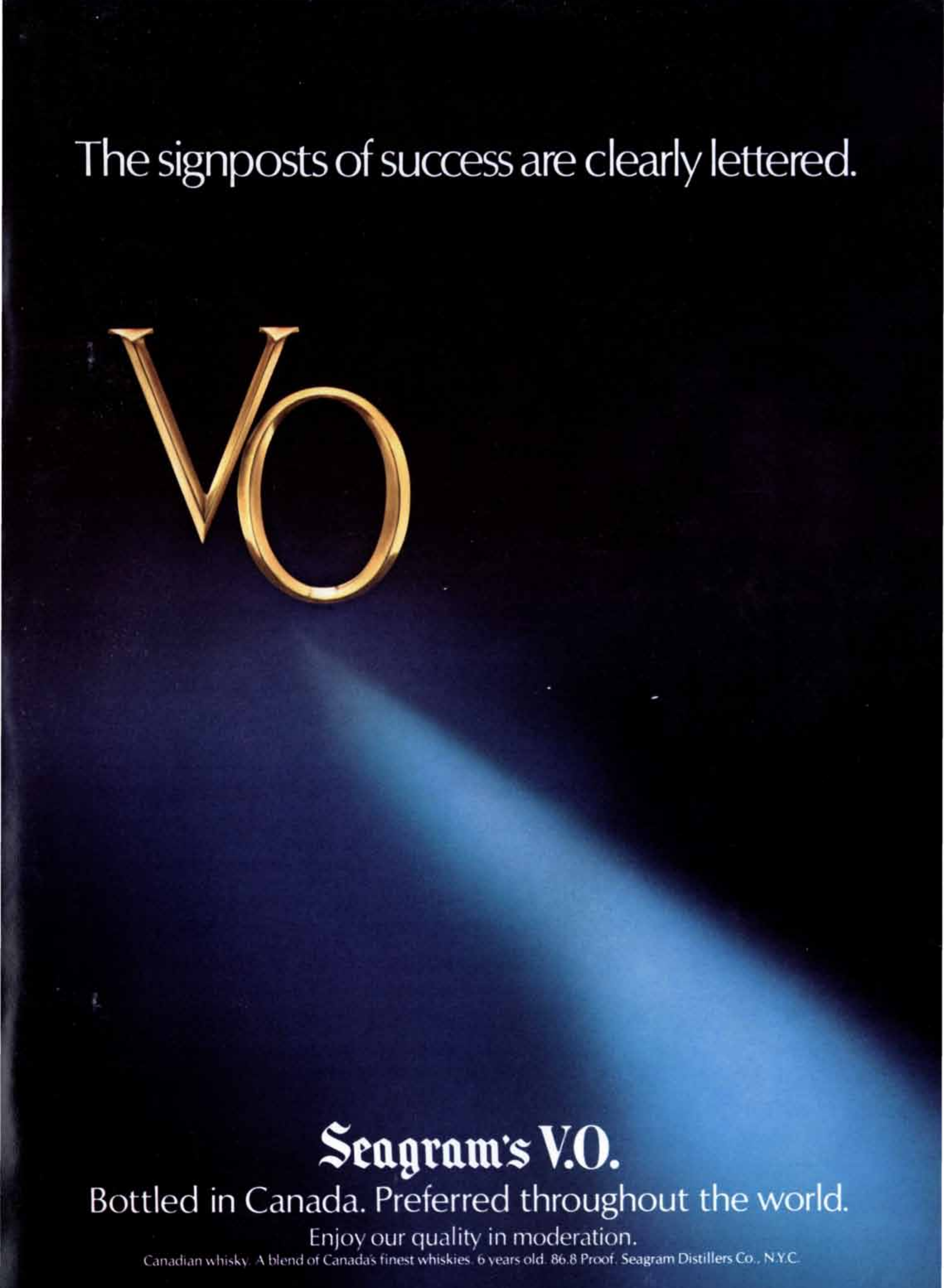
2. Turnablock. John Horton Conway's path-breaking book *On Numbers and Games* (see this department for September, 1976) brought him a flood of correspondence suggesting new games that could be analyzed by his remarkable methods. One such suggestion, from H. W. Lenstra of Amsterdam, led Conway to develop a new family of games, one of the best of which he calls turnablock.

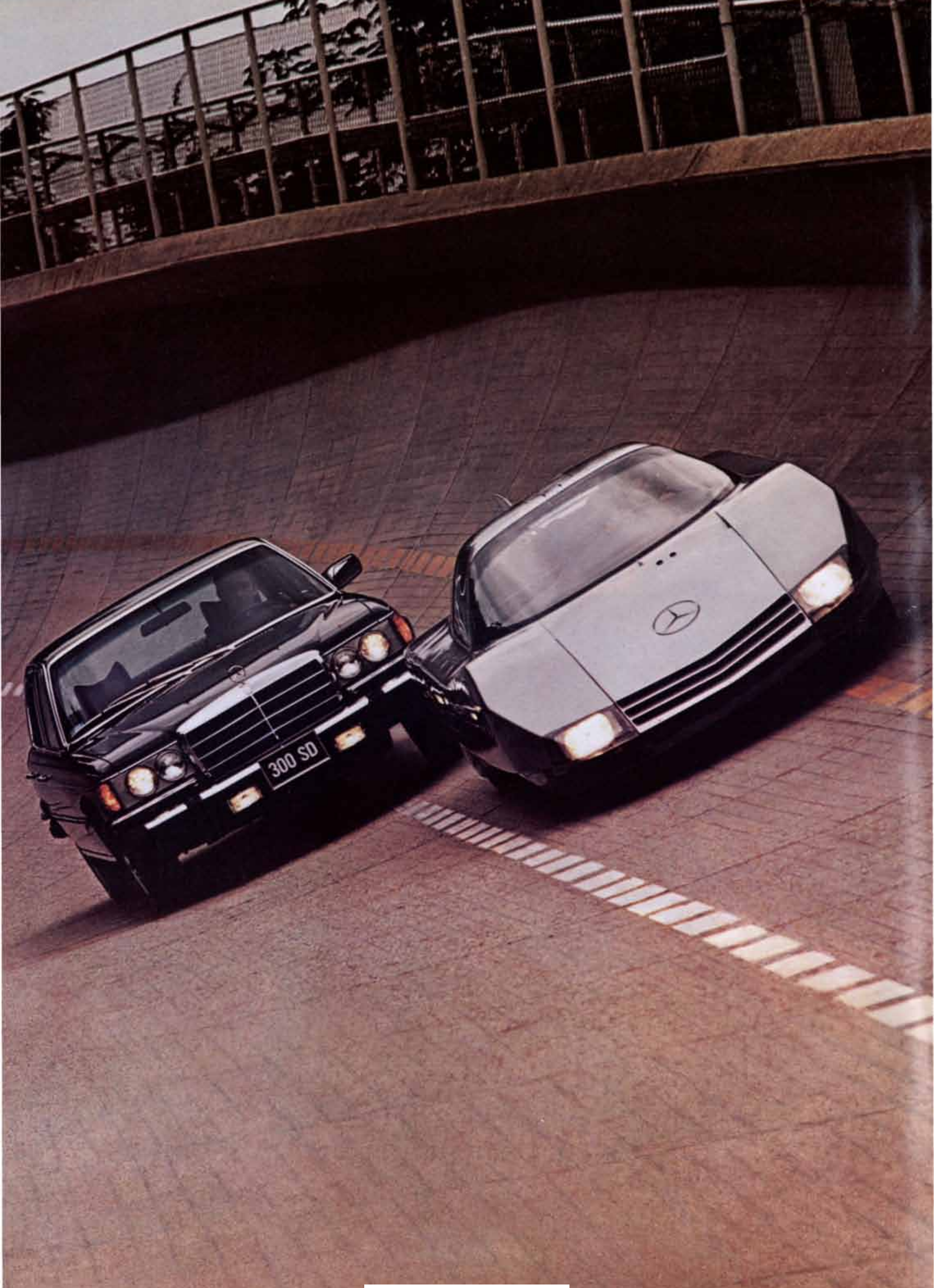
Turnablock is played on any n-by-n checkerboard, using n^2 counters with sides of different colors. The counters included in the popular board game Othello (a new name for the old British game of reversi) can be used for turnablock, or counters can be made by pairing poker chips of different colors. I shall call the two colors black and white. Here I shall describe only the simplest nontrivial version of the game, the one played on the three-by-three board; next month I shall discuss how Conway generalizes it.

At the start of the game the counters may be arranged in any pattern, but for the purposes of this problem assume that the game starts with the alternating-color pattern shown in the illustration on this page. Each player moves by turning over all the counters in any a-by-b rectangular block, where a and b are any two positive integers from 1 through 3. Thus a player's block may be a single counter, a one-by-two "domino" (oriented horizontally or vertically) or any larger configuration up to the entire three-by-three board.

There is one essential rule in turnablock: A block may be reversed only if there is a black counter in its lower right-hand corner. It is assumed that both players are seated on the same side of the board; otherwise the player on one side may turn a block only if there is a black counter in what for him is the upper left-hand corner. The two players take turns making moves; each player must turn a block when it is his move, and the player whose move leaves all the counters with the white side up is the winner. The rule guarantees that eventually all the counters will be turned with their white side up and the game will end. With the starting pattern shown in the illustration the first player can always win if he plays correctly. What are his winning first moves, and what must his playing strategy be?

3. Persistences of Numbers. N. J. A. Sloane of Bell Laboratories, who is the author of the valuable reference work *A Handbook of Integer Sequences,* introduced into number theory the concept of the "persistence" of a number. A number's persistence is the number of steps required to reduce it to a single digit by multiplying all its digits to obtain a second number, then multiplying all the digits of that number to obtain a


The signposts of success are clearly lettered.
VO
Seagram's V.O.
Bottled in Canada. Preferred throughout the world.
Enjoy our quality in moderation.
Canadian whisky. A blend of Canada's finest whiskies. 6 years old. 86.8 Proof. Seagram Distillers Co., N.Y.C.


300 SD

These two Mercedes-Benz cars just rewrote Diesel history–by adding the word "performance."

It was a 1-2 punch that forever shattered the "stodgy Diesel" image. First, that ultra-streamlined 195.4-mph coupe sped to 9 world records in 12 hours on the power of a revolutionary Diesel engine. Then Mercedes-Benz put the same basic engine in a car you can buy: the 300SD Turbodiesel Sedan. At long last, the performance Diesel is here – and from stoplights to filling stations, the benefits must be experienced to be believed.

Do not confuse the Mercedes-Benz 300SD Turbodiesel Sedan with any Diesel automobile you have driven or heard about before.

It is the most technically advanced Diesel that Mercedes-Benz has ever built – but beyond this, it is probably the most *exciting* Diesel automobile anyone has ever built for public sale.

"The most stunning quality of the 300SD, by far, is its acceleration."

– AUTOWEEK

Generating much of that excitement is an engine unlike any ever used in a Diesel car before: a 5-cylinder, 3-liter, *turbocharged* engine.

After testing the Turbodiesel with this power plant, *Road & Track* magazine reported that its 0-60 time of 12.7 seconds eclipsed all other production Diesel cars available today.

Diesel driving, transformed

The effect in everyday driving must be felt to be fully grasped.

You are turbocharged away from stoplights, up long grades, through passing maneuvers, with a feeling of power to spare. In highway traffic, you can maintain the pace with velvet ease.

That newfound sense of power is no illusion: turbocharging this Diesel engine boosts maximum power by 43 percent – and maximum torque by a lusty 46 percent.

195.4-mph Diesel engine

The engine that powers the 300SD Turbodiesel Sedan barely differs in design and major components from the engine that recently made performance history.

The date was April 30, 1978. The place, a circular 7.8-mile track at Nardò, Italy. Driving in relays, dodging fog, fatigue and the odd errant hedgehog, four drivers piloted their fin-tailed C-111/3 research vehicle around and around and around – until, 12 hours later, 9 new world speed records had been set. That C-111/3 had lapped Nardò at an average speed of 195.4 mph for the full 12 hours – including one extraordinary lap at 203 mph.

Diesel efficiency unchanged

Yet the most amazing news about the Turbodiesel engine may not be its performance at all.

Note that while it was averaging 195.4 mph for those 12 hours, that C-111/3 record car was also averaging 14.7 miles per gallon.

In short, you can have your cake and eat it too. Vivid new performance has been injected into the Diesel engine – without subtracting from its legendary fuel efficiency by as much as a single drop.

For example, compare the Turbodiesel with the lively 300D Sedan and its similar but un-turbocharged engine. Though a larger car than the 300D and a good 320 lbs. heavier, the Turbodiesel is not only quicker in acceleration; it uses *even less* fuel. EPA estimates show 29 mpg in highway driving, 24 mpg in the city. (Your mileage of course will depend on the condition and equipment of your car and on where and how you drive.)

"The handling is superb, the suspension always in control. It is a pleasure to drive...and drive ...and drive."

– AUTOWEEK

The 300SD Turbodiesel would be newsworthy even if there were nothing new under its hood.

It is the roomiest Diesel Mercedes-Benz has ever built, the most sumptuously appointed, and the most civilized to ride in.

"On the sound meter," *Road & Track* found, "the 300SD's readings were virtually identical to those of the 450SEL we tested last May."

But despite a plush interior environment of soft velour carpet and hand-finished veneers and thickly padded seats, the Turbodiesel shines in brisk driving; it aims for over-the-road performance seldom found in "luxury" or "prestige" sedans.

Consider some vital statistics. It is less than 18 feet long, needs a mere 2.7 turns of the steering wheel lock-to-lock, and whips around within a turning circle of 38 feet. It is almost playfully maneuverable.

And it needn't be babied along on back roads. The fully independent suspension helps assure that the tires are in constant contact with the road. The car is not knocked off course by every bump and pothole.

The engineers even built a shock absorber into the Turbodiesel's steering system, to help muffle the jolts before they can reach your hands on the wheel. Engineering luxuries abound: gas-pressurized shock absorbers, light-alloy wheels, and four-wheel disc brakes.

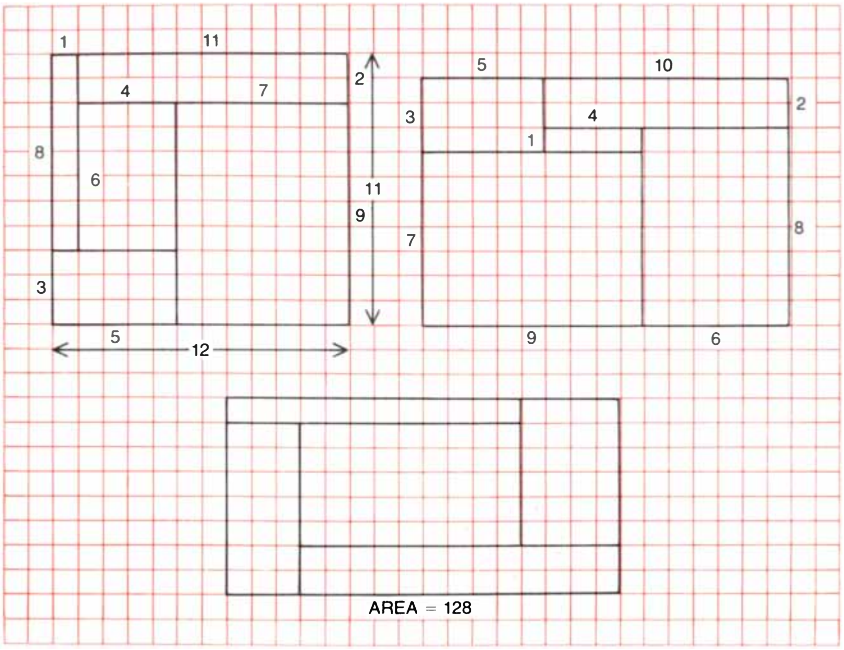
Power-assisted steering and brakes, and a precision-engineered automatic transmission with *four* speeds, help make this performance car almost physically effortless to drive.

Can you resist a test drive?

With the 300SD Turbodiesel, AUTOWEEK concludes, "The Diesel passenger car has come of age."

It seems only logical that the company that pioneered the world's first production Diesel passenger car in 1936 should, two million cars and 50 years of experience later, pioneer the performance Diesel sedan.

The record-breaking Mercedes-Benz C-111/3 (right) and the 300SD Sedan, powered by the same basic turbocharged Diesel engine.



Rectangled rectangles

third number, and so on until a one-digit number is obtained. For example, 77 has a persistence of four because it requires four steps to reduce it to one digit: 77 → 49 → 36 → 18 → 8. The smallest number of persistence one is 10, the smallest of persistence two is 25, the smallest of persistence three is 39 and the smallest of persistence four is 77. What is the smallest number of persistence five?

Sloane determined by computer that no number less than 10^{50} has a persistence greater than 11. He conjectures that there is a number *c* such that no number has a persistence greater than *c*. Little is known about persistences in base notations other than 10. In base two the maximum persistence is obviously one. In base three the second term of the persistence sequence for any number is either zero or a power of 2. Sloane conjectures that in base three all powers of 2 greater than 2^{15} include a zero. Calculations show that Sloane's conjecture is true up to 2^{500}, but there is no formal proof of it. Of course, any number with zero as one of its digits is reduced to zero on the next step. Hence if the conjecture is true, it follows that the maximum persistence in base three is three, as is illustrated by the following sequence: 222,222,222,222,222 → 2^{15} (which equals 1,122,221,122 in base three) → 1,012 → 0. Sloane also conjectures that there is a number *c* for any base notation *b* such that no number in base *b* has a persistence greater than *c*.

Let us call Sloane's persistence a multiplicative persistence to distinguish it from additive persistence, a term introduced by Harvey J. Hindin after he had learned of Sloane's work. The additive persistence of a number is the number of steps required to reduce it to one digit by successive additions. Recreational mathematicians and accountants know this process as "casting out nines" or obtaining a number's "digital root," procedures that are equivalent to reducing the number modulo nine. For example, 123,456,789 has an additive persistence of two: 123,456,789 → 45 → 9.

Unlike multiplicative persistence, additive persistence is relatively trivial and almost everything about it is known. For example, in base two the smallest number of additive persistence four is 1,111,111. In base three the number is 122,222,222. What is the smallest number of additive persistence four in base 10?

4. Nevermore. Three remarkable parodies of Edgar Allan Poe's "The Raven" appeared in recent issues of *Word Ways*, a fascinating journal of recreational linguistics edited and published by A. Ross Eckler (Spring Valley Road, Morristown, N.J. 07960). Each parody is based on a specific form of wordplay familiar to readers of *Word Ways*. All three poems parody the entire Poe poem, but I shall quote only the first stanzas of each.

(1)

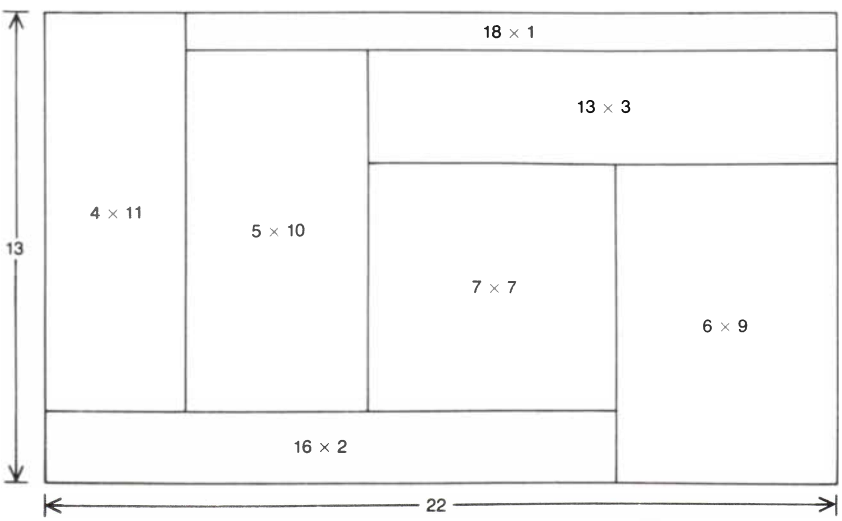
Midnight intombed December's
  naked icebound gulf.
Haggard, tired, I nodded, toiling
  over my books.
Eldritch daguerreotyped dank
  editions cluttered even my bed;
Exhaustion reigned.
Suddenly, now, a knocking, echoing
  door I cognized:
"Eminent Boreas, open up no door!
Go, uninvited lonely frigid haunt!
Avaunt, grim guest—and roar!"

(2)

On one midnight, cold and dreary,
  while I, fainting, weak and weary,
Pondered many a quaint and ancient
  volume of forgotten lore,
While I studied, nearly napping,
  suddenly there came a tapping,
Noise of some one gently rapping,
  rapping at the chamber door.
"Oh, some visitor," I whispered,
  "tapping at the chamber door,
Only one, and nothing more."

(3)

On a midnight, cool and foggy,
  as I pondered, light and groggy,
Ancient books and musty ledgers,
  not remembered any more,
As I nodded, all but napping,
  there I sensed a muffled tapping,

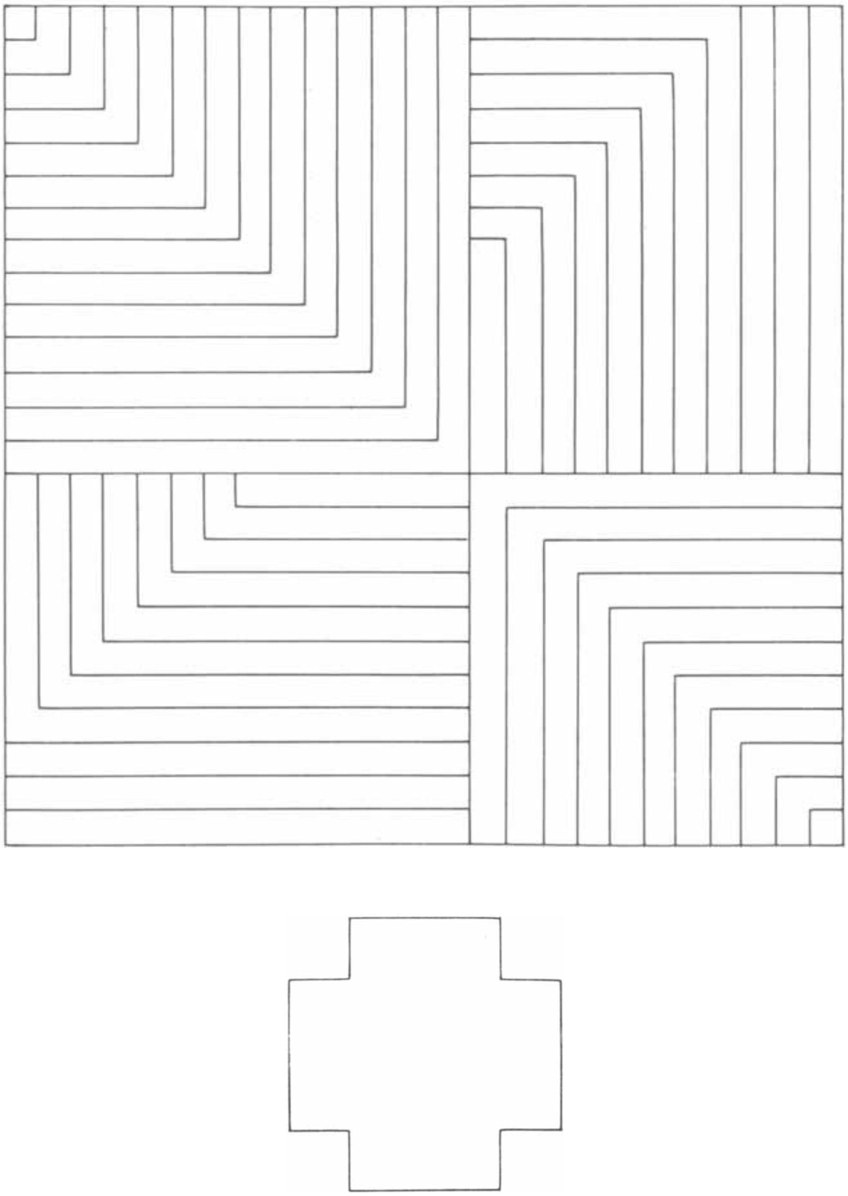


The smallest rectangle that can be tiled with incomparable rectangles

Very much a hushful rapping,
  just behind my attic door.
"'Tis a guest, mayhap," I muttered,
  "knocking at my attic door—
I can't judge it's any more."

The first parody is by Howard W. Bergerson, author of the Dover paperback *Palindromes and Anagrams.* He set himself such a difficult task that it was impossible to retain Poe's original meter and rhyme scheme. The other two parodies are by Eckler. The last was harder to write, but in both cases he was able to preserve the meter and rhyme scheme of the original. What curious linguistic structures underlie the parodies?

5. Rectangling the Rectangle. There is a classic dissection problem known as squaring the square: Can a square be cut into a finite number of smaller squares no two of which are the same size? The solution to this problem was discussed in an article written for this department by graph theorist William T. Tutte (November, 1958) and reprinted in my *2nd Scientific American Book of Mathematical Puzzles & Diversions.* At that time the best solution known ("best" meaning with a minimum number of different squares) required 24 squares. In 1978 this figure was lowered to 21 squares by A. J. W. Duijvestijn, a Dutch mathematician, as was reported in *Scientific American* (June, 1978, pages 86–88). It had been known that 21 was the smallest number possible, and Duijvestijn was also able to show that his pattern for that number of squares is unique. More recently he has found two squared squares containing 22 squares and the best solution, also with 22 squares, for squaring the domino (a rectangle with one side twice the length of the other).



Trace the shape of the smaller figure in the larger one

A somewhat analogous problem is to divide a nonsquare rectangle into the minimum number of smaller rectangles in such a way that no two sides of two different rectangles have the same length. It is not hard to show that the minimum number of internal rectangles is five. Now add the condition that all sides of all six rectangles (including the outer one) are integers. Scott Kim has proved that no solution is possible in which the integers 1 through 12 are used for the 12 different edge lengths, although one can come close. The figure shown at the top left in the top illustration on page 20 is a solution by Kim that includes 11 twice and omits 10. If the sides of the outside rectangle are ignored, the consecutive integers from 1 through 10 will do the trick, as Graham has proved with the rectangle shown at the top right in the illustration.

The figure shown at the bottom in the illustration is reproduced from *Mathematical Puzzles,* a challenging collection by Stephen Ainley (G. Bell and Sons, Ltd., 1977, page 59). It displays one of two solutions in which all 12 edges are different, although not consecutive, and the area of the outside rectangle is reduced to 128. Ainley states that this is a minimal area but gives no proof.

Now relax the conditions slightly so that only the 10 different edges of the five internal rectangles must be distinct and integral. What solution has an outer rectangle of the smallest area? This problem was first solved by C. R. J. Singleton, who sent it to me in 1972. There is only one solution. Can you find it?

The more difficult but closely related problem of finding the smallest rectangle that can be divided into incomparable rectangles was posed by Edward M. Reingold as Problem E2422 in *The American Mathematical Monthly* (Vol. 80, No. 6, June–July, 1973, page 691); a solution was published in the same journal the following year (Vol. 81, No. 6, June–July, 1974, pages 664–666). Two rectangles are called incomparable if neither one can be placed inside the other and aligned so that corresponding sides are parallel. Reingold, Andrew C. C. Yao and Bill Sands, in "Tiling with Incomparable Rectangles" (*Journal of Recreational Mathematics,* Vol. 8, No. 2, 1975–76, pages 112–119), prove many theorems about this problem.

The minimum number of incomparable rectangles needed to tile a larger rectangle is seven. If all sides are integral, the outside rectangle with both the smallest area and the smallest perimeter is the 22-by-13 rectangle shown in the bottom illustration on page 20. It was found by Sands. A square can be tiled with seven incomparable rectangles having integral sides if and only if its side is 34 or larger, but eight rectangles can tile a square of side 27. This square is the smallest one known that can be tiled with incomparable rectangles, but it has not been proved to be the smallest one possible. For details consult the paper by Reingold, Yao and Sands.

In 1975 the first problem, of rectangling the rectangle, was generalized to three dimensions by Kim. There is an elegant proof that a cube cannot be cut into smaller cubes no two of which are alike. (See my *2nd Scientific American Book of Mathematical Puzzles & Diversions,* page 208.) Can a cube be "boxed" by cutting it into smaller boxes (rectangular parallelepipeds) so that no two boxes share a common edge length? The answer is yes, and Kim was able to show that the minimum number of interior boxes is 23. Later William H. Cutler devised a second proof that 23 is minimal, but neither he nor Kim has published his proof. Cutler found 56 essentially different ways to box the cube. If all the edges of such a cube are integral, the smallest cube that can be boxed is not known.

Another unsolved problem is to determine the noncubical box of smallest volume that can be sliced into 23 (or possibly more) boxes with no edge in common and all edges integral. Cutler found a box that is 147 by 157 by 175 and splits into the following 23 boxes:

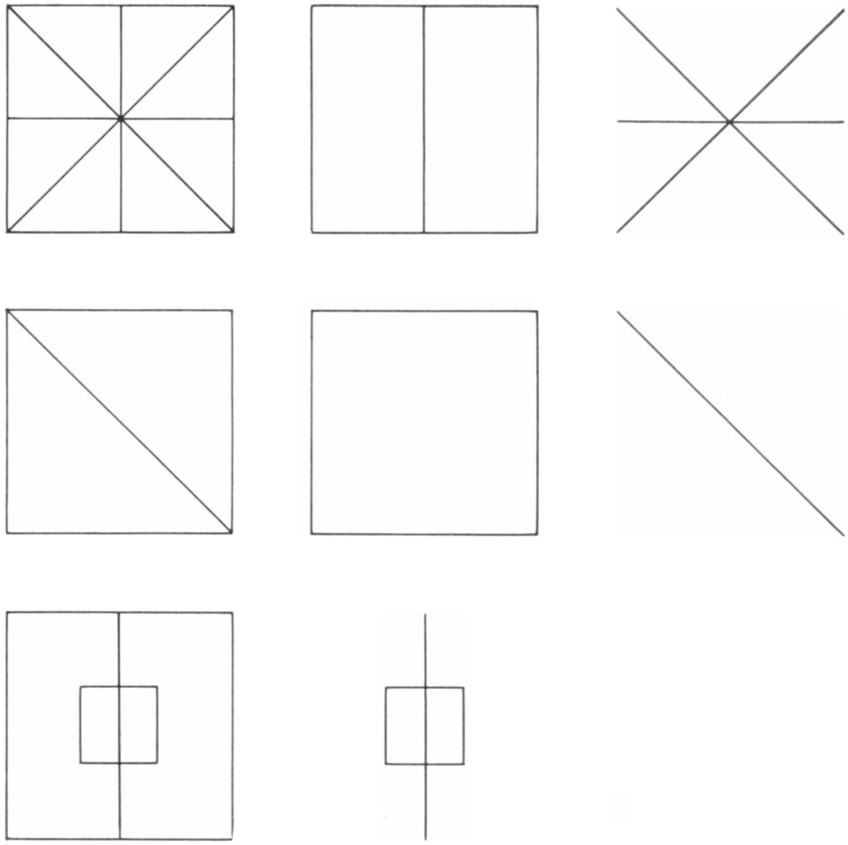
13 by 112 by 141	27 by 36 by 48
18 by 72 by 82	34 by 110 by 135
23 by 41 by 73	57 by 87 by 97
31 by 69 by 78	16 by 74 by 140
38 by 42 by 90	21 by 52 by 65
14 by 70 by 75	28 by 55 by 123
19 by 53 by 86	35 by 62 by 127
26 by 49 by 56	17 by 24 by 67
33 by 46 by 60	22 by 107 by 131
45 by 68 by 85	30 by 54 by 134
15 by 44 by 50	37 by 83 by 121
20 by 40 by 92	

It is not easy to fit these boxes together to make the large box. As far as I know no one has worked on a three-dimensional version of Reingold's incomparable rectangles.

6. Three Geometric Puzzles. (1) The puzzle shown in the illustration on the opposite page is reproduced from last year's September-October issue of the magazine *Games,* a colorful, entertaining new bimonthly now in its third year of publication. (For subscription information write to *Games,* 515 Madison Avenue, New York, N.Y. 10022.) The task is to trace in the larger figure a shape geometrically similar to the smaller one shown below it.

(2) The puzzle shown in the top illustration on this page is from a special issue of the French magazine *Science et Vie* (September, 1978) that was devoted entirely to recreational mathematics. In each row the third pattern is obtained from the first two by applying a rule. What is the rule, and what pattern goes in the blank space in the third row?

(3) The trapezoid shown in the illustration at the right is called a triamond, or an order-three polyiamond, because



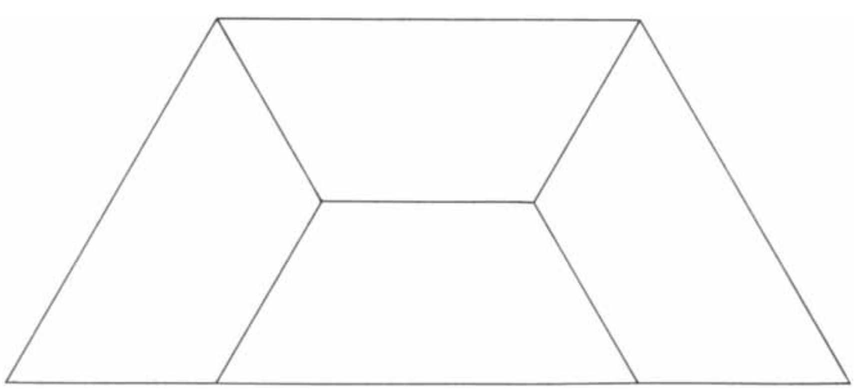
Find the rule

it can be formed by joining three equilateral triangles. In a show last fall at the 55 Mercer Gallery in New York, Denis McCarthy exhibited a striking tessellation made up of 174 of these shapes. They were cut from corrugated cardboard, so that under slanting light their ribbed surfaces would create patterns of light and dark triamonds that would vary with the position of the viewer [*see top illustration on next page*].

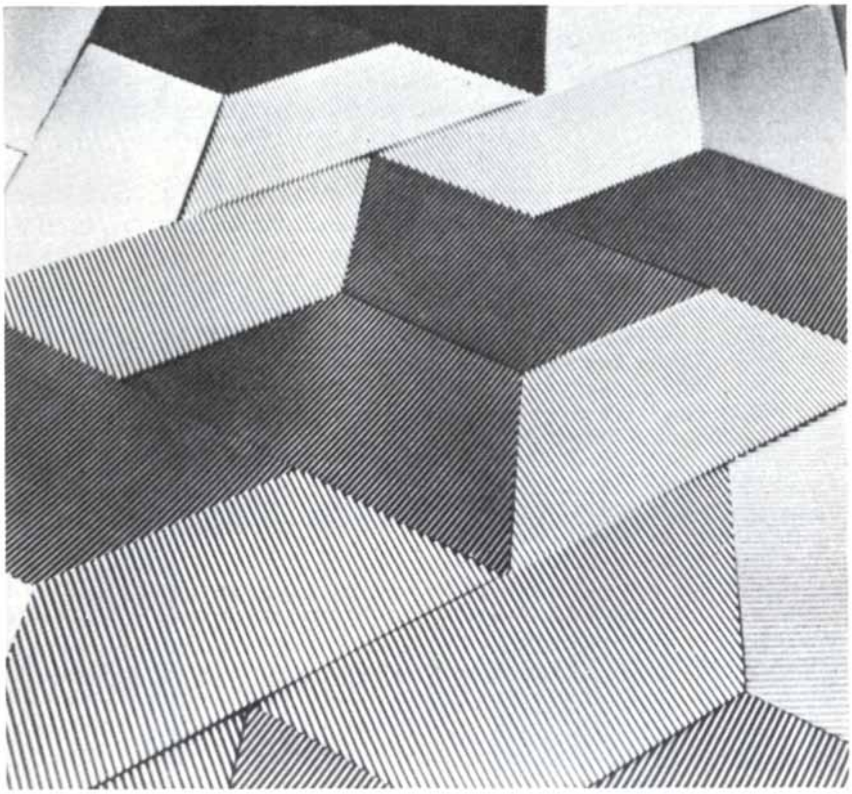
There is an old puzzle that asks how to cut the triamond into four congruent parts. The illustration below gives the traditional solution. Richard Brady, a mathematics teacher in Washington, D.C., tells me that when Andrew Miller, one of his pupils, encountered the problem in Harold R. Jacobs' *Geometry* (W. H. Freeman and Company, 1974, page 188), he found a different solution. In Miller's new solution all four regions do not have the same shape as the larger figure, but they are identical if one or more may be turned over. What is the new solution?

The solution to "The Packer's Secret," an old French puzzle described here last month, is shown in the bottom illustration on the next page. In doing the actual puzzle it is expedient to start with one penny in the center; after 11 more have been placed around it, the center penny can be moved out to the rim.

In my column on twisted prismatic

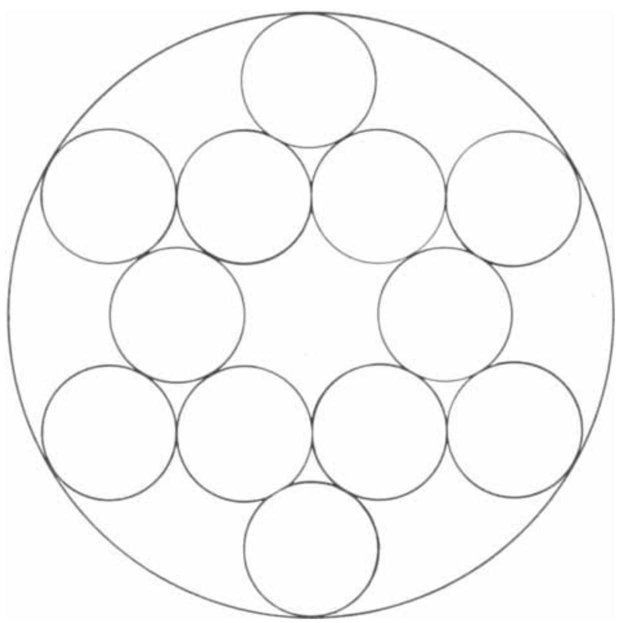


Do it another way



Detail of Denis McCarthy's tessellation of 174 triamonds

rings in August of last year I cited a 1972 article as the only English reference I knew of on the topic. H. T. McAdams sent me a copy of his note "The Sieve of Eratosthenes and the Möbius Strip" (*The American Mathematical Monthly,* Vol. 55, No. 5, May, 1948, pages 308–309), in which the rings are discussed briefly and perhaps for the first time. Shortly thereafter McAdams' table of numbers, which relates the properties of the rings to the ancient sieve for generating primes, appeared on the cover of *The Fantopologist,* a science-fiction "fanzine." It may be from this table that Theodore Sturgeon got the idea of using such a ring in his 1949 science-fiction story "What Dead Men Tell," which I cited at the end of my December column last year.



Solution to "The Packer's Secret" puzzle

Hewlett-Packard computer advances

The HP 300. An innovation that puts high-powered, flexible computing into a small, compact console.

No larger than a free-standing terminal, the HP 300 is a new, low-cost business system that achieves a new level of computing concentration. Once programmed, it is friendly enough for people with no computer experience to use, yet packs a breadth of capabilities that will surpass normal expectations of a sophisticated applications designer.

The HP 300 computer is designed to simplify the development and control of dedicated on-line business applications. It incorporates HP's proprietary silicon-on-sapphire (SOS) integrated circuit technology: Exhaustively tested and refined, SOS semiconductor chips give the HP 300 the related benefits of smaller size, greater performance, and lower cost than computers using traditional technologies. The HP 300 is designed for office environments and plugs into a standard electrical outlet.

The HP 300 can handle up to 16 terminals in transaction processing environments, and its "virtual memory" allows the use of extremely large programs and data sets without being confined to physical memory size. A powerful new operating system, Amigo/300, in conjunction with an integrated display system (IDS), greatly simplifies user interface, and enables the HP 300 to perform both as a multiprogramming and multitasking system. For example, the HP 300 can print reports at the same time that higher-priority data entry or data base management inquiry operations are taking place at the system's terminals.

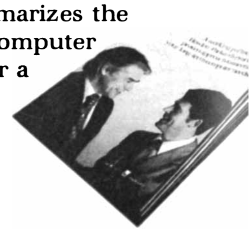
Eight programmable soft keys on the side of the video display screen can be captioned to give the operator a series of choices in performing a job. The display itself can be divided into multiple "windows" to view and control different parts of an application. These windows can even be scrolled individually, both vertically and horizontally, to a width of 160 columns and a length of several thousand lines.

Two widely used business languages, RPG II and Business BASIC, are available, as are IMAGE/300 data base management with an English-like inquiry capability and TYPIST/300 text editing. There's even a HELP key that gives the user direct access to an on-line reference manual for the system. Full control of all job priorities and applications is maintained at the HP 300's central console.

The price of the system is $36,500 with 256k bytes of main memory, 1 megabyte of mass storage on one flexible disc drive, and one 12-megabyte fixed disc. This can be expanded to 1 megabyte of main memory and 240 megabytes of disc storage. In brief, the HP 300 offers the capabilities, flexibility, and growth potential of computers costing two or three times its price.

A working partnership with HP.

This informative management booklet outlines HP's approach to doing business. It summarizes the expertise, resources, support, and computer products we bring to customers. For a free copy write to A.P. Oliverio, Vice President, Marketing, Hewlett-Packard Co., 1505 Page Mill Rd., Palo Alto, CA 94304.



extend your possibilities.

The HP 3000 Series 33 computer system. Another innovation that, through advanced LSI technology, delivers major business system capability in a desk-styled package.

With the Series 33, Hewlett-Packard adds a low-cost, compatible member to its 3000 computer family: powerful systems that adapt to your operational requirements for the management of manufacturing and other business information.

Another dramatic example of HP's advanced silicon-on-sapphire (SOS) technology is the HP 3000 Series 33.

This general-purpose major business system is especially well suited to on-line transaction processing applications such as order processing, inventory control, and material management. It is a relatively low-cost answer for many companies that can afford only one computer, or for larger organizations implementing distributed processing.

Its power can be simultaneously tapped by users at up to 32 terminals, and the Series 33 runs the same multiprogramming executive operating system—MPE-III, as the larger HP 3000 systems. This means the new Series 33 can immediately share in the thousands of man-years of software development for HP 3000 systems. The Series 33 also provides IMAGE/3000, HP's award-winning data base management software, and its easy-to-use QUERY capability. A new data entry software package, VIEW/3000, makes it easy for experts and beginners alike to create business forms on system video terminals and to enter data into files without writing a program.

The Series 33's system console includes a built-in diagnostic microcomputer for on-the-spot diagnostic testing. With a modem, HP service specialists can even communicate with the Series 33 by telephone for long-distance troubleshooting.

A base system includes the computer—now only three SOS chips on two printed circuit boards instead of nine printed circuit boards—256k bytes of error-correcting main memory, a 1 megabyte flexible disc, a 20 megabyte cartridge disc, and the system/maintenance console. The Series 33 can grow with your needs to 1 megabyte of main memory, and can use up to 8 disc drives with 960 megabytes of disc storage. A wide range of peripherals—magnetic tape drives, printers, and terminals—further increases a user's ability to tailor an HP 3000 Series 33 to a specific application and changing it as operations demand. Prices start at $70,000*.

HEWLETT hp PACKARD

1505 Page Mill Road, Palo Alto, California 94304

For assistance call: Washington (301) 948-6370, Chicago (312) 255-9800, Atlanta (404) 955-1500, Los Angeles (213) 970-7500

Mail to: Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304
Please send me further information on

( ) HP 300 computer system
( ) HP 3000 Series 33 computer system

Name ____________________
Company ____________________
Address ____________________
City ____________ State ________ Zip ________

*Domestic U.S. prices only

00942

BOOKS

Beer made naturally, the strange story of the poison dioxin and other matters

by Philip Morrison

BEER NATURALLY, by Michael Hardman, photographs by Theo Bergström. Bergström & Boyle Books Limited, London. Two Continents, 30 East 42nd Street, New York, N.Y. 10017 ($4.50). Why was Burton-on-Trent built on Trent? "Because the gypsum content of the local well water helped to produce a first-rate clear beer." Throughout the world brewers treat their local water to imitate the gypsum content in the well water of the lovely Trent valley. Less prosaically, the reader can follow the large photographs, some in color, from the green spires of the hop field through the golden rows of barley falling before the reaper to the barmaid at the tap and the happy men, the day's work behind them, calling for "a pint of beer, naturally."

This paperbound album of handsome photographs and its brief yet detailed text sum up an understandable and interesting industry that produces the classical malt beverages with a technology halfway between craftsmanship and chemical engineering—the Georgian and early Victorian brewing industry that is still active in Britain. Here are its tools: the wood paddle for skimming the yeast, the wood shovel for turning the germinating grains of barley, the big coppers and the warehouses of cooperage, even the conical drying house for the conical fruit of the hops. Here are its people (the hop picker and the master brewer) and here is its jargon (firkin and oast house, wort and spile, isinglass and stillage).

Water is of course the base; it is 90 percent of any beer. For 1,000 years the Germans and their neighbors have added the bitter hop fruit, not only a flavoring but also a bacteriostatic agent. For a while hopped beer (by way of the Dutch) struggled for English acceptance; what the English then drank unhopped was called ale, and beer was its hopped Continental competitor. Hopped beers have been ascendant ever since the Flemish came to grow hops in Kent in Shakespeare's time. Today the two words, beer and ale, have the same meaning. A natural beer (the ingredients are legally specified in Germany today) is brewed solely from barley, hops, yeast and water. There are two enzymatic stages. The first is malting, in which the living barley grains germinate and the enzyme diastase, among others, converts the starch of the cereal into various sugars that are more accessible to the next microbial stage. That stage is fermentation by yeasts, which yields carbon dioxide and ethyl alcohol, the two staple molecules in the complex, flavorful and aromatic mix that is beer as a whole.

The darkness of some natural beers comes from the browning of the sprouted barleycorns as they dry in the hot air of a kiln. When the barley has germinated (formerly by being turned over and over again with a wood shovel, now more easily in rotating drums), it is sieved, cleaned and ground. Brewing proper is the steeping of the malt grist in hot water (in this trade water is always called liquor) until the mixture, a mash as thick as porridge, has yielded its solubles in just amount. The process may be automated to the hilt, with recording thermometers and flashing lights, or it may be left to the ripe judgment of the mash master. After a couple of hours there pours out of the mash tub the sweet liquid called wort. Once that was sent straight to fermentation. Now it is boiled with the dried hop fruits, again a matter of skill and taste, with many subtle sources of distinction in time, type of hops and so on. Powdered and pelletized hops are often used in the quest for surety.

Next the wort is fermented by yeasts, an exacting piece of microagriculture. Here the action is that of a live microorganism, and the options for culture multiply; national beers reflect yeast types and treatments. Only scrupulous cleanliness makes possible adequate control; infection is always a hazard, and the hops narrow the range of infectants. The time is now measured not in hours but in days or weeks. Then the beer is cleared by settling and by its own bubbling, and finally most of the remaining yeast is brought down by the addition of finings (the viscous isinglass derived from sturgeon swim bladders is traditional). Now is the time of the tasting, a social occasion in the brewery.

The beer is casked and shipped still alive with residual yeast, whose continued growth is controlled by the pub. By manipulating leaks from the cask with pegs driven through a soft-cored bung in the cask top, the production of carbon dioxide is controlled, guarding against the Scylla of flatness and the Charybdis of fizziness. More than a few days at cellar temperatures just below 60 degrees Fahrenheit and the air-containing cask will go "off." The good landlord anticipates the demand; he serves mild and bitter at the needed rate with a beer engine (a hand-driven suction pump) or with a power assist.

Most bottled beer and beer in pressurized kegs is different. It is meant to travel well and keep longer; the yeasts are filtered out, no settling time is needed and the liquid is often pasteurized. It can make no more gas and so it is stored pressurized, with carbon dioxide added. The well-chilled American beers belong to this large class. The reformist authors of *Beer Naturally* combine a touch of nostalgia with their sense of gourmetry: for them the consistently processed chemical-engineering beers, fermented on a large scale by continuous flow and not in batches, pressurized, filtered and pasteurized, "are not the same.... They are missing the taste of the barley field and the hop garden."

Indeed, even though Americans drink three times as much beer as the British, the U.S. barley crop is about the size of the British one. The label on the six-pack explains: American brewers add to their malted barley much extra starch from rice, maize or wheat. (One of the leading national American beers is in fact mainly a rice beer, although with not much taste of rice or anything else!) The disposable container has long favored centralized large-scale brewing; nearby breweries are relatively better off with reusable containers. Quickly perishable natural beers are no longer much seen across the U.S., even when the beer is drawn from a tap and the sign proclaims an ancient European name.

It is illegal even today to make beer in Germany of ingredients other than the classic four. The old law was probably intended to conserve the finer cereals for bread more than to ensure the barley-malt flavor. It may be that today we win our grain all too easily.

DIOXIN: TOXICOLOGICAL AND CHEMICAL ASPECTS, edited by Flaminio Cattabeni, Aldo Cavallaro and Giovanni Galli. SP Medical & Scientific Books, a division of Spectrum Publications, Inc., distributed by the Halsted Press, John Wiley & Sons ($20). The molecule nicknamed dioxin, whose proper title is 2,3,7,8-tetrachlorodibenzo-para-dioxin

(acronym TCDD), is not a very complex substance. It is elegantly symmetrical: two benzene rings joined side by side, linked in their own plane by a double bridge of oxygen atoms and fringed on each outer edge by two chlorines substituted for the outermost hydrogens. The molecule's pattern is neat but its work is deadly. A fatal dose of such powerful agents as heroin or cyanide for a person of normal size has about the bulk of half a tablet of aspirin; dioxin is more toxic by a rough factor of 100, the most poisonous small molecule we know.

It is a synthetic product derived more or less inadvertently from the large-scale manufacture of polychlorinated single benzene rings, an essential step in the production of the widely used herbicides 2,4-D and 2,4,5-T. From 1949 on there have been severe accidents in five countries during industrial preparation of the intermediate, several hundred workers having been affected as a consequence of such "uncontrolled conditions" as fire or explosion. What substance was to blame became known only in 1957, when research laboratories in Germany and the U.S. synthesized dioxin (the American workers actually made the bromine analogue) in conventional 20-gram lots. "In both cases [a] laboratory worker was hospitalized." Remarkably, the German chemist entered, for treatment of his extreme case of acute acne with pustules forming at all the little oily glands around the hair follicles of the skin, the very Hamburg hospital where, a few years before, a number of chemical-factory workers had appeared with the same affliction. The identification was confirmed.

The effect is form-specific: somewhere in life the poison fits. Modify the molecule by replacing four more hydrogens with chlorine and the lethal dose rises (in rats) 50,000-fold. Try only two or three chlorines in the molecule and virulence is similarly reduced; keep the fourfold form only (but all bromine atoms instead of chlorine) and the dose is little changed. The substance has many effects. Acne arises from contact, even in rabbits, and the stuff is fatal to rats in a single oral dose; the animals show little interest in food and waste away in a month. A similar total dose administered chronically at the rate of a few micrograms a day kills laboratory animals quite surely.

The effects vary from species to species. "At postmortem the most striking finding is a loss of body fat. In fact, in many animals it is difficult to find enough remaining fat for analysis." The substance is strongly toxic to the embryo in chickens, rats and mice; it is mutagenic in bacteria, but it has shown no effect on human cells in culture, at least not on certain abnormal cell strains. There is evidence of strong carcinogenicity. Just how the drug acts remains entirely conjecture.

The symposium on dioxin recorded in this book was not, as it might well have been, coolly motivated by the singular quality of the drug in the laboratory. Rather the 20 papers on chemistry and analysis, on toxicology and on field decontamination were delivered by American and European experts to a Milan workshop in late 1976, a few months after the most acute of dioxin spreads. A midnight explosion of a trichlorophenol reactor in a chemical plant near Milan had resulted in the diffusion of about a kilogram of the stuff, as part of a toxic cloud, to an area near the plant covering a few square kilometers. The disaster was half-concealed for days. The square kilometer most affected housed some 200 families, who were evacuated a couple of weeks after the event. There has been put in motion an elaborate five-year plan for decontamination, with vacuum cleaning and washing of walls, floors and ceilings, removal of topsoil and careful monitoring and control.

The key technique is a reliable analysis in the field. To detect micrograms ingested by people who daily take in kilograms of foodstuffs, air and water one needs to analyze small samples for specific molecules at parts per trillion. It can be done, and done reliably. The techniques are elaborated here. They depend on gas chromatography for initial enrichment of the substance. Helium carrier gas flows slowly through a 30-meter glass capillary coil that is kept hot and lined with the right solvent (a silicone is preferred). Into the flow through this selective filter the prepared sample fraction, dissolved in another just-right solvent (TCDD is hardly soluble in water and seeks oils and lipids), is steadily vaporized.

Within a time defined to a couple of seconds out of an overall five-minute delay, the TCDD molecules finally leave the coil. They are led through an ionizing beam, and the known ions of most likely charge and mass (about 320 units) are selected by a focusing magnetic mass spectrometer, accelerated to higher energy and counted in a many-stage secondary-electron-multiplier tube. The entire system, obviously an instrument of the highest technology, is nowadays controlled by computer. The digital signal is manipulated for background reduction and the like and is plotted out as a narrow telltale peak or pair of peaks. The system can meet every demand of sophisticated control, in particular spiking with isotopically loaded tracer samples. This extraordinary sensitivity represents a kind of bioassay by the mind.

TCDD has been spread far and wide, in low concentrations to be sure, over the U.S. and Vietnam. Agent Orange and its stateside analogues, much used defoliants for weeds and forest cover, seem to have had an unwanted and often unknown contamination by TCDD since they were first made. The Air Force delivered from the air about 40,000 tons of Agent Orange (derivatives of 2,4,5-T and 2,4-D) to the forested roadsides of Vietnam. TCDD was often there in concentrations well below one part in 10 million. A trillion is a very big dilution indeed, however, and clinical reports and field samples indicate some TCDD effects in areas of Vietnam. The situation in the U.S., where fodder crops more than forest have been sprayed over a much longer period, is not yet fully known. Samples of mother's milk are now being studied under rigorous controls. The results should bear sharply on this example of an uneasy balance between the visible economic benefits of agricultural chemistry and uncertain risks to public health.

At least photodegradation by solar ultraviolet radiation and decomposition by soil microorganisms do slowly remove the danger. Contact with hydrogen donors (a decontaminating olive-oil spray was being tested in Lombardy) greatly speeds the sunlight effect. It seems that the carrier formulation used in the Agent Orange spray may fortuitously have resulted in swifter disappearance of its unheeded TCDD.

ATLAS OF WORLD POPULATION HISTORY, by Colin McEvedy and Richard Jones. Penguin Books ($4.95). THE TIMES ATLAS OF WORLD HISTORY, edited by Geoffrey Barraclough. Hammond Incorporated ($50). The number of titles of books in history is yearly so large as to equal all those in science and technology. These pages, therefore, ordinarily report only on works bearing quite directly on the history of science and technology. The two reference books listed above, one small and one large, are offered as exceptions to the rule.

Population statistics are such a simple and quantitative index to human affairs that they partake of the analytic simplicity of much physical science. How many people make a hunting band, how many a great empire? The number alone fixes much. In *Atlas of World Population History,* a tour de force of a volume, two British historians present in all modesty, but with decisiveness and courage, a quantitative muster of human heads from the time of the glacial retreat to A.D. 1975. They put their estimates forward not only as a global sum but also graphically as a function of time, from 400 B.C. to the present, continent by continent and country by country. Each graph is accompanied by a page or so in elucidation of the estimates and their sources, and there is a good, but not a number-by-number, bibliography. The authors argue engagingly and well; their

case is strong, although clearly and necessarily a matter of reasonable estimation; cautious authors would never have tried this valuable task. What they have done makes sense, although "consistency . . . provides comfort rather than proof and we wouldn't attempt to disguise the hypothetical nature of our treatment of the earlier periods. But we haven't just pulled figures out of the sky. Well, not often."

They come to real conclusions. World population when the ice retreated was some four million. The area under crops was too small to mean much globally until about 5000 B.C. Then it began to rise, and it rose at an increasing rate through the Iron Age in the Near East, in about 5000 B.C. Growth slowed and then ceased with the simultaneous decline of Han China and Rome, the slaveholding empires. The Dark Ages followed in both of these widely separated lands; then, from the 10th century to the Mongol wars, numbers rose, topping out in A.D. 1200. Modernization took over. First there was a fast but variable rate of rise continuing until about 1700, and from then on there were "no more of these hiccups": a transition spreading from Europe to the other continents, first a high birthrate with a low death rate and then a lowered birthrate. The final birth adjustment has yet to arrive globally; the authors foresee a plateau about a century hence at between eight and nine billion people, well below the 20 billion theoretical limit.

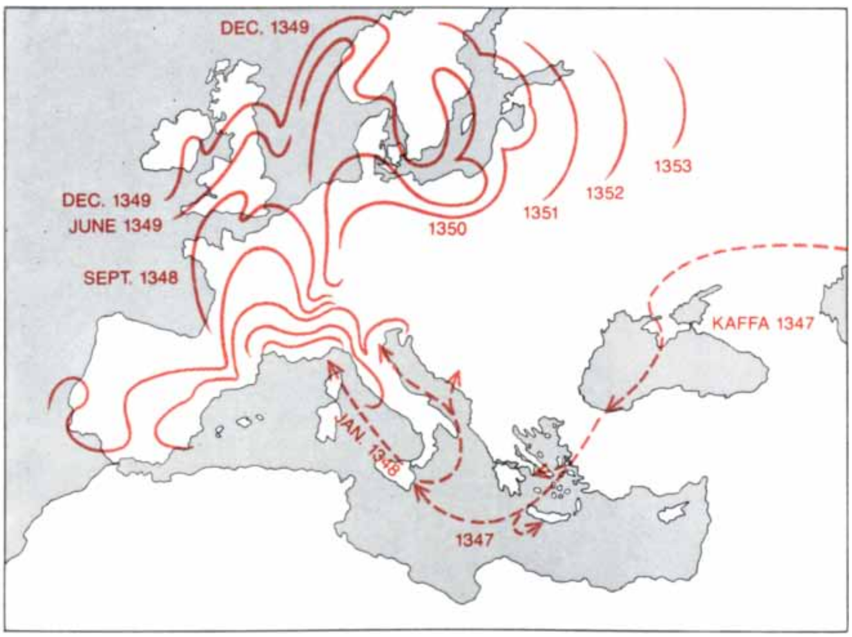
Causes? The authors admit the influence of climate and crops, but they are less than convinced: the "little ice age" shows a fall in population in the south of Europe rather than in the colder north. A wider net must be cast to catch the meaning of the cycles of growth and pause.

The 50 individual regions discussed are full of interest, for reading once over and for reference. They circle the globe from the British Isles to New Zealand. The Irish graphs and text page are of special interest. The population of Ireland doubled in the 17th century and more than doubled in the 18th. But 19th-century Ireland, neither industrialized nor modernized, could not keep up the rate. The impoverished and landless people emigrated to British factories and to the New World. Even the high-yielding potato had left them without work, and almost without food during the winter. A million and a half had left by 1845, when the potato failed and famine came; the stream of emigration rose then in flood, and the excess deaths totaled at least 750,000 in some five years.

The Times Atlas of World History cannot be described as compactly as a count of heads worldwide. It shares somewhat the broad organization: each of the 130 double-page spreads treats, by means of colorful maps and charts interspersed with a page or two of text, a period and a place. First comes early man and his world, continent by continent. Then the sections include the early empires and the classical ones, the Dark Ages that divided the old centers and the rise of new structures in Islam, Africa, Turkey, India and the Americas. The scene narrows as the West expands, at first caught in war and revolution and then as dominant empires until World War I. The last section offers a score of spreads ranging from the world economy of 1870 to the final maps showing the rich and the poor lands of 1975.

The chief novelty of this excellent compendium is its earnest and successful effort to be ecumenical. We see the Warring States become Ch'in and Han as clearly as we watch the rise and fall of Rome. India, Africa and the Americas are well treated also; the Kievan state, Champa, Khmer and Dai Viet are bordered and described. The ancient Nile is mapped in some detail, allowed to flow horizontally across the two-page spread to give more room for place-names. Consider in detail one sample spread describing the emergence of states in Africa from A.D. 900 to 1500. There is a small painting in color from a medieval Catalan atlas showing the golden wealth of Mansa Musa of the empire of Mali through contemporary eyes. A fractional-page color map displays the main caravan routes of the Sahara and a full-page color map of Africa marks the products and the old territories of Oyo, Benin, Songhai, Ghana, Zimbabwe and Axum. In their turn such more conventional subjects as the Hapsburgs and the Peace of Westphalia are not neglected, but no question of it, this stage is wider than all Europe.

The work is new, and so ambitious that perhaps it lacks the magisterial sense of authority one would hope for in such a substantial volume. For example, it gets the date of Sputnik wrong, it is a little confused about the Japanese civil wars and it celebrates Abdera not for Democritus but for Protagoras. All the same it is a fine and useful work, the latest in a long tradition of historical atlases.



SPREAD OF THE BLACK DEATH in Europe from 1347 to 1353 is shown on this map adapted from *Atlas of World Population History,* by Colin McEvedy and Richard Jones. On land the plague spread most readily in densely populated areas. The low-population-density areas of the Balkans and southern Russia acted as a firebreak. Although the plague got to Moscow, it did so by way of the Mediterranean, France, the North Sea and the Baltic, not up the Volga.

COPERNICUS: ON THE REVOLUTIONS OF THE HEAVENLY SPHERES, a new translation from the Latin by A. M. Duncan. Barnes & Noble Books ($26.50). NICHOLAS COPERNICUS ON THE REVOLUTIONS, edited by Jerzy Dobrzycki, translation and commentary by Edward Rosen. The Johns Hopkins University Press ($67.50). The flood of scholarly and national celebrations of the 500th anniversary of the modest and learned canon of Frombork has subsided, but it brought us much of value. In our near-Latinless age the original work, available now in facsimile and even in a facsimile of the autograph manuscript, is beyond the reach of many who might want to follow for

Full employment. That's our goal for the working tree.

America's need for wood is growing.

By the year 2020, government and industry sources expect demand for wood fiber to nearly double.

That's one reason why, at Potlatch, we feel a responsibility to utilize each harvested tree to the fullest.

For example, improved sawmill technology enables us to get more end product from every log we process.

We use that part of the tree crop not suitable for conversion into lumber—the tops of pine and hardwood trees, slabs, edgings and sawdust. We process them into pulp for paper and paperboard products.

We also use the bark and scrap residuals as fuel to supply part of our power needs.

Full employment of our working trees.

It's no small task. But it's the one that pays off in jobs and paychecks today, and homes and forest products tomorrow.

Potlatch Corporation, Public Information Office, P.O. Box 3591, San Francisco, CA 94119.

Potlatch

themselves its pathbreaking arguments. One translation into English did exist; it was prepared just before World War II and is available as one volume in the large set *Encyclopaedia Britannica Great Books of the Western World*. To the impediment in accessibility one should add that this still useful translation, by the late Charles Glenn Wallis, is original to the point of eccentricity; it impresses some readers as being unrecognizably modernized and others as being filled with curious archaisms. In the past year or two we have been offered a wider look at Copernicus in English.

The first of the new examples listed above is a clear, workmanlike and easily available translation by a historian of science at Loughborough University of Technology in England. Duncan's offering is presented succinctly and modestly in a book of ordinary scale, with half a dozen pages of notes for the unspecialized reader that are helpful as far as they go (as in explaining the Latin wordplay Copernicus enjoys on his first page of text). A student or any other reader can confidently follow this text, most profitably with the aid of a teacher or the literature.

The second volume listed is a veritable Rolls of vehicles. Sponsored by the Polish Academy of Sciences and lovingly produced in Poland (with few errors), it is a strikingly large and handsome volume with an inset portrait of the author in color on the binding, with superb gravure endpapers that are real wide-field photographs of the night sky, and with a high price. The content is also de luxe, the work of the most experienced and learned Copernican scholar of them all. Professor Rosen gives us useful indexes (prepared by Erna Hilfstein) and a small treatise in notes on the text, 100 pages and more in close-set type.

Here we see the text through the eyes of a meticulous scholar-translator eager to find the discrepancies, the misstatements, the subtle allusions we might otherwise miss. Copernicus tells us in the preface that his manuscript had been buried "not merely until the ninth year but by now the fourth period of nine years." It was his friends who urged publication. Yet the famous first published edition held many errors. "The painful truth was that it was not really ready for publication even when it was released to the printer." It "was composed . . . in the driblets of leisure available to a member and occasional official of a harassed Cathedral Chapter concerned with preserving the sources of its income."

Rosen's view is wide. We read of what many contemporaries made of the text, sometimes even of its misprints and omissions (which are controlled in our time by the autograph version). Rosen lets us into what Galileo said, for instance, on the lack of fit between the small observed changes in the brightness of Venus and the large differences in the distance between the earth and Venus. Copernicus alluded to the phases of planets but simply ignored this difficulty; it was left to the telescope to show the changing phases of Venus, which tend to iron out the variations in apparent brightness to the unresolving eye. Our translator-guide is by no means a detached commentator; he heatedly defends, with the aid of Henri Poincaré, the reality of the terrestrial motion against those who airily dismiss it even today by invoking the powerful transformations in coordinates opened up by modern gravitational theory. Their case is not illogical, but it makes coincidences out of phenomena the realistic theory can ascribe to a single physical cause.

Now we have both a serviceable volume and a luxurious and definitive one; the generous Copernicus would not complain.

COMMUNICATING WITH DEAF PEOPLE: A RESOURCE MANUAL FOR TEACHERS AND STUDENTS OF AMERICAN SIGN LANGUAGE, by Harry W. Hoemann. With illustrations by Shirley A. Hoemann. University Park Press, 233 East Redwood Street, Baltimore, Md. 21202 ($10.95). A fascinating new natural language has arisen here in the U.S. and is still growing apace. It is at the point of generating a literature of its own; already a play and a work of poetry have been published that were "originally conceived in Sign." (American Sign Language is otherwise known as ASL or "Sign.") This small book is not yet truly a general reader's introduction; it is intended for teachers of Sign, mainly as a second language for people with hearing who speak English. It is thin in vocabulary examples, but by that economy it is made unusually readable. Its concern with general structures, with the wide variety of strategies and with the recent history of Sign and of Sign studies gives it lively interest. Most of the literature on Sign has so far been either highly utilitarian and detailed or linguistically technical; this book presages wider treatments to come. The drawings are helpful.

The immediate message of the book is the striking richness of Sign. This rule-governed language is not made up of finger spelling, English speech translated word by word into signs or even the deft use of isolated conventional signs, although they all have roles. The language is rather the living production of a big human community, the community (one almost could say the nation) of half a million people in the U.S. It bears the marks of any working language, with dialects, dictionaries (and off-color words not listed in the dictionaries) and of course a complex and powerful set of grammatical rules. It is less than 20 years since the first analysis of the signs themselves was made by William C. Stokoe, Jr., who found that combinations of a dozen locations of the hand with similar numbers of configurations and movements were the elements from which all the signs are built. This syllabary has led to computer simulations, to a plausible notation and to a useful sequential dictionary. Recently it has been shown that the scheme has weaknesses; for example, it neglects such side strategies as expressions of the face and certain imitative gestures. It nonetheless stands as a firm working basis for study.

ASL proceeds in space as well as in time. Its multidimensionality has added to its grammar much that the linear flow of speech cannot match. Kinship words are few in ASL, as they are in English, yet in Sign one can distinguish female cousins from male ones. Gender is indicated by the placement of the sign: the right forehead and the temple for male, the right cheek and the chin for female. Location at the forehead adds to the quick recognition of signs for "know," "wise" and so on; the heart accommodates signs such as "grief," and perceptual acts are signed near the eye, the mouth or the ear. Plurals? Well, as in Chinese, add the sign for "group": a small gesture for a small group, an exaggerated sign for a big group and a double clawing downward and away for a "surprisingly large number of items." Or point out one by one the several places where a soldier is standing.

All kinds of references, pronouns, agent-patient distinctions, conjunctions and other devices in the string of more algebraic languages are carved out of the space where signs are formed: with "coffee" on the left, "tea" on the right, "which" signed over both, the use of "or" and "and" is sparing. Tense too takes spatial form: the past is thrown out of the palm back over the right shoulder, whereas the future is grasped forward. Changes in stance and expression mark boundaries of phrase and clause; explicit role playing helps the structure of complicated citations. The conditional and the subjunctive do not mimic the English forms of modified verbs but often employ a questioning expression for real conditions and negating gestures for contrary-to-fact statements. "If HAVE MONEY is signed with lowered brows, a slight squint and slightly pinched lips," the penniless inference is plain.

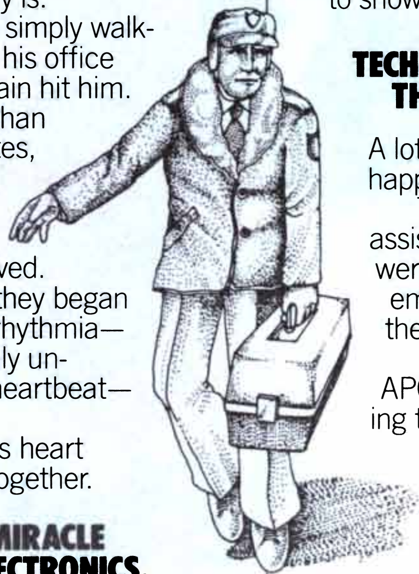
There is too much richness here in a little room to cite more fully. Certainly ASL and its counterparts in other lands are now prime material for the study of language itself, and they threaten to break the mold so strongly pressed on all forms of speech in this babbling world.

MOTOROLA HELPED BRING

His name isn't important.

His story is.

He was simply walking back to his office when the pain hit him.



In less than three minutes, an ambulance with two paramedics arrived.

But as they began to work, arrhythmia—erratic, wildly uncontrolled heartbeat—set in.

Then his heart stopped altogether.

A MIRACLE OF ELECTRONICS.

Fortunately, the ambulance was equipped with a Motorola APCOR Coronary Observation unit.

It put the paramedics in immediate touch with an emergency physician at the hospital miles away.

Not just by voice communication; that doesn't give a doctor all the information he needs.

But at the very same instant, by Motorola telemetry, an electrocardiogram of the activity of the man's heart was being transmitted.

A miracle of electronics—microelectronics—was about to show what it could do.

TECHNOLOGY DOES THE TALKING.

A lot of things were happening at once.

The medical assistants at the scene were talking to the emergency room at the hospital.

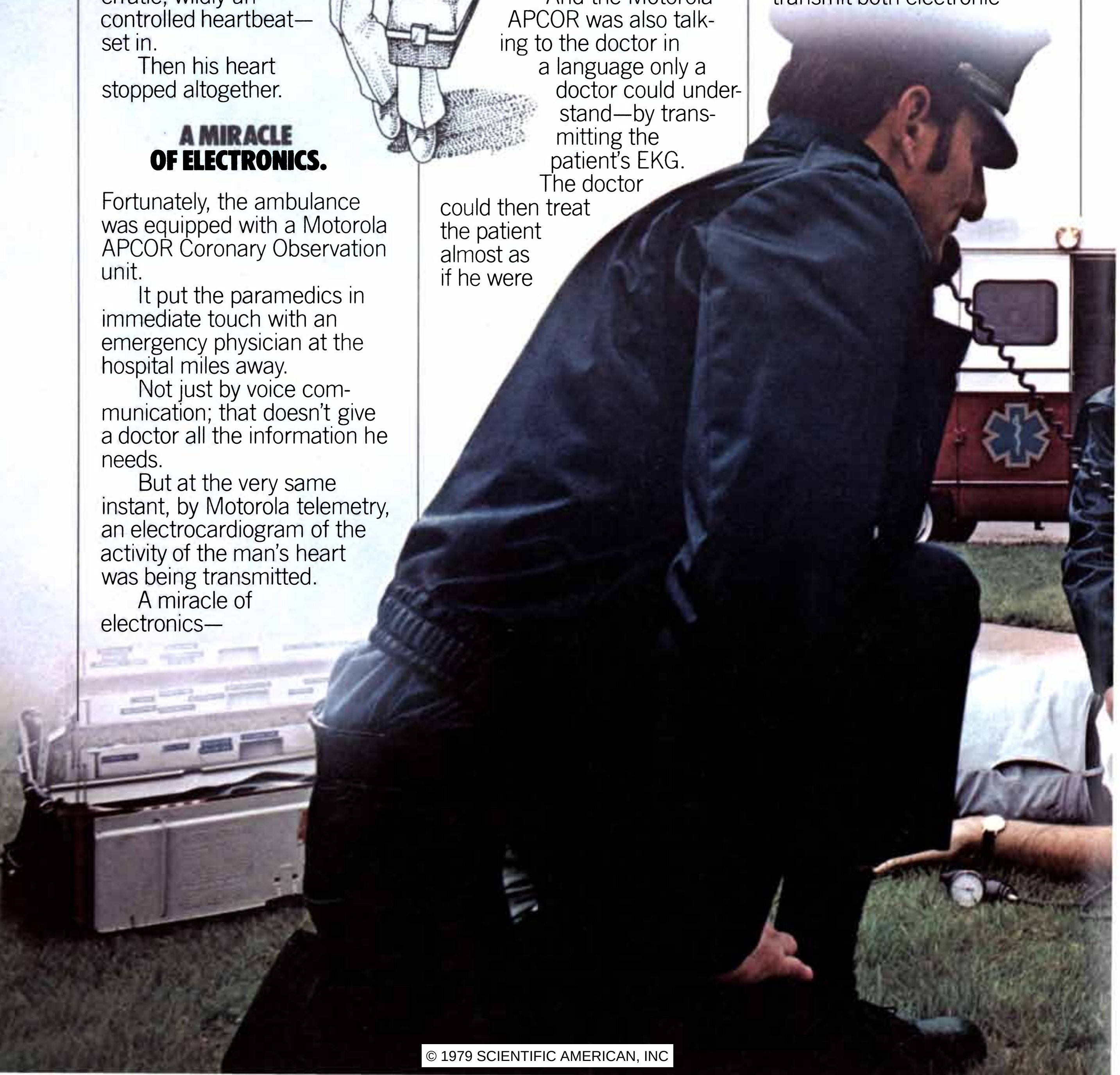
And the Motorola APCOR was also talking to the doctor in a language only a doctor could understand—by transmitting the patient's EKG.

The doctor could then treat the patient almost as if he were on the spot.

In a few minutes, the man's heart began to beat. He was once again on his way to being alive, in every sense of the word.

MICROCIRCUITS MAKE IT HAPPEN.

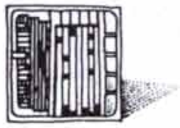
APCOR is made possible by Motorola microcircuits, tiny information processors that transmit both electronic



THIS MAN BACK TO LIFE.

signals and the human voice.

Microcircuitry is also at the heart of the many other kinds of two-way communications equipment we make.



A microcomputer, drawn larger than life.

But communications equipment is only part of what contributes to our nearly two billion dollars a year in sales.

MAKING ELECTRONICS HISTORY.

Thanks to Motorola microelectronics, we create all kinds of remarkable systems that would have been inconceivable not long ago.

A system to help power companies handle peak loads without danger of blackouts.

An electronic car-engine management system that can save gasoline.

Even a communications system to help probe Mars.

But then, we've come a long way from the time we first made history by putting radios into cars (we went on to put alternators and electronic ignitions into them) and later put popular-priced TV sets into homes (they're a product we don't make here at all anymore).

Today, Motorola is one of the world's largest manufacturers dedicated exclusively to electronics, as well as one of its foremost designers of custom and standard semiconductors.

Many of the things we make are changing people's lives.

Others are actually saving them.

MOTOROLA

Making electronics history since 1928.

APCOR, Motorola and Ⓜ are registered trademarks of Motorola, Inc.

For further information, write Public Affairs Office, Corporate Offices, Motorola, Inc., 1303 E. Algonquin Road, Schaumburg, Illinois 60196.

Established 1845 SCIENTIFIC AMERICAN February 1979 Volume 240 Number 2

The Verification of the SALT II Agreement

The U.S. has at its disposal ample "national technical means" of surveillance to detect any attempt by the U.S.S.R. to gain a significant military advantage by violating a new arms pact

by Les Aspin

The keystone of any international arms-control agreement is the ability of each side to make sure the other side abides by it. Without adequate verification of compliance agreements such as the bilateral strategic-arms pacts between the U.S. and the U.S.S.R. are bound to collapse. As the Senate prepares to debate the ratification of the new treaty emerging from the second round of strategic-arms-limitation talks (SALT II) between the two superpowers, charges that the Russians will seek to evade its provisions are beginning to be heard. The charges raise several fundamental questions of verification: How could the U.S.S.R. go about cheating? How could the U.S. discover such violations? What would the U.S.S.R. stand to gain and what would the U.S. stand to lose if the Russians were to violate the SALT II treaty? Let us examine each of these questions in turn to ascertain just what the real problems of verification are.

The SALT II agreement will apparently consist of two basic parts: a treaty lasting through 1985 and a protocol lasting through 1982. A summary of the various provisions included under these two headings is given in the illustration on the opposite page.

The new SALT treaty will provide in the first instance for a gradual reduction in the total number of strategic offensive-weapons launchers allowed on each side, from 2,400 at the time of ratification to 2,250 by 1982. For the purpose of the treaty strategic launchers are defined in such a way as to include land-based intercontinental ballistic missiles (ICBM's), submarine-launched ballistic missiles (SLBM's) and long-range heavy bombers.

The U.S.S.R. could try to evade the ceiling on the total number of strategic launchers in three ways: by deploying new types of strategic weapons, by deploying more weapons of the existing types or by converting nonstrategic, tactical weapons into strategic ones (for example by increasing their range).

The first of these cheating methods—deploying new types of strategic weapons—is perhaps the least feasible and most easily detectable way in which the U.S.S.R. could violate the SALT II total-launcher ceiling. The introduction of a new strategic weapon involves at least five stages: research, development, testing, production and deployment. At any one of these stages the present ability of the U.S. to detect clandestine activity on the part of the U.S.S.R. ranges from fair to excellent. The key point, however, is that the Russians would have to disguise all five stages, and the odds against their successfully doing so are extremely high.

Consider the ways in which the U.S. is currently able to monitor just one of these stages: the testing of strategic launchers. U.S. line-of-sight radars can identify the distinctive "signature" of reflected microwaves associated with each major type of Russian missile. In addition over-the-horizon radars can penetrate deep into the interior of the U.S.S.R. and recognize the characteristic pattern each type of missile makes when it disturbs the earth's ionosphere. Early-warning satellites, originally designed to detect a Russian ICBM attack, can also serve to monitor missile tests; the infrared sensors on these satellites can identify the rocket-exhaust plume of a missile as it is being test-fired. Finally, the U.S. has a complex array of sensors, including assorted photographic gear, on ships and planes that routinely monitor missile-test impact areas on the periphery of the U.S.S.R. and in the Pacific. The information gathered from these sources can be used to distinguish new types of missiles from old ones.

In short, the "national technical means" of surveillance available to this country for observing Russian missile tests are multiple, redundant and complementary. They enable the U.S. to detect all long-range missiles fired from test sites in the U.S.S.R. They are, in fact, far more reliable than most human intelligence gathering (that is, spying), which may yield second-hand, dated information or even false, planted information.

To repeat, testing is only one of the five steps that must be taken before a new weapon is ready to be introduced to the strategic arena. Other means of detection could uncover a Russian attempt to evade this particular treaty provision either before testing (during the research and development stages) or after testing (during the production and deployment stages).

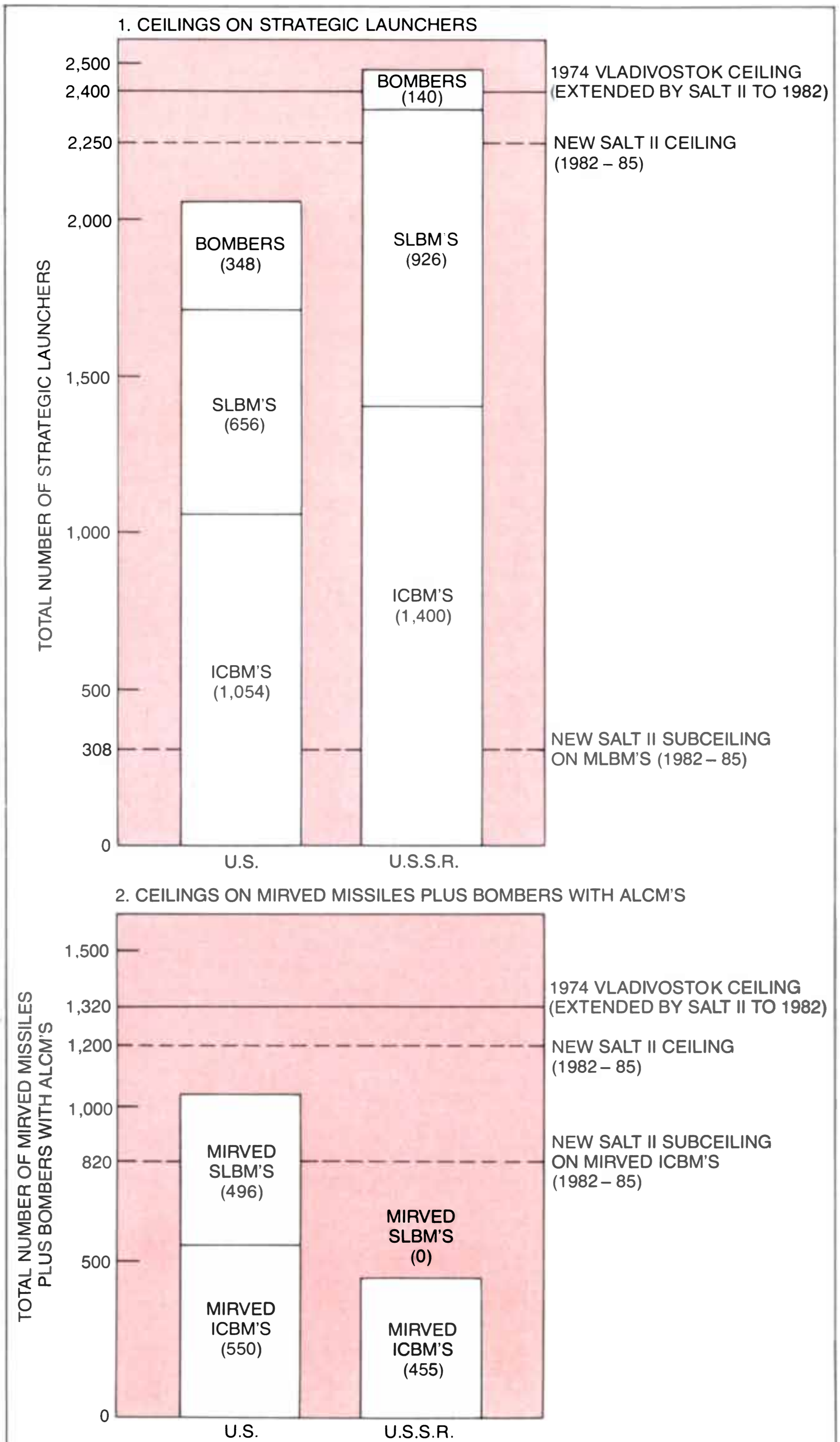
The second method potentially available to the U.S.S.R. for cheating on the overall strategic-launcher ceiling—deploying additional weapons of existing types—is more difficult to monitor than the first cheating method, but here the detection capabilities of the U.S. are still very good indeed. The national technical means of surveillance adopted by this country are particularly effective in detecting the production and deployment of additional missile-carrying submarines and heavy bombers.

Strategic submarines are large ships, displacing between 8,000 and 9,000 tons and measuring more than the length of a football field. They are hard to hide. Moreover, there are only two shipyards in the U.S.S.R. that currently build submarines of this type. Both these two sites and other potential sites for constructing submarines are constantly watched. Even if a new missile-carrying submarine were to be built under impenetrable cover (assuming the simultaneous successful disguising of all support activities), the new submarine must eventually leave its shipbuilding ways. From that point on there would be innumerable opportunities for observing it. The deployment of additional missiles on Russian submarines also cannot go undetected.

Similarly, the detection of any increase in the number of heavy bombers in the Russian air force is a fairly straightforward task. The production lines for the two existing types of heavy bomber, referred to by U.S. military analysts as the Bear and the Bison, have long been closed. Renewed production of these two aircraft could be detected by the U.S. with a high degree of confidence, as is evident from the demonstrated ability of this country to monitor the production of much smaller pieces of Russian military equipment, such as tanks. The additional deployment of heavy bombers would also be hard to hide. There are no more than 10 heavy-bomber airfields in the U.S.S.R., and all of them are closely watched by U.S. surveillance satellites. If the Russians were to decide to build new airfields equipped to handle such planes, their detection would be easy; runways cannot be hidden underground. If, as some strategic planners expect, the U.S.S.R. were to deploy a new type of heavy bomber by the early 1980's, it too would be readily detectable.

Detecting the deployment of addi-

PART I: TREATY



3. SUBCEILING ON MLBM'S

4. BAN ON RAPID-RELOAD SYSTEMS

PART II: PROTOCOL

1. BAN ON DEPLOYMENT OF MOBILE ICBM LAUNCHERS AND ON THE FLIGHT TESTING OF ICBM'S FROM SUCH LAUNCHERS.
2. BAN ON THE FLIGHT TESTING AND DEPLOYMENT OF LAND-BASED OR SEA-BASED CRUISE MISSILES CAPABLE OF RANGES IN EXCESS OF 600 KILOMETERS.
3. LIMITATIONS ON THE FLIGHT TESTING AND DEPLOYMENT OF NEW TYPES OF BALLISTIC MISSILES.

MAIN PROVISIONS of the SALT II agreement worked out in recent months by negotiators for the U.S. and the U.S.S.R. are contained in two forms: a treaty lasting through 1985 (*Part I*) and a protocol lasting through 1982 (*Part II*). Abbreviations are resorted to in several of the illustrations accompanying this article for the following types of strategic weapons: intercontinental ballistic missiles (ICBM's), sea-launched ballistic missiles (SLBM's), modern large ballistic missiles (MLBM's), multiple independently targetable reentry vehicles (MIRV's) and air-launched cruise missiles (ALCM's). In effect the SALT II treaty subceiling on MLBM's applies to the U.S.S.R. only, since the U.S. has no plans for such large land-based missiles.

tional land-based strategic missiles would be only a little more difficult. Given the current Russian practice in such matters, the U.S. can reliably identify by means of satellite photography such telltale activities in the U.S.S.R. as the construction of new ICBM silos and the transport of missiles to new deployment sites. The emplacement of command-and-control systems and associated support equipment can also be detected.

If the Russians were to attempt to hide these activities, they would probably have somewhat more luck than they would with strategic submarines. After all, land-based ICBM launchers are smaller than submarines, and there are vast interior areas of the U.S.S.R. in which they could be built or based. Even in the case of additional ICBM's, however, the likelihood of successful large-scale deception is very small. The Russians might try to build more ICBM launchers in the open, on the assumption that the vastness of their country would hide the violation, but that would be a mistake. U.S. surveillance satellites currently provide complete photographic coverage of the U.S.S.R. at frequent intervals. If suspicions are aroused by the regular large-area survey photographs, "close look" cameras can be ordered to rephotograph the area in question, providing more detailed information. The present generation of high-resolution cameras on U.S. surveillance satellites are theoretically capable of making a clear photograph of an object one foot across from an altitude of 100 miles. Photographs at this resolution would leave little doubt about the nature of the activities in question.

Alternatively the Russians could attempt to deploy additional ICBM's under camouflage or at night. U.S. satellites, however, are now equipped with multispectral sensors that can penetrate camouflage and can also observe nighttime activity. Infrared sensors are particularly good at detecting underground missile silos and silos that have been camouflaged. As long as the ground in the immediate vicinity of the suspicious object is at a different temperature from that of the surrounding terrain (or has different infrared-emission characteristics) an underground or camouflaged missile silo will stand out in the infrared image.

Primarily because of the time it takes for U.S. photoanalysts to process the data contained in such satellite pictures, small-scale violations might be hard to identify. Any sizable effort to cheat (say a clandestine addition of 100 ICBM's), however, would surely be detected.

The third method of cheating on the overall strategic-launcher ceiling—converting nonstrategic weapons into strategic ones—presents a more challenging problem of verification. Two notable examples of Russian intermediate-range weapons that could be made long-range (and therefore strategic) weapons are the Backfire bomber and the SS-20 intermediate-range ballistic missile (IRBM).

There is little disagreement within the U.S. intelligence community that the primary purpose of the Backfire is to carry out missions in areas peripheral to the U.S.S.R. (such as Europe and China). Roughly half of the Backfires deployed so far have been assigned to naval-aviation missions, and the rest are part of the U.S.S.R.'s medium-bomber force. There is also little question, however, that the Backfire has some intercontinental capability, specifically for one-way missions with recovery in a third country, for round-trip attacks against the western U.S. and, provided the bombers are refueled in flight, for even longer round-trip missions.

Although the SALT II treaty will exclude Backfires from the overall count of strategic launchers, the treaty will be accompanied by a variety of assurances (some in the form of unilateral statements) that will limit the strategic value of the aircraft. These assurances could include limits on the production and deployment of the Backfire, restrictions on the employment of the bomber in conjunction with tanker aircraft capable of in-flight refueling and limits on the bomber's range and payload. Of these assurances the easiest to verify would be the limits on production and deployment (even assuming deception), simply because of the size and complexity of these activities. These constraints are as applicable to the Backfire as they are to the strategic Bear and Bison bombers discussed above.

Verifying tanker restrictions would be a little harder. U.S. Air Force pilots testify to the difficulty of midair refueling. It is extremely doubtful that the Russians would actually try to refuel Backfires during a war without having attempted some practice runs, and practice runs can be monitored by a variety of means, including listening in on aircraft communications. If the Russians wanted to take their chances and attempt wartime refuelings without rehearsals, however, there is no guaranteed means of verifying any such restrictions on tanker employment.

The most difficult of the SALT II Backfire assurances to verify involve the plane's characteristics, specifically its range and payload. Even with unhindered surveillance there has already been some dispute among U.S. analysts over the range of the Backfire. Assuming skillful and determined cheating on the part of the U.S.S.R., both the range and the payload of the Backfire could probably be disguised.

The other intermediate-range weapon that could be converted into a strategic weapon is the SS-20 IRBM. The SS-20 is not covered by the SALT II treaty, since its present range (3,000 kilometers) is less than the 5,500-kilometer lower limit used to define ICBM's. The potential problem stems from the fact that the SS-20 comprises the first two stages of the advanced three-stage SS-16 ICBM; moreover, the mobile launcher for the SS-20 is identical with that for the SS-16. By surreptitiously stockpiling SS-16 third stages and payloads the Russians could at some point in the future be in a position to upgrade SS-20's into SS-16's on short notice. This course of action could provide them with a significant

	STRATEGIC-WEAPONS LAUNCHERS	NUMBER OF LAUNCHERS DEPLOYED	NUMBER OF WARHEADS PER LAUNCHER	TOTAL NUMBER OF WARHEADS
MIRVED	MINUTEMAN III ICBM	550	3	1,650
	POSEIDON C-3 SLBM	496	10	4,960
	SUBTOTAL	1,046		6,610
NOT MIRVED	TITAN II ICBM	54	1	54
	MINUTEMAN II ICBM	450	1	450
	POLARIS A-3 SLBM	160	1 (3 MRV'S)	160
	B-52D BOMBER	79	4	316
	B-52G, B-52H BOMBERS	269	4	1,076 (+1,500 SRAM'S)
	SUBTOTAL	1,012		3,556
	TOTAL	2,058		10,166

U.S. STRATEGIC ARSENAL is broken down in this table into two broad categories: weapons launchers that carry multiple independently targetable reentry vehicles and those that do not. The number of nuclear warheads indicated for the Poseidon C-3 SLBM is believed to be an average figure. The number of B-52's available for strategic missions is an estimate. The number of warheads per strategic B-52 listed represents the standard loading; the B-52H and B-52G aircraft carry an additional complement of nuclear-armed short-range attack missiles (SRAM's). The three reentry vehicles on the Polaris A-3 SLBM are not independently targetable and hence are not counted as MIRV's. Cruise missiles and FB-111 bombers are omitted.

numerical increase in their ICBM force with very little warning.

The SALT II agreement, however, will specifically ban the further production, testing and deployment of the SS-16 ICBM and will further require that existing SS-16's be dismantled. Thus if the Russians were to try to augment their ICBM force by adding a third stage and a different payload to the SS-20's, in effect making them SS-16's, they would be doing so without any opportunity for testing the new system. The existing prototype models of the SS-16 have not been tested in almost two years, and the last test was apparently a failure.

In short, the issue of upgraded SS-20's turns out to be another testing issue. To have any confidence in upgraded SS-20's, particularly enough confidence to satisfy traditionally conservative Russian military planners, the U.S.S.R. would have to test some of them, and as was pointed out above surreptitious testing of new strategic missiles by the U.S.S.R. is a practical impossibility.

The problem of preventing the Russians from converting nonstrategic weapons into strategic ones is complicated by one other factor: the existence of between 90 and 100 "reconfigured" heavy bombers in their arsenal. These aircraft have been modified to serve in reconnaissance and antisubmarine-warfare roles, but they still retain their bomb bays. Any further permanent change in the configuration of most of these aircraft could be detected in time. Nevertheless, about a dozen of these planes were originally built in such a way that they could be rapidly converted into heavy bombers at their airfields, making prompt verification of their nonstrategic roles extremely difficult.

Besides limits on the number of launchers, the SALT II treaty will contain numerical limits on missiles equipped with multiple independently targetable reentry vehicles (MIRV's) and on bombers equipped with strategic air-launched cruise missiles (ALCM's). According to the treaty, the sum of these two types of systems will not be allowed to exceed 1,320. Furthermore, no more than 1,200 MIRVed missiles will be allowed on each side, and MIRVed ICBM's will be limited to 820.

There are four ways the Russians might try to increase their combined MIRV/ALCM total beyond the treaty limits: by constructing new ICBM silos and SLBM submarines for the additional MIRVed missiles; by substituting MIRVed missiles for unMIRVed ones in existing missile silos or submarines; by deploying MIRVed payloads on unMIRVed missiles in existing silos or submarines, and by placing strategic ALCM's on additional bombers.

The first way the U.S.S.R. might try to evade the MIRV/ALCM ceiling—constructing new silos and submarines for MIRVed missiles—would clearly be unfeasible, since (as was pointed out above) any cheating on the total number of such strategic launchers can be detected by the U.S. with a very high degree of confidence.

The second way the Russians could exceed the MIRV/ALCM ceiling would be by substituting MIRVed missiles for unMIRVed ones in existing silos or submarines. The U.S.S.R. currently has a number of silos and submarines containing unMIRVed missiles. Detecting their surreptitious replacement with MIRVed missiles requires that the U.S. know which Russian missiles are MIRVed and which silos and submarines contain which missiles.

In the SALT II negotiations both sides have agreed that all missiles of a type that has been tested in a MIRVed mode or has been fired from a launcher with a MIRVed warhead would be counted against the MIRV ceiling. The U.S. proposed this counting rule precisely because it facilitates verification. U.S. analysts already know from extensive observation which of today's Russian missiles are "MIRV-capable," and future MIRV-capable ICBM's and SLBM's can be detected at the test stage.

Although the U.S. knows which Russian missiles are MIRVed, another question remains: Is it possible to tell which silos and which submarines contain which missiles? The answer is again provided by the known differences among missile systems. First, Russian silos that contain MIRV-capable missiles are significantly different in appearance from those that contain unMIRVed missiles. Second, MIRVed launchers require different command-and-control systems, support equipment and other facilities, all of which are observable with existing U.S. satellites.

The various types of missile-launching tubes on strategic submarines can also be identified by U.S. surveillance satellites. Any attempt by the Russians to install existing MIRVed SLBM's on submarines with unMIRVed missiles would require the alteration of the launching tubes, the replacement of fire-control systems and other extensive modifications. These would take time; even a routine overhaul of a nuclear submarine takes from 30 to 36 months. Under the circumstances the changes would certainly be detectable.

Another method of evading the MIRVed-missile limits would be to take an unMIRVed missile and replace just its warhead. If the Russians were to deploy MIRVed payloads onto unMIRVed missiles in existing silos or submarines, that would be very hard to detect. Fortunately no such transferable payloads exist now, and the current generation of Russian missiles have design

	STRATEGIC-WEAPONS LAUNCHERS	NUMBER OF LAUNCHERS DEPLOYED	NUMBER OF WARHEADS PER LAUNCHER	TOTAL NUMBER OF WARHEADS
MIRVED	SS-17 ICBM	100	1 OR 4	~ 2,500
	SS-18 ICBM	170	1 OR 8	
	SS-19 ICBM	320	1 OR 6	
	SUBTOTAL	590		~ 2,500
NOT MIRVED	SS-9 ICBM	130	1	130
	SS-11 ICBM	620	1 (3 MRV'S)	620
	SS-13 ICBM	40	1	40
	SS-16 ICBM	20	1	20
	SS-N-6 SLBM	528	1 (2 MRV'S)	528
	SS-N-8 SLBM	286	1	286
	SS-N-17 SLBM	16	1	16
	SS-N-18 SLBM	96	1	96
	BEAR BOMBER	100	2	200
	BISON BOMBER	40	2	80
	SUBTOTAL	1,876		2,016
	TOTAL	2,466		~ 4,500

RUSSIAN STRATEGIC ARSENAL is estimated in this table on a similar basis. In accordance with the new SALT II "counting rules" the MIRV subtotals shown here include some 135 ICBM's that have not yet been MIRVed and hence still carry single warheads. In addition the intermediate-range bomber referred to by U.S. military analysts as the Backfire is omitted. In general numerical tallies of this kind fail to reflect substantial U.S. advantages over the U.S.S.R. in terms of missile accuracy and reliability. Moreover, such tables do not include the fact that the U.S. has thousands of tactical nuclear weapons capable of reaching targets in the U.S.S.R., whereas the U.S.S.R. has none in a comparable position to reach targets in the U.S.

characteristics that make it virtually impossible to transfer MIRV payloads from the new MIRVed missiles onto the old unMIRVed ones.

A final method of evading the MIRV ceiling would involve placing strategic cruise missiles on additional bombers. The treaty places a ceiling of 1,320 on the total number of MIRVed missiles plus bombers equipped with ALCM's. Could the Russians exceed that ceiling by producing more than the allowed number of ALCMed bombers?

For the foreseeable future the U.S. will be able to tell which Russian bombers are equipped with cruise missiles, since Russian cruise missiles are externally mounted and therefore visible. Internal mountings would present a problem, but so far the Russians have none. Internally mounted cruise missiles would be detected most readily at the time of their introduction, because the aircraft involved would presumably have to be sent to some central facility to be modified, and the U.S. follows the activities at such facilities quite closely.

The verification of the limit on internally mounted cruise missiles will be eased considerably by the adoption in the SALT II accord of a "type" rule, which states simply that if one bomber of a given type carries ALCM's, all bombers of that design would be counted as ALCMed bombers. Counting which bombers have cruise missiles, however, is not the same as verifying which bombers are strategic. Bombers might be fitted with cruise missiles that had short ranges, which would not qualify them as bombers armed with strategic cruise missiles. Bombers might also be fitted with long-range cruise missiles carrying non-nuclear payloads. Hence a separate and far more complicated problem is determining whether cruise missiles on bombers have strategic ranges (in this case more than 600 kilometers) and strategic payloads.

Under normal conditions the U.S. can obtain adequate if rough estimates of these characteristics, but there is no systematic way of verifying the range of deployed cruise missiles. Significant differences would be revealed neither by the missile's exterior nor by its flight test. Unlike ballistic missiles, cruise missiles do not have to be tested at full range or even near it for the military to have confidence in their performance. Like aircraft, they can be flown for a limited time under cruise conditions, and their range can be estimated on the basis of the amount of fuel consumed. As it happens, the U.S. Joint Chiefs of Staff do have good estimates of the range of existing Russian cruise missiles, and the U.S.S.R. does not yet have air-to-surface cruise missiles capable of strategic ranges. Moreover, the Russians are not expected to have many long-range ALCM's for a number of years, and so it is unlikely that they could exceed the numerical restrictions in the SALT II treaty before it expires in 1985.

As for the payloads of cruise missiles, there is no way at present to distinguish a nuclear-armed cruise missile from a non-nuclear one by external observation. Once again, however, the Russians are not expected to have long-range ALCM's with either nuclear or conventional warheads for a number of years, so that violations of this provision before the treaty expires are unlikely.

Assuming that the Russians do perfect strategic ALCM's, the U.S. would still not be at a loss. If the Russians were to begin refitting existing aircraft with new ALCM's, suspicions would be aroused and the U.S. would be aware of the potential for cheating. Even if a new Russian bomber were equipped with ALCM's but the U.S.S.R. falsely asserted it was not strategic and the U.S. was not able to contradict the assertion, it is doubtful that more than about 120 bombers would be available before the expiration date of the treaty.

The SALT II treaty will also contain a sublimit on the number of modern large ballistic missiles (MLBM's) allowed on each side. Any missile larger than the Russians' SS-19 (which has a "throw weight" of about 8,000 pounds) will count as an MLBM; any missile larger than the largest ICBM currently in the Russian inventory (the SS-18, with a throw weight of roughly 16,000 pounds) will be prohibited. The debate over the substitution of SS-l9's for SS-11's after the signing of SALT I provides ample evidence of the sophistication of U.S. monitoring techniques. The dispute turned on the question of whether the installation of SS-19's in SS-11 silos violated the SALT I provisions covering the substitution of "heavy" missiles for "light" ones. The consensus following the debate was that the substitution by the Russians did not violate the letter of the SALT I treaty but that it was inconsistent with one of the unilateral statements made at the time by the U.S. The main point here, however, is what the discussion revealed, namely that the U.S. knew precisely how much larger the SS-19 was than the SS-11.

The SALT II treaty will also prohibit "rapid reload" systems. The purpose of this provision is to protect against the possibility that the U.S.S.R. would stockpile extra ICBM's and fit them into existing launchers once a first salvo had been fired. Loading a 50-ton missile into a silo is considerably more complicated than putting a cartridge into a rifle. The elaborate equipment around existing silos necessary for such a system to work, to say nothing of the storage sites for extra missiles, would certainly be detectable with existing satellites. The Russians could scatter the equipment and extra missiles far from the silos and probably avoid detection in that way as long as they did not test the resulting system; they would then, however, not have a rapid-reload capability, and so there would be no violation of this particular provision.

In addition to the treaty lasting until 1985 the SALT II agreement will contain a protocol lasting until 1982, and there will be verification issues in the protocol too. One part of the protocol will ban the deployment and testing of mobile ICBM launchers. The potential for violation of this section lies in the possible deployment of the existing SS-16 ICBM in a mobile mode; no other mobile ICBM is expected before the protocol expires. There is no question of the ability of the U.S. to ascertain that the Russians have deployed a mobile land-based system. Nevertheless, under certain deceptive basing schemes such as the multiple-aim point, or "shell game," options discussed recently (which involve the construction of hundreds or even thousands of shelters, only a fraction of which contain missiles), verifying the actual number of missiles deployed would be very difficult.

The SALT II protocol will also ban the flight testing and deployment of ground- and sea-launched cruise missiles capable of ranges in excess of 600 kilometers. Since the ranges of cruise missiles cannot be determined accurately in the event of conscious deception, such a ban will not be verifiable.

The flight testing of U.S. cruise missiles has only recently begun, however, and these weapons are not scheduled to be deployed in militarily significant numbers until after the SALT II protocol expires. Current Russian cruise missiles are primitive technologically. The U.S. is far more advanced in the development of compact warheads, computer-guidance systems and small turbofan engines, the technologies that are the key to small but long-range cruise missiles. The U.S. Department of Defense has stated that in cruise-missile technology the U.S. is "10 years ahead of the Russians" and that U.S. cruise missiles now under development are "two or three generations" ahead of current Russian weapons.

There are nevertheless some existing Russian sea-launched cruise missiles that exceed the 600-kilometer limit by as much as 250 kilometers. Because of their primitive design, however, they are very large. Since any attempt to begin new deployments is observable, and since the Russians have no capability for deploying new, long-range ground- and sea-launched cruise missiles until after

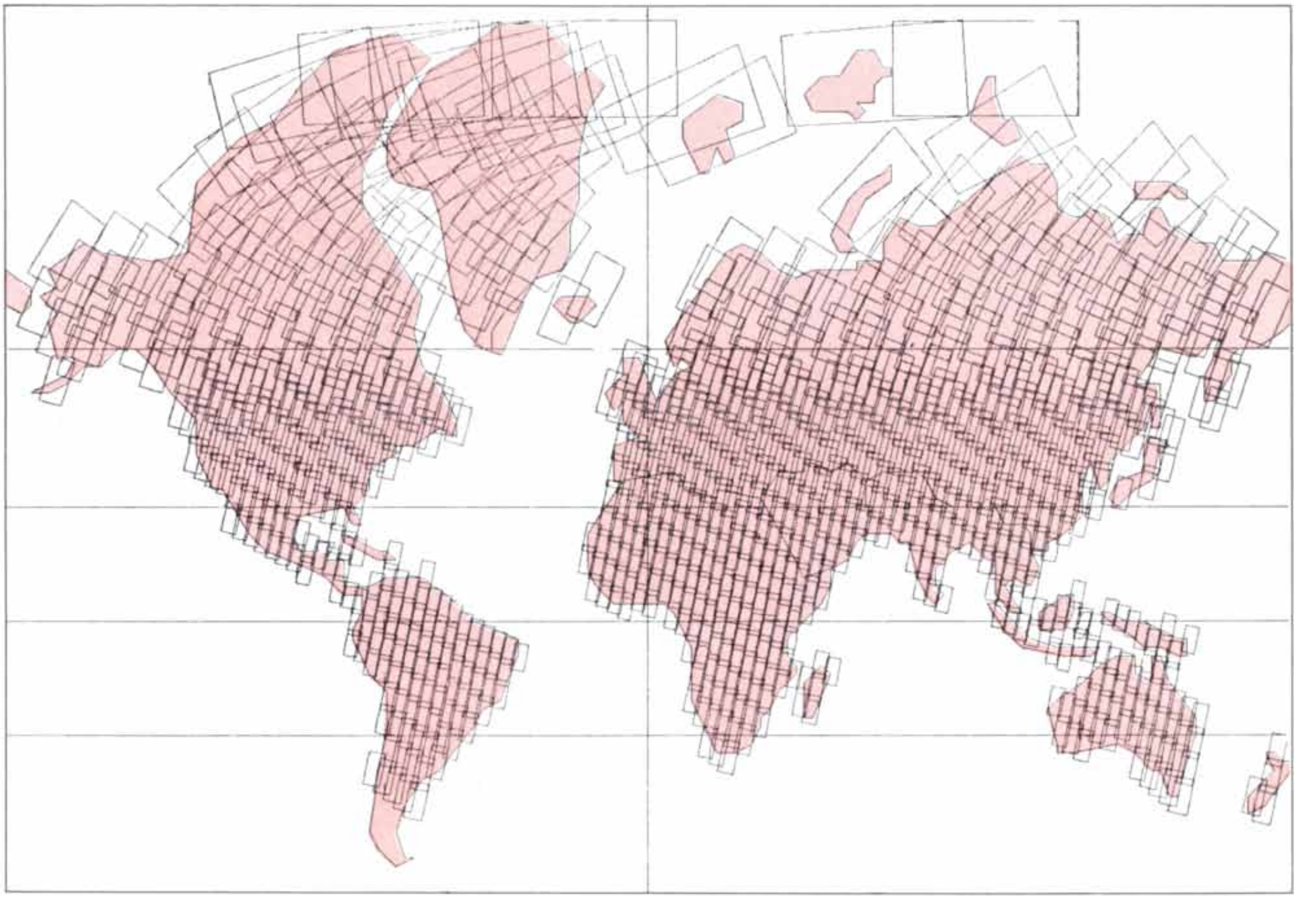
1982, there is virtually no potential here for violations by the U.S.S.R.

The SALT II protocol will include certain limitations on the flight testing and deployment of new types of ballistic missile. Although a full assessment of the associated verification problems depends on a detailed analysis of these limitations, there is reason for optimism. New ballistic missiles can be detected at the test stage, and added deployments of new missiles would be one of the easiest violations to detect.

The ability of the U.S. to detect potential violations of the SALT II agreement by the U.S.S.R. can be summarized in terms of three broad levels of confidence. First, there are the numerous cheating methods for which the verification capabilities of the U.S. are excellent, and the possibility of successful evasion on the part of the U.S.S.R. is remote. These contingencies include all the areas in which major violations by the U.S.S.R. could upset the present strategic "balance of terror": the deployment of new strategic weapons, the addition of even small numbers of bombers and SLBM's, the deployment of additional ICBM's on either a large scale or a moderate scale, the upgrading of the SS-20 missile to the status of a strategic weapon (unless it is upgraded without ever being tested), the deployment of additional MIRVed missiles or ALCMed bombers (in the absence of radically different systems), the upgrading of smaller missiles into modern large missiles and the introduction of rapid-reload systems.

Second, there are several areas in which the verification capabilities of the U.S. are at present quite weak. In all these cases, however, the possible cheating is not militarily significant. The problems of verification include detecting the small-scale deployment of additional ICBM's, monitoring the operational characteristics of the Backfire bomber, verifying that an untested SS-20 upgrade system does not exist and verifying the status of a limited number of heavy-bomber variants.

Third, there are a few areas, mainly those involving cruise missiles and transferable MIRV payloads, in which the U.S. may face serious verification problems at the next stage of the SALT negotiations. Although the Russian cruise missiles of today are primitive, at



WORLDWIDE COVERAGE available with current military photoreconnaissance satellites is suggested by this computer-generated map prepared by researchers at the Itek Corporation's Optical Systems division in connection with their work on the development of a new large-format camera for the National Aeronautics and Space Administration's Space Shuttle project. According to Itek, the new camera is designed "to assist in worldwide exploration for oil and mineral resources, mapping and monitoring of the earth's environment" by providing black-and-white and color imagery "with a coverage, clarity and fidelity that has not been available in the past" from nonmilitary systems. On a typical high-altitude mission, represented here, the Space Shuttle carrying the survey camera would be launched into a roughly circular polar orbit from Vandenberg Air Force Base in California. The rectangular "footprints" outline 3,468 individual photographic frames "exposed during daylight passes in an orbit inclined 97.6 degrees to the Equator and 294 nautical miles high." (The apparent enlargement of the frames at higher latitudes is a distortion arising from the Mercator projection adopted by the mapmakers.) Under suitable conditions the entire globe could be covered by two eight-day Space Shuttle flights, requiring approximately a mile of film contained in two recoverable film loads. On such a mission the photographic ground resolution would average "better than 75 feet (compared with 250 feet for future LANDSAT missions)." The current generation of U.S. military photoreconnaissance satellites are believed to fly in similar polar orbits, although at lower altitudes. In addition to their large-format cameras, such satellites carry "close look" cameras capable in theory of making photographs with a ground resolution of about a foot from an altitude of 100 miles.

some point in the future it will be virtually impossible to determine whether the range of a given Russian cruise missile is long enough to be considered strategic. The U.S. will also not be able to determine whether the payload of a given Russian cruise missile consists of conventional explosives or nuclear explosives. Furthermore, counting cruise missiles accurately will never be easy.

At present the payloads of the MIRVed Russian missiles are so different from their single-warhead payloads that the MIRV's cannot be installed on the old missiles. If the Russians eventually develop a transferable warhead, the U.S. could then face a serious verification problem.

The issues of strategic cruise missiles and transferable payloads are not problems of any great magnitude for the duration of the SALT II agreement. Looking forward to a SALT III pact in the mid-1980's, however, the verification problems are certain to be much more difficult.

In general the surveillance capabilities of the U.S. cited in this article are almost certainly underestimated. The reason is that the assessment considers only those methods of intelligence gathering that can be firmly relied on, such as satellite photography. Actually there is a great deal of other information the U.S. intelligence community receives that cannot be guaranteed in advance. For example, U.S. monitoring of internal communications and signals within the U.S.S.R. might pick up evidence of some activity that is not detectable by satellite photography. An undetected violation might even be revealed by a defector, whose defection could never be assumed in advance. Data obtained under such fortuitous circumstances would undoubtedly reduce even further the chances for successful violations.

The potential for violations is also overstated here because inordinately skillful cheating by the U.S.S.R. has been assumed throughout, a routine assumption in assessing one's own verification capabilities. It may not necessari-

PART I: TREATY

PROVISION	CHEATING METHOD		POTENTIAL FOR UNDETECTED ACTIVITY
1. Ceiling on total number of launchers (2,400 – 2,250)	Deploying new strategic systems		None
	Deploying more of existing systems	SLBM's	None
		Bombers	None
		ICBM's	Maximum of 100
	Converting non-strategic systems to strategic systems	Backfire, new production and deployment	None
		Backfire, employing tankers for inflight refueling	Minor
		Backfire, upgrading range and payload	Sizable
		SS-20, upgrading to SS-16	Minor
		Converting reconfigured bombers	Maximum of 12
2. Ceiling on MIRVed ICBM's and SLBM's plus bombers armed with ALCM's (1,320 – 1,200)	Constructing new missile silos or submarine launching tubes		None
	Substituting MIRVed missiles for unMIRVed ones in existing silos or submarines		None
	Deploying MIRVed payloads on unMIRVed missiles in existing silos or submarines		None with present systems; potentially large with future systems
	Placing ALCM's on strategic bombers		None in near future; minor in early 1980's
3. Ceiling on MLBM's (308)	Upgrading non-MLBM's to MLBM's		None
4. Ban on rapid-reload systems	Deploying rapid-reload systems		None

PART II: PROTOCOL

1. Ban on mobile ICBM's	Deploying mobile ICBM's	None
2. Ban on strategic cruise missiles	Deploying cruise missiles on land-based or sea-based launchers with a range in excess of 600 kilometers	None during protocol
3. Limitations on new types of ballistic missiles	Flight testing and deploying new types of ballistic missiles	Probably none

POTENTIAL VIOLATIONS of the main provisions of the SALT II agreement by the U.S.S.R. fall into three broad categories. First, there are the numerous cheating methods for which the verification capabilities of the U.S. are excellent and the possibility of successful evasion on the part of the U.S.S.R. is remote (*light-color boxes*); these contingencies include all the areas in which major violations by the U.S.S.R. could upset the present strategic balance. Second, there are the areas in which the verification capabilities of the U.S. are at present quite weak but in which the possible violations are not militarily significant (*medium-color boxes*). Third, there are a few areas in which the U.S. does not face serious verification problems at present but may at the next stage of SALT negotiations (*dark-color boxes*).

ly be a realistic assumption, however, because even the best-laid plans of a nation attempting to cheat can go awry.

Take, for example, the one known case in which the U.S.S.R. attempted a significant covert strategic-arms build-up: in Cuba before the missile crisis of October, 1962. Although the Russians clearly wanted to hide the emplacement of offensive missiles in Cuba, they were quite inept at doing so. On several occasions standard operational procedures and routines, which are necessary for the functioning of any large organization and are notoriously inflexible, betrayed their plans. For cheating to be successful everything must work perfectly. In the real world, however, unforeseen events upset plans. To be sure, one cannot count on any particular scheme's going awry, but any nation would be foolish to count on its not happening.

There has already been a heated debate in the U.S. concerning possible Russian violations of the SALT I agreement. That experience has raised a number of questions about the intentions of the Russians regarding their compliance with existing treaties, but it has not raised any questions about the ability of the U.S. to monitor what the Russians are doing. Indeed, the very basis for the allegations that violations have taken place is the detailed information the U.S. intelligence services have gathered on Russian actions since the SALT I pact was signed. The debate has centered not on what the Russians' actions have been but rather on what their actions mean. If the U.S.S.R. had engaged in illegal behavior that had gone unnoticed, this would obviously raise doubts about the detection capabilities of the U.S., but no one has even hinted that this might be the case.

So far only the potential for undetected violations has been considered here. An equally important issue is whether the Russians would attempt to cheat if they felt they could get away with it. The potential for violations is small; the likelihood of violations seems even smaller.

First, the SALT II framework provides enormous leeway for both sides to pursue strategic programs without cheating. Although the Russians would be able to build substantially larger forces without SALT II, they can still do much under the terms of the treaty. They can scrap existing missiles and replace them with more reliable and more accurate models. They can greatly increase their inventories of multiple-warhead missiles. They can direct a greater effort into areas not prohibited by SALT II, such as antisubmarine warfare, that could be perceived in the U.S. as threatening.

Second, even if the Russians became dissatisfied with the SALT II agreement after signing and ratifying it, they still would not necessarily cheat. Several alternatives might seem at least as attractive, if not more so: seeking the renegotiation of certain provisions, seeking to modify the terms of the SALT II pact in the SALT III negotiations, reneging on a part of the treaty (or even withdrawing from the treaty altogether), partly modifying their programs to comply with the treaty and so on.

Third, there is the question of what benefits would accrue to the U.S.S.R. from cheating. There could be no political gain unless the Russians made their transgressions public. No one is intimidated by weapons that are not known to exist. Yet if the Russians did make public the fact of their cheating, there would be enormous political repercussions. The U.S. Government, for example, might find itself pursuing an unprecedented arms buildup in response to the expressed demands of an aroused American public.

The real dangers stemming from Russian violations of SALT II would arise only if there were a significant military advantage to be gained by cheating, for example, if the Russians, after cheating for a few years, could then unveil a devastating superiority that would force the immediate surrender of the U.S. That, however, is impossible. Under the terms of the SALT II agreement the U.S. will still have a formidable strategic arsenal: almost 2,000 launchers and roughly 10,000 independently targetable warheads. To upset the strategic "balance of terror" the Russians would require much larger numbers of weapons than they are now allowed, and it would be impossible for them to acquire enough additional weapons without cheating on such a massive and pervasive scale that it would be detectable with certainty.

It helps to consider a number of plausible "worst cases" in which the U.S.S.R. could actually cheat on certain SALT II provisions and evade detection. The Russians might, for example, add as many as 100 ICBM launchers to their strategic arsenal clandestinely, but that would amount to an increase of less than 5 percent in their launcher force and would yield no discernible advantage. The Russians now have almost 2,500 missiles and bombers. Under the terms of the SALT II pact this total would drop to at most 2,250, a cut of about 250. Hence cheating would be more than outweighed by the reduction in forces required by the treaty.

The Russians might also be able to divert some Backfires to strategic missions in case of war. This substitution would add marginally to their second-strike forces but would correspondingly diminish their antiship capability and undercut their capability against enemies on their borders, which would hardly be a fair trade from their point of view.

The Russians might already have an untested SS-20 upgrade potential. Even if this potential were realized, the resulting SS-16 missiles would be the least accurate and least powerful ICBM's of the current generation. The diversion of SS-20's to intercontinental attack missions would also substantially reduce the threat to Western Europe and to China.

The Russians might convert some of their naval aircraft into long-range bombers. Again, this would marginally increase their strategic retaliatory strength while substantially diminishing the threat to the U.S. Navy.

The Russians might also develop an untested, nonrapid-reload capability. The benefit from having a launcher reloaded (at the optimum) 12 hours after a first firing is questionable; the silo could be destroyed in the interim and by that time the reloaded missile is likely to be no more than a potential "rubble-bouncer" anyway.

In other words, even if the Russians were to cheat in every way that might evade detection, they would add little to their strategic power, and they might actually reduce their military strength in other areas.

To sum up, the ability of the U.S. to verify Russian compliance with the SALT II accord is clearly essential to a successful outcome of the agreement. On close consideration, however, it becomes evident that the much-touted problems of verification are more imagined than real. The multiple and duplicative methods of detection at the disposal of the U.S. are sufficient to reveal any cheating on a scale adequate to threaten this country militarily. Certain small violations of the treaty could be achieved by the Russians without detection, but a handful of additional missiles or bombers would add too little to their arsenal to be militarily significant. In the political realm the Russians would stand to lose more than they would gain by violating the single most important treaty they would have with a foreign power.

It is in the future that verification problems might become critical. Technological advances, particularly those involving cruise missiles and transferable MIRV payloads, will stretch the monitoring capabilities of both sides once the SALT II protocol and treaty expire. Dealing with these systems under a SALT III agreement may well require a substantial lowering of the present standards of confidence for detecting violations. At that point a renewed examination of the entire verification issue will be in order.

How Bird Eggs Breathe

Bird embryos take up oxygen and discharge carbon dioxide by simple diffusion through microscopic pores in the eggshell. The process is regulated largely by pore geometry, which varies among bird species

by Hermann Rahn, Amos Ar and Charles V. Paganelli

The bird egg is a self-contained life-support system for the developing bird embryo. All the nutrients, minerals, energy sources and water utilized by the embryo during its incubation are already present in the freshly laid egg, so that the egg requires only warming by the parents and periodic turning to prevent the adhesion of the embryo to the shell membranes. Still, the egg lacks one crucial requirement: oxygen, which drives the metabolic machinery of the embryonic cells so that they can execute the complex maneuvers of development. How does the egg breathe, taking up oxygen from the surrounding atmosphere and discharging carbon dioxide, the waste product of respiration?

Gas exchange is usually associated with the periodic inhalation of a fluid medium (air or water), which carries oxygen to the capillaries of the lungs or the gills and removes carbon dioxide from the respiratory organ with each exhalation. The lungs or the gills are driven by muscles whose rate of pumping is determined by metabolic demand and controlled by the nervous system. Yet the eggs of birds and other organisms (such as insects, spiders, amphibians and reptiles) show no respiratory movements, and there are no air currents within the egg that could transport oxygen to the capillaries of the growing embryo. Instead the egg "breathes" by diffusion through thousands of microscopic pores in the shell.

The existence of these pores was first demonstrated in 1863 by John Davy, a member of the Royal Society from Edinburgh, without the aid of a microscope. He placed an egg in a jar of water and evacuated the air above the surface of the water with a vacuum pump. He noted that small bubbles of gas formed on the surface of the egg, and he deduced that there are minute openings in the shell.

Gas moves through the pores by the passive process of diffusion: the tendency for a high concentration of a molecule to run downhill to an area of lower concentration. Diffusion takes place because of the kinetic energy of gas molecules and does not require the direct expenditure of metabolic energy by the embryo; the lower concentration of oxygen inside the egg brings new oxygen molecules in through the pores from the outside, where the concentration is higher. Conversely, the concentration of carbon dioxide inside the egg causes those molecules to diffuse toward the outside, where there are essentially none. These diffusion processes are governed by the available pore area of the shell, the length of the pores and the concentration differences of the gases diffusing across the shell.

The water content of the air within the egg is greater than that of the air outside it, and so the pores will also allow water molecules (which are smaller than oxygen molecules) to diffuse out. Animals have evolved many specialized adaptations for conserving water, but bird eggs seem designed to lose it at a controlled rate. Most of the energy needed for embryonic development is taken from the fat stores of the yolk, and for every gram of fat burned an almost equal mass of metabolic water is generated. Therefore the relative water content of the egg will increase during incubation unless water is lost. If the relative water content at hatching is to equal that of the freshly laid egg, about 15 percent of the initial mass of the egg must be lost as water. As breeders of domestic fowl well know, this amount of water loss is essential for successful hatching.

Over the 21 days of its incubation a typical chicken egg weighing 60 grams will take up about six liters of oxygen and give off 4.5 liters of carbon dioxide and 11 liters of water vapor. Because of the water loss the egg will weigh only 51 grams after 21 days of incubation, and the newly hatched chick will weigh about 39 grams (subtracting the weight of the shell and the respiratory membranes). Investigations in our laboratories and in others have been aimed at elucidating the diffusive mechanisms responsible for this transfer of gases.

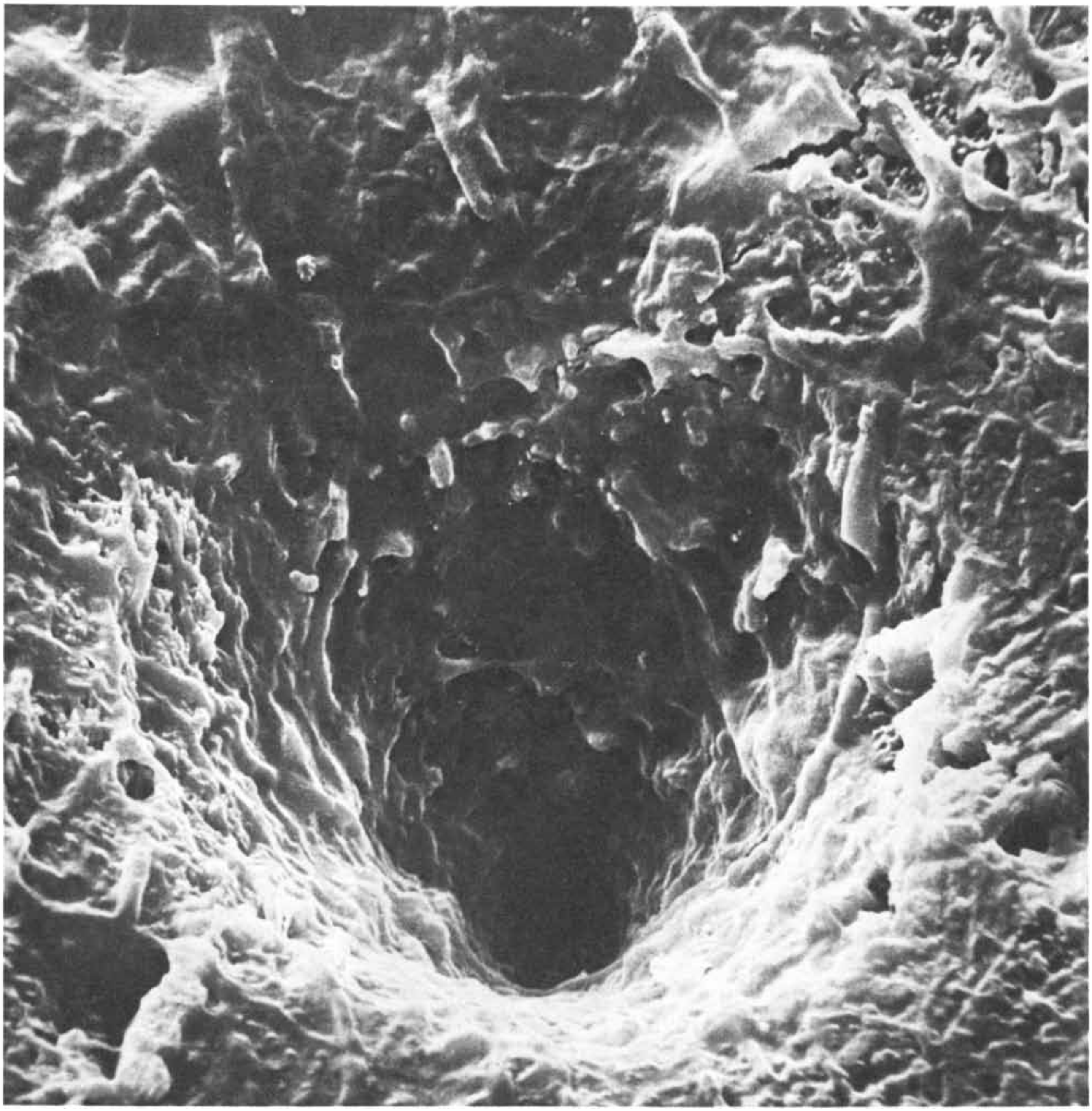
The three layers of the egg's integument—the shell and the two underlying membranes—should be familiar to anyone who has ever peeled a boiled egg. All three layers are laid down in less than 24 hours as the naked egg cell passes down the oviduct of the hen into the shell gland or uterus [see "How an Eggshell Is Made," by T. G. Taylor; SCIENTIFIC AMERICAN, March, 1970]. The eggshell itself has an outermost layer called the cuticle (it is missing in some species), which is a very thin sheet of organic material, often cracked to expose the surface of the eggshell proper. The shell consists of secreted calcium carbonate, which forms columnar calcite crystals incorporating a small amount of organic material. Imperfect packing of the calcite crystals leaves spaces between them that traverse the thickness of the shell, giving rise to the microscopic pores.

The outer and inner shell membranes consist of a meshwork of organic fibers. The fibers of the outer shell membrane are attached to the inside of the eggshell through the mammillary cones, the centers of crystallization of the eggshell during its formation. The two membranes differ in diameter of fiber, coarseness of weave and total thickness. The inner side of the inner membrane is lined with a thin "film" that appears to be continuous rather than simply an extension of the fibers. Soon after the egg is laid the spaces between the fibers in the two membranes are filled with gas.

The pores in the eggshell are the only path of gaseous communication between the outside environment and the embryonic membranes. They are cylindrical in shape and their openings are sometimes covered with secreted organic or inorganic particulate material. Careful etching of the shell with acid followed by staining makes the pores visible to the unaided eye: the typical chicken egg has about 10,000 distributed over its surface. The dimensions and numbers of the pores are established in the shell gland before the egg is laid; thereafter they remain unchanged. As we shall see, the shape, size and number of the pores in the eggshell vary among bird species, and it is the aggregate geometry of the pores that determines the diffusing capacity of the eggshell.

Since the bird embryo cannot directly control its gas exchange, the permeability or conductance of the shell and membranes to gases must be delicately adjusted to meet the embryo's metabolic needs. If the gas conductance is too high, the oxygen requirement of the embryo will be amply met but too much water will be lost, resulting in dehydration. If the gas conductance is too low, the embryo will either suffocate for lack of oxygen, be poisoned by its own carbon dioxide or drown in its own metabolic water. A happy medium must be struck that provides optimal gas pressures in the embryo and ensures a finite loss of water from the egg.

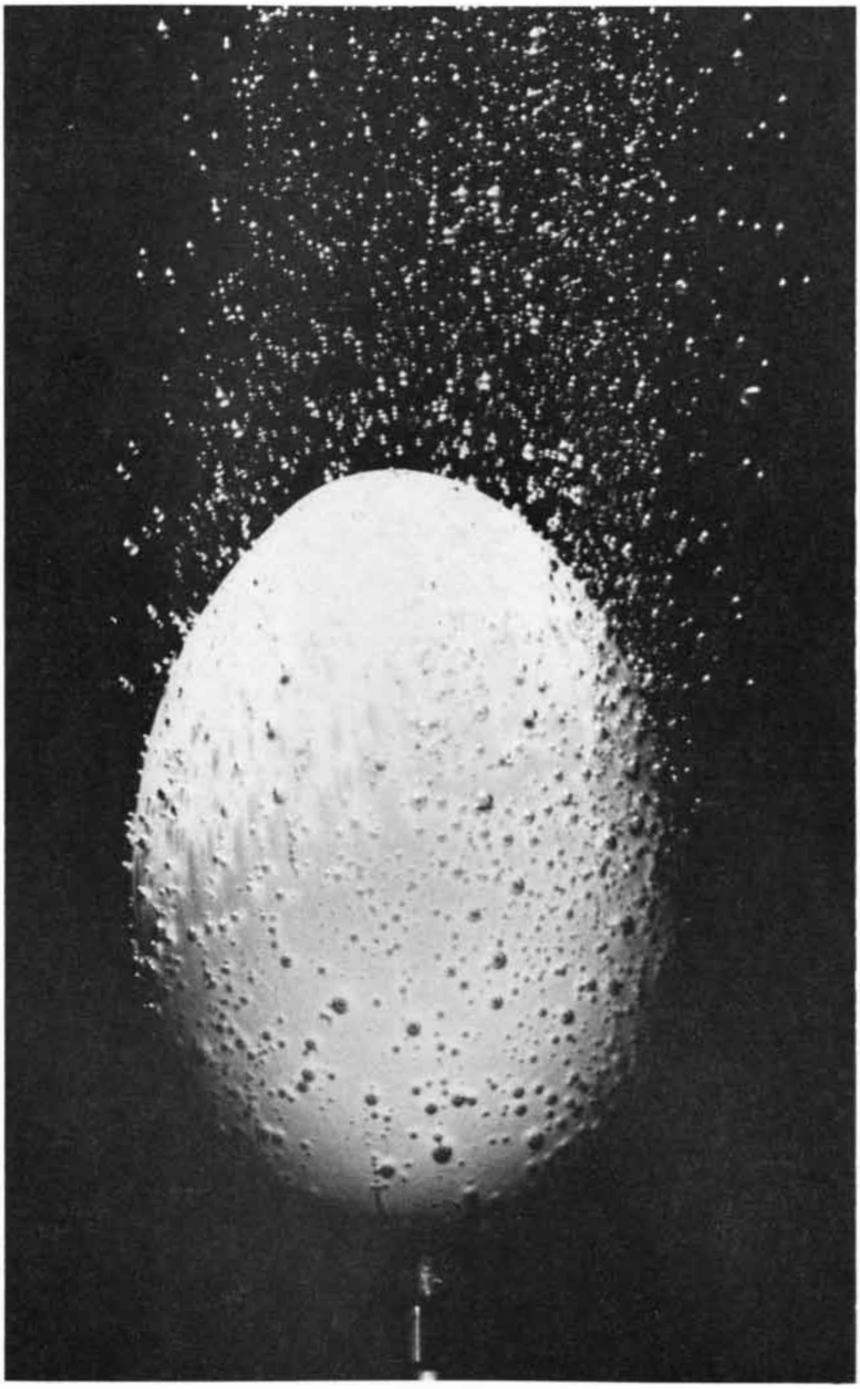
Water is lost from the egg at a constant rate, but the uptake of oxygen increases considerably over the course of incubation. In the chicken the first 18 days of incubation are designated the prenatal period. On days five and six the chorioallantois (the respiratory organ of the embryo, analogous to the placenta of mammals) extends out from the embryo, making contact with the continuous film lining the inner membrane and establishing a network of capillaries. By day nine the chorioallantois has invested about half of the inner surface of the



PORE IN THE EGGSHELL of the plaintive cuckoo (*Cacomantis merulinus*) of Java is magnified 3,800 diameters in this scanning electron micrograph. The conical pore opening narrows into a cylindrical tube a few micrometers in diameter that passes entirely through the eggshell. Atmospheric oxygen enters the egg and metabolic carbon dioxide leaves it by passive diffusion through thousands of similar pores distributed over the egg surface. This view of the pore opening is unusually clear; in most eggs the pores are concealed by secreted organic or inorganic material. Micrograph was made by J. H. Becking of Institute for Atomic Sciences in Agriculture in the Netherlands.

shell, and by about day 12 it has completely covered the inner surface. (The inner and outer shell membranes lie between the chorioallantois and the eggshell proper.)

During the prenatal period oxygen and carbon dioxide are exchanged across the chorioallantois and water vapor is continually lost. Since the shell is rigid, gas enters to replace the lost water and forms an air cell at the blunt end of the egg. The air cell increases steadily in size until it occupies about 15 percent of the internal volume of the egg at the end of incubation. The gas within the air cell is continuous with the gas found between the fibers of the outer and inner membranes. This continuity can be demonstrated by injecting pressurized air into the air cell underwater, causing air bubbles to escape from pores all over the surface of the egg. The gas tensions in the air cell are nearly identical with those in the air spaces of the shell membranes, so that the air cell provides a convenient site for sampling the gases within the egg.



DEMONSTRATION OF PORES in the eggshell without the need for a microscope can be achieved by injecting pressurized air into the air cell at the blunt end of a chicken egg. Bubbles emerge from the entire surface of the egg, indicating that it contains thousands of minute openings. This experiment is similar to one first performed by John Davy of Edinburgh in 1863. Photograph was made by Dennis R. Atkinson of State University of New York at Buffalo.

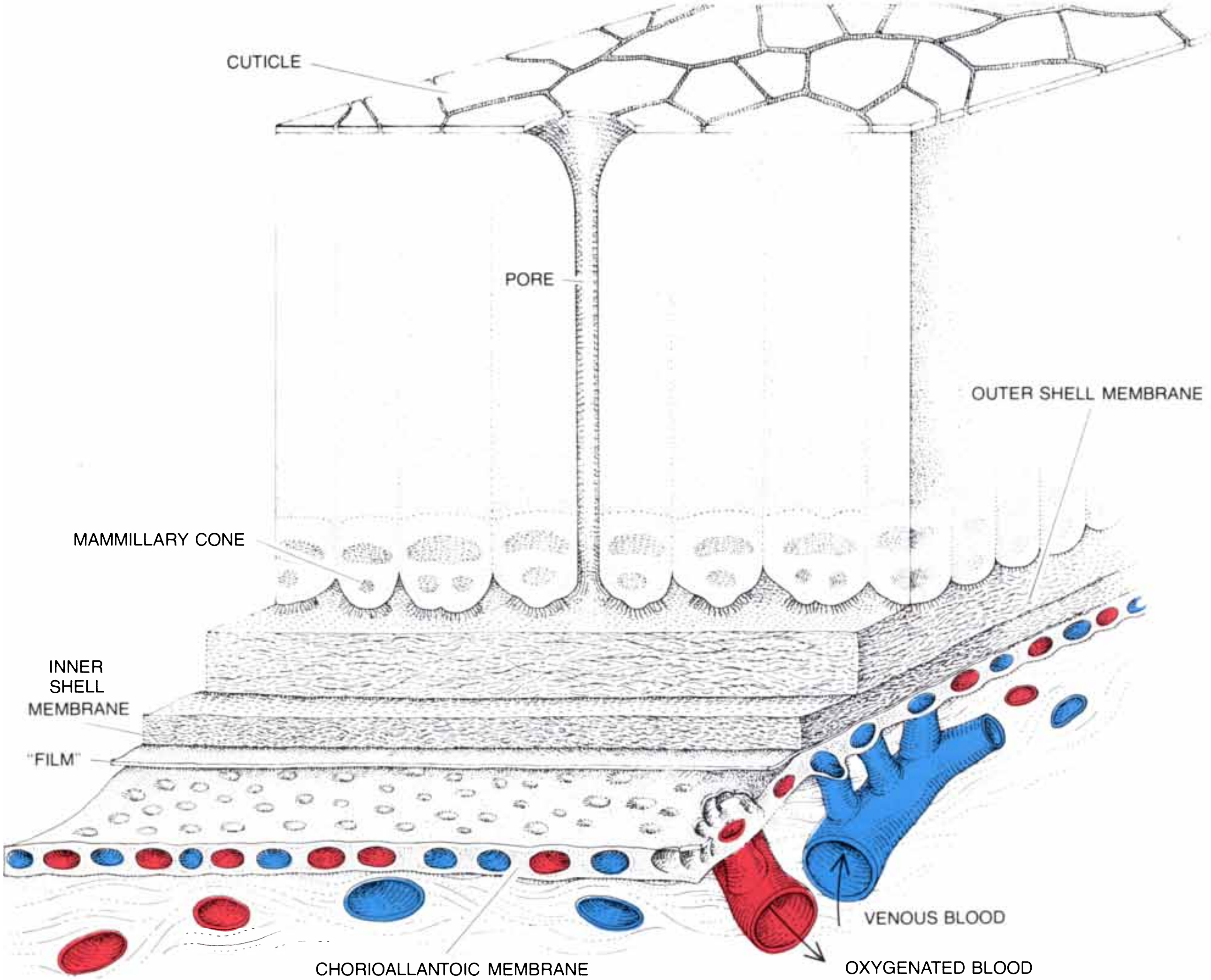
The oxygen consumption of the embryo increases slowly during the first week and a half of incubation. Between days 10 and 14 it rises steeply to reach a plateau of 600 milliliters of oxygen per day prior to hatching. This value is the maximal amount of oxygen that can be obtained by passive diffusion through the fixed pores of the shell. Six hundred milliliters of oxygen per day may not seem very impressive to the reader (who will consume that amount in two minutes), but the molecular traffic through the 10,000 pores in the eggshell is remarkably intense. Every second a net of about 20 trillion (20×10^{12}) oxygen molecules flows into the egg through each pore and 14 trillion molecules of carbon dioxide and 12 trillion molecules of water vapor flow out.

Nevertheless, the rigors of hatching require more oxygen than simple diffusion can provide. Where does the additional oxygen come from? Nature has provided a simple solution to this problem: on about day 19 the chick penetrates the air cell at the blunt end of the egg with its beak, a process called internal pipping. The chick then begins to breathe from the air cell, ventilating its previously unused lungs. Internal pipping is a crucial event: it allows the chick's lungs and their air sacs to be inflated so that oxygen is transported to the lungs by convection, or bulk flow.

The period of active breathing from the air cell within the confines of the intact egg is called the paranatal period, because the chorioallantois is still functional. During this period oxygen is therefore delivered by both diffusion and convection.

About six hours after the chick has penetrated the air cell it pips the shell, that is, it makes a small hole with the egg tooth on its upper beak and breathes atmospheric air for the first time. By now lung function is well enough established to allow the significant increase in oxygen consumption required for the final effort of hatching. At the same time chorioallantoic function begins to wane, although it persists to the end, when the respiratory membranes are left clinging to the inside of the shell as the chick emerges and the postnatal period begins. The smooth transition from passive gas transport to active gas transport is accomplished in from 24 to 36 hours.

It is now known that the flow of oxygen from the atmosphere to the blood



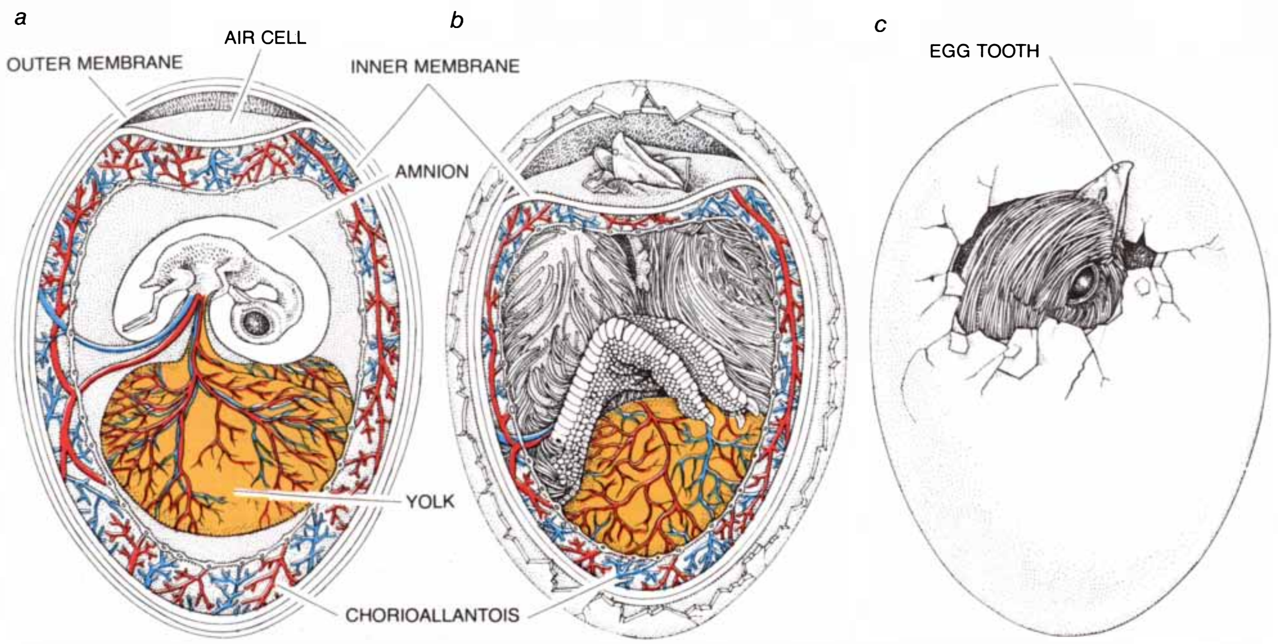
OUTER LAYERS OF THE EGG participating in gas exchange are shown in this cross section covering a depth of about .4 millimeter. The outermost layer is the cuticle, a thin organic sheet. Under it is the shell proper, made up of columns of calcite crystals and traversed by pores. The pores terminate in the loose, fibrous outer membrane; the outer and inner membranes differ in fiber diameter, coarseness of weave and total thickness. The very thin inner surface of the inner membrane, termed the "film," is most probably a continuous sheet. Attached to the film is the chorioallantois (the respiratory organ of the embryo), which is analogous to the placenta of mammals. Venous blood (*blue*) pumped by the embryonic heart flows to the chorioallantoic membrane, where it is replenished with oxygen that has diffused through the pores. Oxygenated blood (*red*) then travels to embryonic tissue. Simultaneously carbon dioxide diffuses out of venous blood.

capillaries of the chorioallantois is regulated by a series of barriers to diffusion: the shell and the inner and outer shell membranes. Respiratory physiologists express the permeability of a membrane to gas in terms of its conductance: the reciprocal of its resistance to diffusion. The partial pressures of the respiratory gases are commonly expressed in torr, or millimeters of mercury, one torr being equal to 1/760 of standard atmospheric pressure (the combined partial pressures of all the gases in the atmosphere at sea level, including nitrogen, oxygen and carbon dioxide).

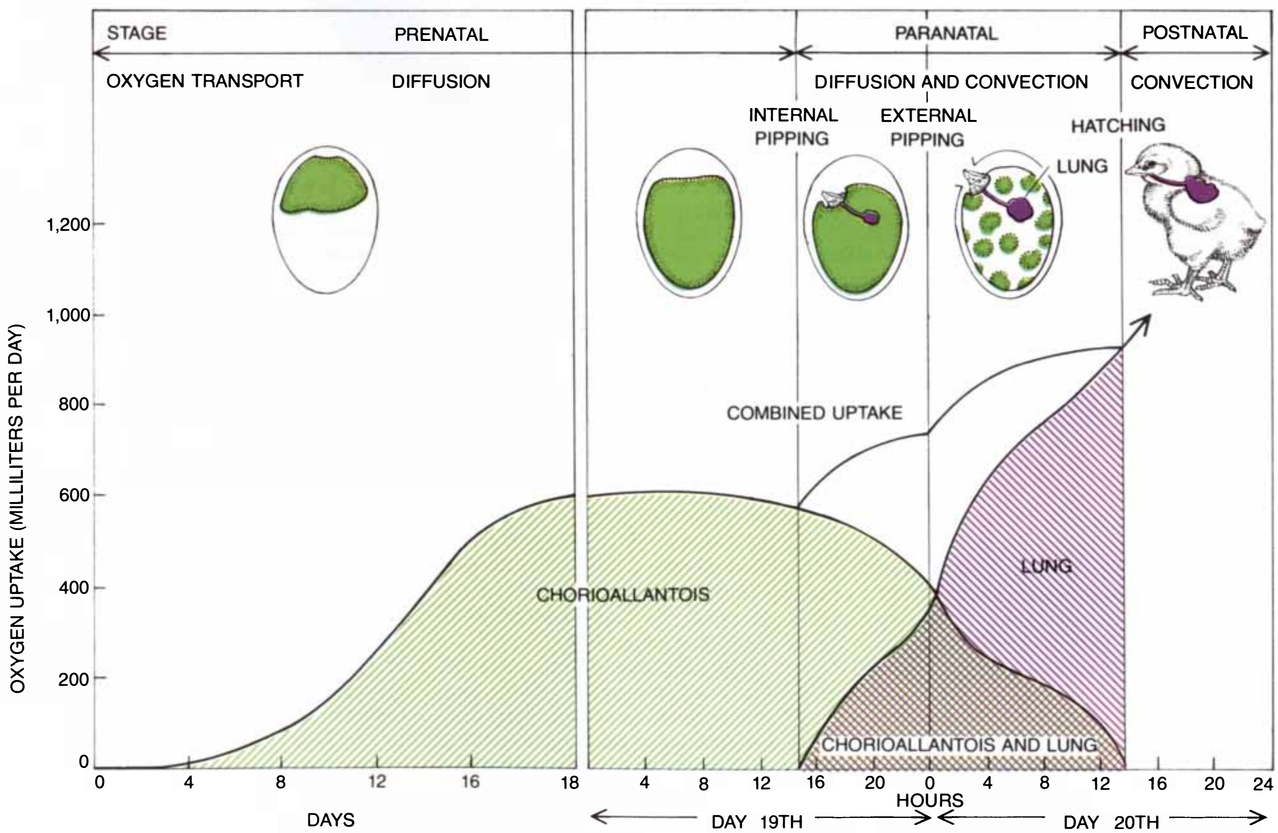
There is a drop of some 100 torr between the partial pressure of oxygen in the atmosphere (154 torr) and that in the newly oxygenated blood of the chorioallantois (58 torr). Nevertheless, an oxygen pressure of 58 torr in the blood is nearly sufficient to saturate the blood with oxygen, which is carried to all the tissues of the embryo. The venous blood returning from the embryo has an oxygen pressure of 22 torr. It passes through the chorioallantois, where it is replenished with newly diffused oxygen to a pressure of 58 torr. Conversely, the partial pressure of carbon dioxide drops from 47 torr in the venous blood to 38 torr in the oxygenated blood.

Measuring the gas conductance between the inner membrane and the chorioallantois proved to be extremely difficult because the gas partial pressures in the air cell and in the oxygenated blood must be measured simultaneously. This experimental problem was solved by Hiroshi Tazawa of Yamagata University, with whom two of us worked last year in the laboratory of Johannes Piiper at the Max Planck Institute for Experimental Medicine in Göttingen.

To measure the partial pressures of oxygen and carbon dioxide in the egg, Tazawa glued to the shell over the air cell a hypodermic syringe that was partly filled with air, so that the air in the syringe was in direct contact with the gas in the air cell. Over a period of a few hours the partial pressures of oxygen and carbon dioxide in the syringe and the air cell came into equilibrium. By analyzing the gas partial pressures in the syringe it was therefore possible to determine those in the air cell. To measure the gas pressures in the oxygenated blood Tazawa made a small hole in the eggshell and implanted an ultrathin catheter, or plastic tube, in the chorioallantoic blood vessel that carries oxy-



SHIFT TO ACTIVE BREATHING during the 21-day incubation period of the chicken embryo takes place in two steps. The prenatal period (*a*) is the first 18 days of development. Beginning at day five the chorioallantois protrudes from the embryo and covers the inner shell membrane with a net of capillaries that supply the embryo with oxygen and remove carbon dioxide. Water vapor also diffuses continually from the egg; the liquid water that is lost is replaced by gas to form an air cell at the blunt end of the egg. The paranatal period (*b*) begins on day 19, when the embryo penetrates the air cell with its beak, a process called internal pipping. The embryo then begins to breathe from the air cell, inflating its lungs and air sacs for the first time, although the chorioallantois still continues to function. About six hours later the chick breaks through the eggshell with the egg tooth at the tip of its upper beak (*c*) and starts to breathe atmospheric air. At this time chorioallantoic function begins to wane. Illustration is based on one by Hans-Rainer Duncker of the University of Giessen.



OXYGEN CONSUMPTION of the chicken embryo increases considerably over the course of incubation. The increase is gradual for the first week and a half, but between days 10 and 14 oxygen consumption rises sharply. On days 19 and 20 there is a gradual switchover from diffusive gas exchange through the chorioallantois to active breathing through the lungs. By the time the chick emerges from the egg its lungs are already working efficiently. This graph is based on work done by A. H. J. Visschedijk of the University of Utrecht.

genated blood to the embryo. The air in the syringe and a sample of the oxygenated blood were removed simultaneously and analyzed for their oxygen and carbon dioxide levels.

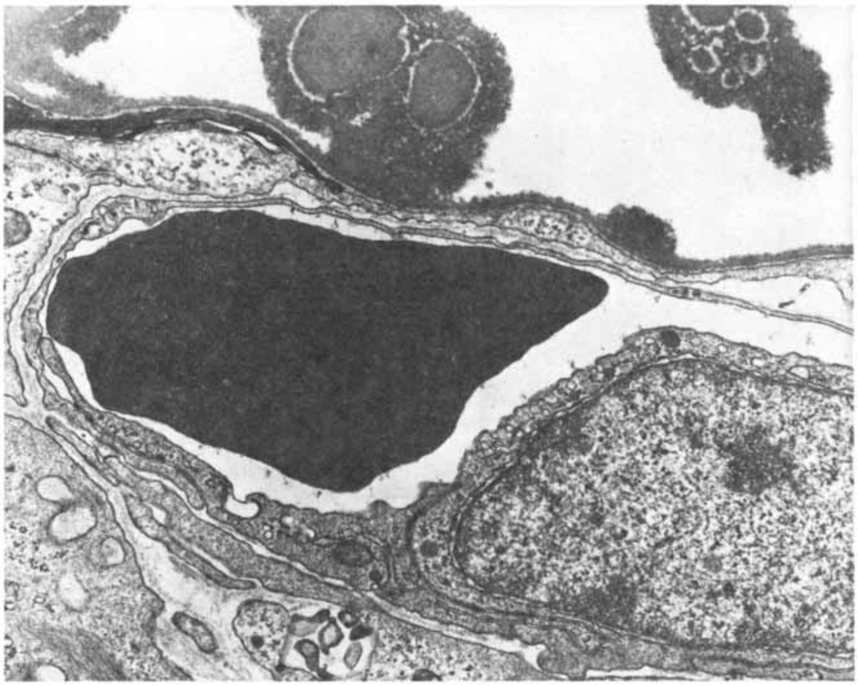
With this approach Tazawa found that the partial pressure of oxygen in the air cell is about 50 torr higher than that in the oxygenated blood, whereas the carbon dioxide pressure is 2.5 torr higher in the oxygenated blood than it is in the air cell. The fact that there is a 50-torr difference in oxygen pressure between the air-cell gas and the blood calls for explanation. It is unlikely that the air-filled spaces normally present between the fibers of the inner shell membrane are responsible. The difference may be caused partly by the resistance to diffusion of the thin continuous film between the inner shell membrane and the chorioallantoic epithelium. Another possibility is that some venous blood (with its low oxygen pressure of 22 torr) bypasses the capillaries where gas exchange takes place and mixes with fully oxygenated blood, thereby lowering its oxygen pressure to 58 torr. Of course, both phenomena may operate.

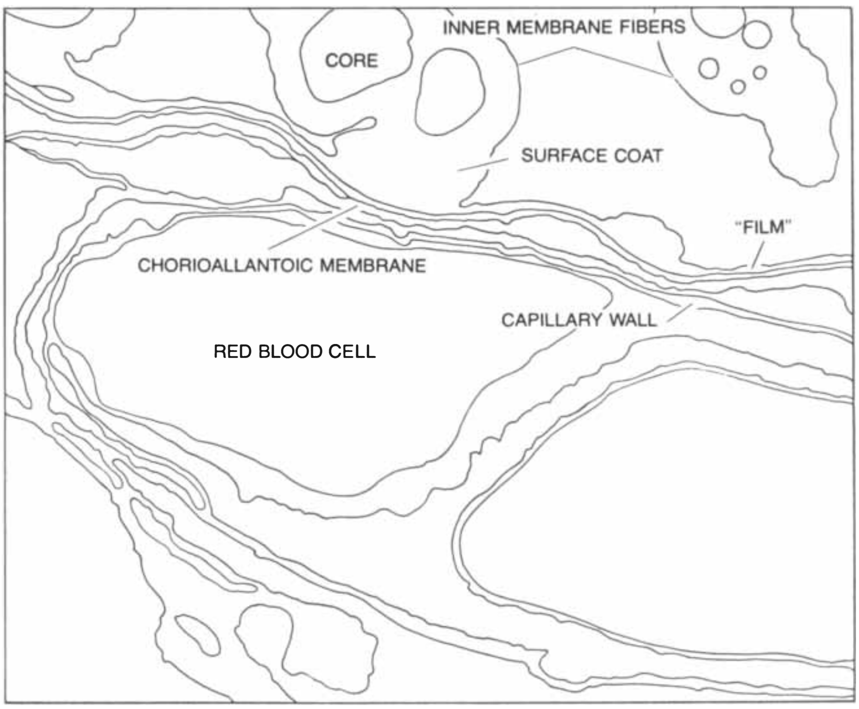
In summary, as one traces the partial pressure of oxygen from the atmosphere to the oxygenated blood leaving the chorioallantois of the chick, one finds only two areas where large differences in oxygen pressure exist: the shell proper and the inner shell membrane. The shell alone, however, is responsible for almost the entire partial-pressure difference for both carbon dioxide and water vapor.

Given the task of designing an efficient gas-exchange system, the respiratory physiologist would logically try to maximize oxygen conductance. In the egg, however, such purely respiratory considerations must be balanced against other factors important for the survival of the embryo. These factors include the need for an eggshell thick enough to provide mechanical protection for the embryo, to prevent harmful bacteria from invading the egg, to conserve essential fluids and to maintain the proper carbon dioxide pressure for a normal acid-base balance.

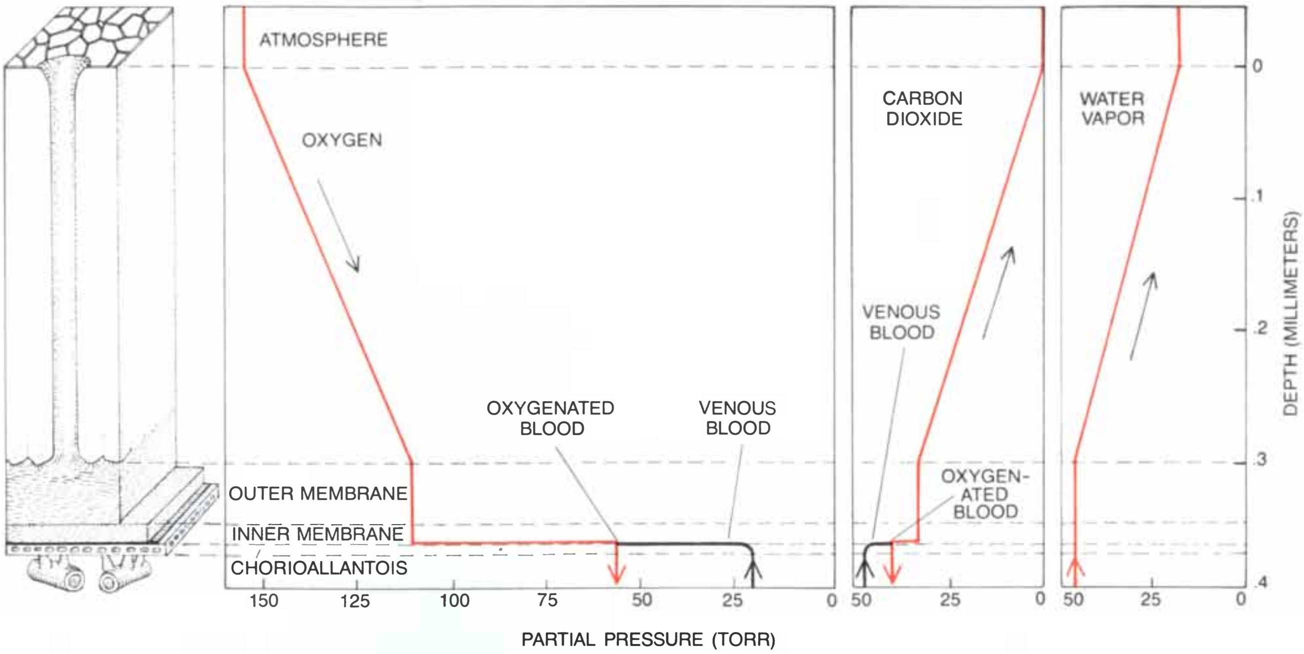
Having learned some of the principles that govern gas exchange in the chicken egg, we became curious about some of the other 8,500 species of bird eggs, which cover a broad range of shapes and sizes. The smallest egg known, laid by one of the 150 species of hummingbirds, weighs a quarter of a gram; it takes 240 of these eggs to equal the weight of a single 60-gram chicken egg. At the other end of the range is the egg of *Aepyornis*, the recently extinct elephant bird of Madagascar, which had an average weight of nine kilograms, equivalent to 150 chicken eggs. The *Aepyornis* eggshell alone weighed two kilograms, more than the fresh weight of an entire ostrich egg, which at 1.5 kilograms is the largest egg laid by an existing bird.

How do the shape and dimension of the pores differ in small and large eggs? Cyril Tyler and K. Simkiss of the University of Reading answered this question by impregnating eggshells with plastic and then dissolving away the shells, thereby obtaining microscopic casts of the pores. These casts have demonstrated that the pores are widely variable in shape, even within a single species. We also became interested in other aspects of respiration in the eggs of different species. For example, how does the gas conductance of the shell vary with egg size? Is there a relation between

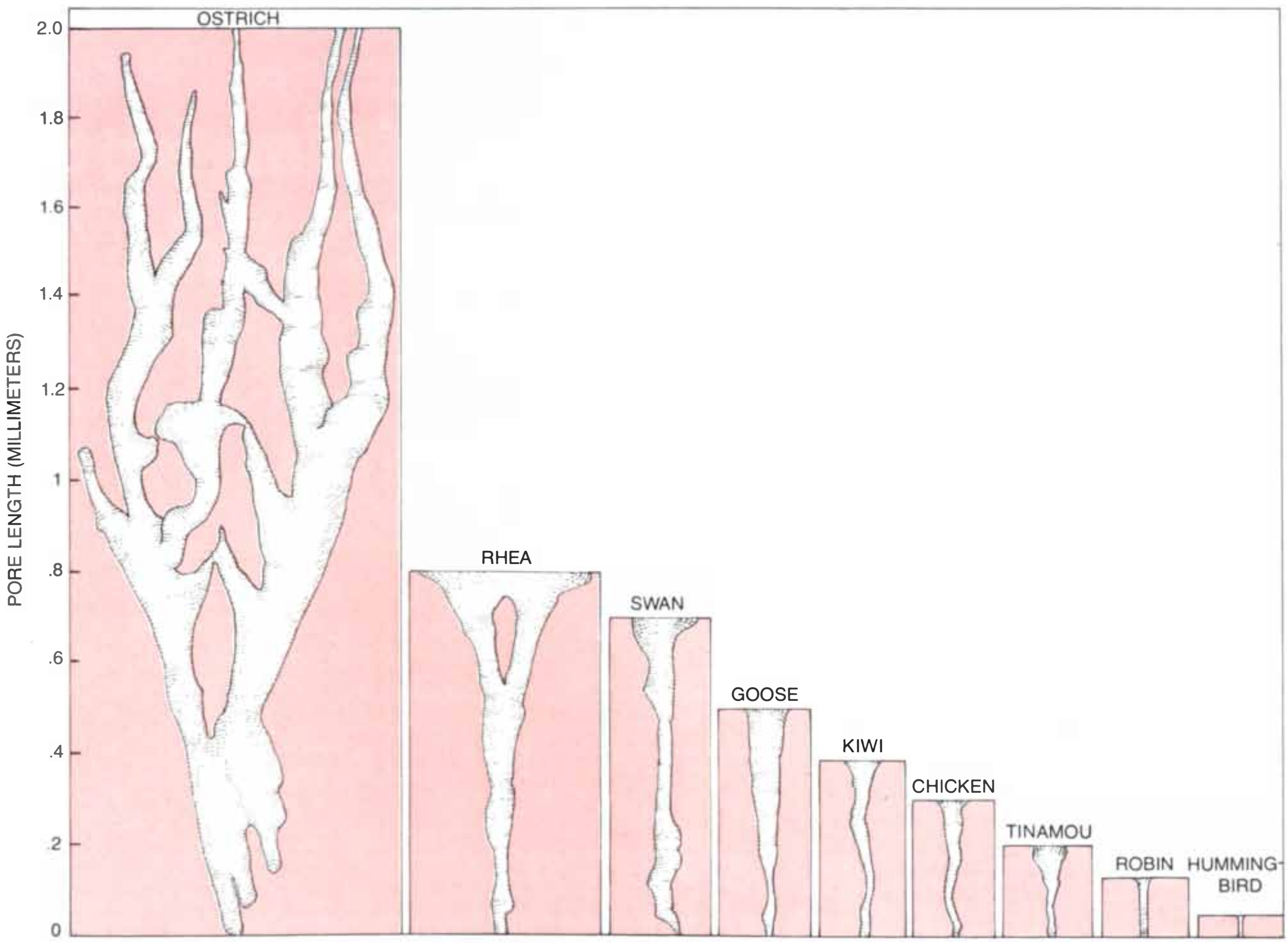




INTERFACE separating the inner shell membrane and the chorioallantois appears in this electron micrograph. At the top some of the fibers of the inner shell membrane are shown in section, coated with a fluffy material. The continuous line below the fibers is the thin film lining the inner membrane, to which the outer membrane of the chorioallantois is attached. The dark, irregular object in the center is a red blood cell shown in oblique section. Micrograph, the magnification of which is 18,400 diameters, was made by Ewald Weibel of University of Bern.



PROFILE OF GAS PRESSURES across the shell and the respiratory membranes is depicted for an 18-day-old chicken egg immediately prior to internal pipping. The various layers provide a series of barriers to diffusion that differ considerably in their conductance. A large change of gas pressures across a layer means that it has a low conductance; a small change means that the layer has a high conductance. The graphs indicate that both pore and film have a low conductance for oxygen; only the pore has a low conductance for carbon dioxide.



SHAPE OF SHELL PORES changes with increasing egg size. The length of the pores ranges from two millimeters in the ostrich egg to .04 millimeter in the hummingbird egg. Illustration is based on pore casts made by Cyril Tyler and K. Simkiss at University of Reading.

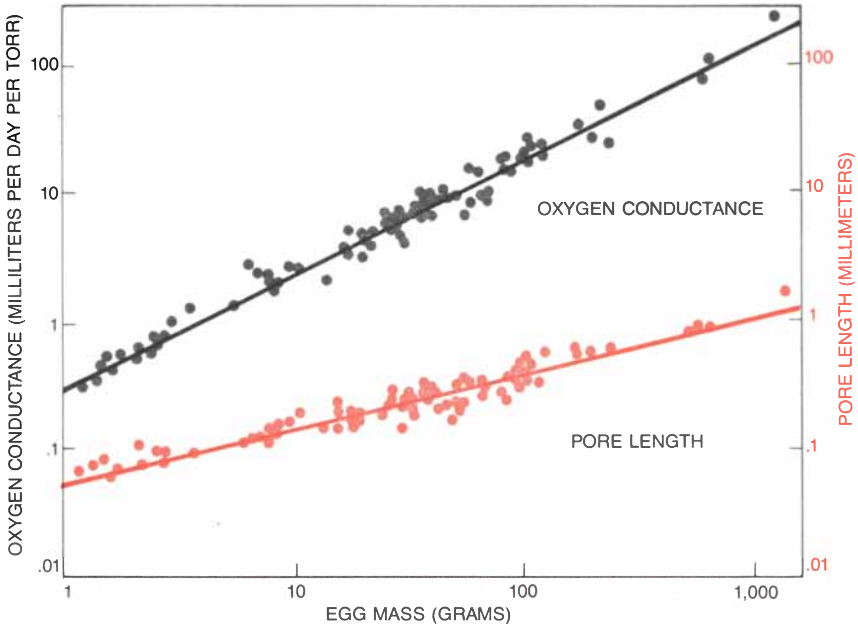
egg size and the oxygen concentration in the air cell at the end of incubation? How does oxygen consumption vary as a function of egg size? And finally, how is water loss related to egg size and the length of incubation?

Knowing that gas exchange across the eggshell proceeds entirely by diffusion, we first sought to develop a method to measure the gas conductance of intact eggs that was simple enough to use in field laboratories. The solution to this problem was provided by the laws of diffusion published in 1855 by Adolph Fick, who was professor of physiology at Zurich and later at Würzburg. Fick had the perceptiveness to realize that the mathematical description of heat flow in solids developed 30 years earlier by Jean-Baptiste Fourier could equally well apply to molecules diffusing through solids, liquids or gases.

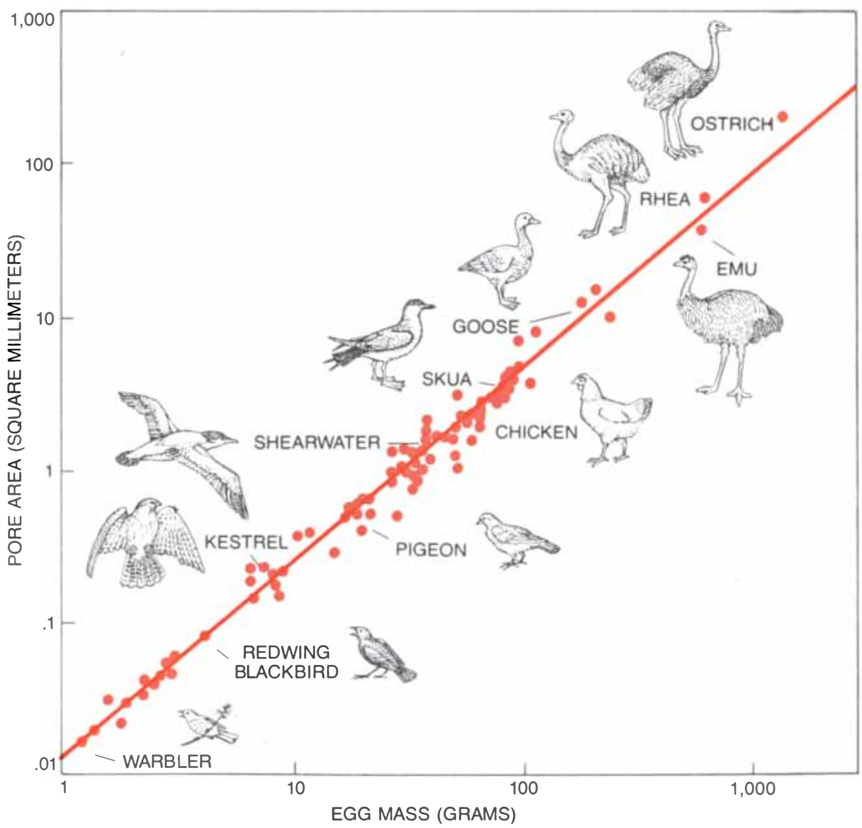
The movement of gas by diffusion through a permeable barrier depends on the random motion of the gas molecules and the difference between the concentration of the diffusing species on one side of the barrier and the concentration on the other side. Since collisions among gas molecules are more frequent in a concentrated gas than in a dilute one, the molecules will tend to move from the side of higher concentration to the side of lower concentration.

A simplified version of Fick's law of diffusion states that the quantity of a given gas diffusing in a unit of time through the pores of an eggshell will be directly proportional to the area of the pores available for diffusion and to the difference between the concentration of the diffusing gas at one end of the pore and the concentration at the other end. On the other hand, the rate of diffusion will be inversely proportional to the length of the diffusion path (in this case the length of the pores through the eggshell). In other words, the gas conductance of the eggshell depends on the ratio of pore area to pore length. Doubling the area available for diffusion of a gas or doubling the concentration difference of that gas across the shell will double the rate of passage, whereas doubling the pore length will halve the rate of passage, all other factors remaining equal. Thus if one could measure the flux of a gas and divide it by the concentration difference of that gas across the pores, one would be able to calculate the conductance of the shell for that gas.

After trying several methods for measuring shell conductance we finally hit on a very simple one. If eggs are kept in a desiccator at a constant temperature and are removed only briefly once a day for weighing, they lose weight at a rate that remains constant over many days. This weight loss results entirely from the diffusion of water vapor through the pores of the shell into the dry atmo-



PORE LENGTH AND OXYGEN CONDUCTANCE increase at different rates with increasing egg mass, as is shown in this graph encompassing data from the eggs of some 90 species from different parts of the world. For every tenfold increase in mass the oxygen conductance of the eggshell increases 6.5 times but the pore length increases only 2.7 times. Pore length probably increases slower because the eggshell must be thin enough for the embryo to hatch.



TOTAL PORE AREA, the sum of all the pores available for diffusion, increases almost 18 times with every tenfold increase in egg mass. For example, the fraction of total shell area occupied by pores jumps from .02 percent in the chicken egg to .2 percent in the ostrich egg. Rapid rise in pore area with increasing egg mass accounts for the fact that oxygen conductance (ratio of total pore area to pore length) increases 6.5 times for every tenfold increase in egg mass.

sphere of the desiccator. Dividing the daily weight loss by the water-vapor pressure between the inside of the egg and the environment yields the water-vapor conductance of the shell. Since the diffusion path through the shell for oxygen and carbon dioxide is the same as that for water, the water-vapor conductance can be converted into the oxygen conductance. Determining the oxygen conductance in this way is relatively easy: the only necessary tools are a desiccator, an accurate balance and a thermometer.

Adding calipers to this armamentarium enables one to measure the thickness of the shell, which is equivalent to the length of the pores. Knowing the conductance and the pore length, one can calculate from Fick's law the total functional pore area: the sum of all the individual pores available for the diffusion of gases. (In the chicken egg the 10,000 pores have an aggregate cross-sectional area of two square millimeters.)

Over the past seven years we and others have determined the conductance of hundreds of freshly collected eggs in various parts of the world. The eggs, ranging in mass from one gram to 1,500 grams, were obtained from 90 species and 15 orders of birds. We have consistently found that the gas conductance of eggs increases with the size of the egg. This trend was expected, because the oxygen demand of the embryo prior to internal pipping is greater in large eggs than it is in small ones. The rate of increase is not, however, directly proportional to the mass of the egg: with every tenfold increase in mass there is only a 6.5-fold increase in oxygen conductance.

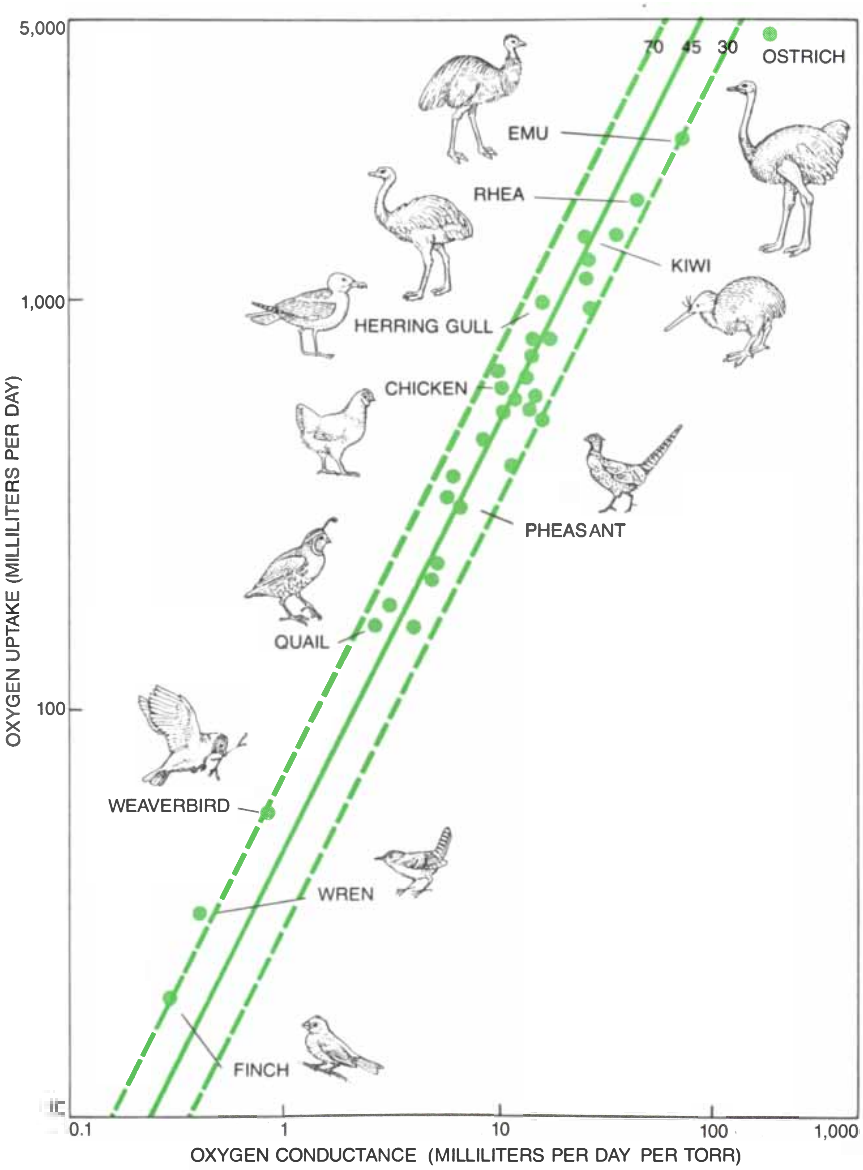
A second interesting observation is that with every tenfold increase in egg mass the length of the pores increases only 2.7 times. This trend can perhaps be explained in terms of a balance of adaptive factors operating in the course of evolution. The thickness of the shell (which determines the length of the pores) is established to withstand the gravitational stresses of the contents of the egg and the weight of the incubating parent, but the thickness is also limited by the requirement that the embryo be able to break through it during hatching. This last presumably explains why eggshells increase in thickness in less than direct proportion to the size of the egg.

What are the structural consequences of a gas conductance that increases 6.5-fold and a pore length that increases 2.7-fold with each tenfold increase in egg mass? Fick's law states that the total functional pore area is directly proportional to the product of the gas conductance and the pore length. The pore area will therefore increase about 18-fold (6.5×2.7) for each tenfold increase in egg mass. For example, a 600-gram rhea egg, with a shell about three times thicker than that of a 60-gram chicken egg, will have a total pore area that is about 18 times larger.

We can now appreciate that the structural elements of the shell (pore area and pore length) determine its gas conductance, a functional property that can be related to the embryo's metabolic requirements and the oxygen-pressure difference across the eggshell. Here again Fick's law gives us the quantitative relation among these variables: the oxygen-pressure difference across the shell is equal to the oxygen consumption of the egg divided by the oxygen conductance of the shell.

Let us take as an example a hypothetical egg whose oxygen conductance is 10 milliliters per day per torr (that is, 10 milliliters of oxygen diffuse across the shell each day for each torr of oxygen-pressure difference). If the egg consumes 500 milliliters of oxygen per day, the oxygen-pressure difference across the shell must be 50 torr. Since the partial pressure of oxygen in the atmosphere is about 150 torr, the partial pressure of oxygen in the air cell of the egg (to which the blood of the embryo is exposed) will be 100 torr, or about 14 percent.

In order to explore the relation between shell conductance and the metabolism of the embryo, we compared the oxygen consumption of the egg just before internal pipping with the oxygen

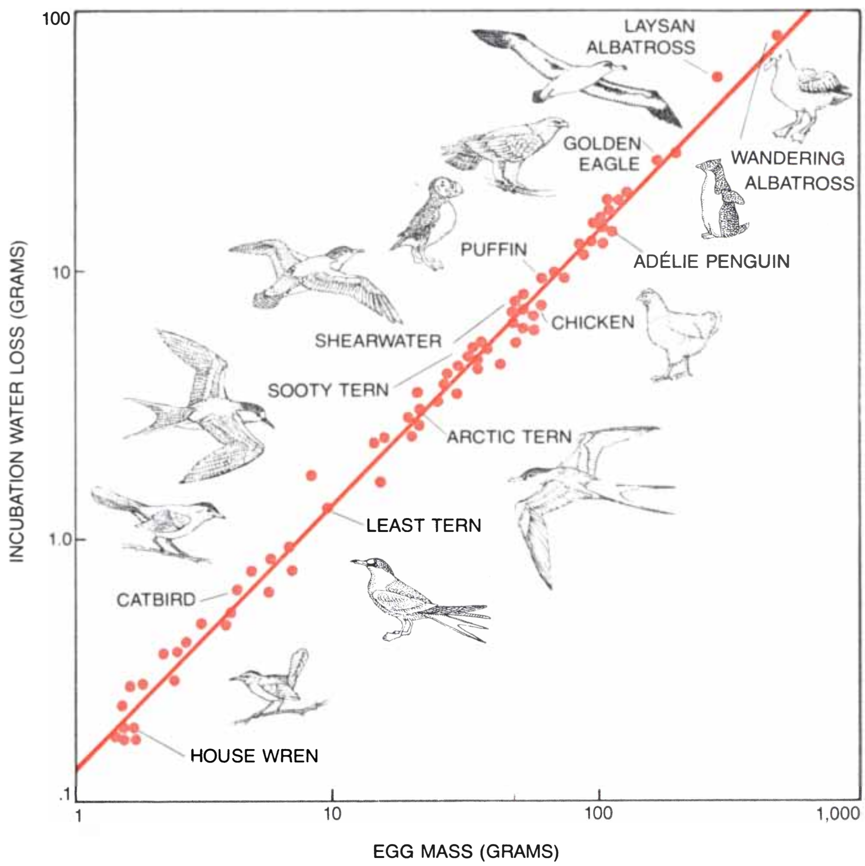


DIFFERENCE IN OXYGEN PRESSURE between the air cell and the atmosphere is nearly the same in eggs of different sizes and shapes. The difference in oxygen pressure was calculated from this graph, which for 28 species plots the oxygen consumption of the embryo before internal pipping against the oxygen conductance of the shell. The slopes of the broken lines represent oxygen-pressure differences of respectively 70 and 30 torr. Most of the experimental points fall within these boundaries; the average value is 45 torr. The pressure of oxygen in the atmosphere is 150 torr, so that oxygen pressure in air cell before internal pipping is 105 torr.

conductance of the eggshell in 28 species [*see illustration on opposite page*]. The list of known values for the oxygen consumption of the eggs of many species, large and small, was greatly expanded by the work of three investigators at the University of California at Los Angeles, Donald F. Hoyt, David Vleck and Carol Vleck. Knowing the oxygen consumption of the eggs and the conductance of the eggshells, we could then utilize Fick's law to calculate the oxygen pressure in the air cell. We found that the value was approximately the same for all 28 species: about 105 torr of oxygen, or 15 percent. This value was verified by directly sampling the air cell in 13 species. In addition the measured value of carbon dioxide in the air cell averaged 35 torr, or about 5 percent.

These values of oxygen and carbon dioxide in the air cell just prior to internal pipping are nearly identical with those found in the lungs of adult birds. Therefore the oxygen conductance of the eggshell appears to be matched to the egg's uptake of oxygen to "anticipate" the oxygen demand of the embryo prior to internal pipping and to provide in the air cell the oxygen and carbon dioxide pressures characteristic of the adult bird. A. H. J. Visschedijk of the University of Utrecht has suggested that these gas pressures initiate the final act of hatching and prepare the embryo for its postnatal existence.

It is remarkable that the gas conductance of the eggshell is calibrated to the mass of the embryo to yield nearly the same final concentrations of oxygen and carbon dioxide in eggs of different sizes. The general plan for shell porosity and function is even more impressive if one considers that the incubation period of bird eggs may vary from 11 days in some of the smaller species to 70 or more for the eggs of the tube-nosed birds, such as the wandering albatross.



LOSS OF WATER during incubation is independent of the metabolic rate of the embryo, yet it appears to be essential for successful hatching. Here the total amount of water lost during incubation has been plotted against the initial mass of the egg. The graph includes data obtained from 65 species of eggs ranging in size from one gram to 500 grams, with incubation times ranging from 11 to 70 days. There is a remarkably consistent trend: regardless of egg mass or incubation time, the typical egg will lose 15 percent of initial mass during natural incubation.

So far we have not considered in detail the loss of water from the egg, which proceeds at a constant rate throughout incubation. Water loss in any given egg is independent of metabolic rate and yet appears to be necessary for successful hatching. The reason for this is not fully understood but may concern the state of hydration of the embryo.

Rudolf Drent of the University of Groningen was the first to point out that the daily rate of water loss during natural incubation is related to egg mass. Surveying data in the biological literature, he demonstrated that in 45 species of birds the daily rate of water loss increases 5.6 times for every tenfold increase in egg mass. We subsequently noted that an even more precise relation between water loss and egg mass could be derived if we calculated the total amount of water lost during the incubation period and plotted this value against the initial mass of the egg. Our survey covered the eggs of some 65 species of birds, ranging in mass from 1.5 to 500 grams and in incubation time from 11 to 70 days. The results showed a remarkably consistent trend: regardless of mass or incubation time, a typical egg will lose 15 percent of its initial mass during natural incubation. This loss gives rise to the air-cell volume, whose functional importance to the embryo we have already discussed.

The gas spaces within the egg are essentially saturated with water vapor, so that at the average incubation temperature for most species (35.6 degrees Celsius) the vapor pressure within the egg is 44 torr. For the egg to lose the requisite amount of water the vapor pressure in the microclimate of the typical bird's nest must be maintained at about 15 torr, equivalent to a relative humidity of 45 percent. That humidity level can be maintained only by periodically ventilating the nest with drier atmospheric air. It is not known how the parent bird senses the humidity of the nest, but through its behavior it is able to regulate the passage of water vapor from the nest into the drier ambient air. The parent therefore has two major functions during incubation: it not only warms the eggs to an optimal temperature but also maintains the humidity of the nest air within close tolerances.

The length of embryonic development, or incubation time, is probably determined by genetic factors, but the metabolic rate of the embryo must be tuned to this incubation time so that by the end of the period—long or short—the embryo is mature enough to hatch and survive. One might therefore predict the intriguing possibility that the exact geometry of the eggshell's pores (and hence the shell's oxygen conductance) is calibrated both to the egg's mass and to its incubation time.

It seems clear that the eggshell and its associated membranes are closely adapted to the respiratory needs of the embryo, although certain compromises have been made in the course of evolution to ensure the mechanical stability of the egg and the fluid and acid-base balances of the embryo. Our studies have also demonstrated that the bird egg provides an ideal model for studying the diffusion of gases, which may eventually help in understanding the more complicated diffusion processes in the airways and air sacs of the human lung.

Lake-Bottom Archaeology

Mountain lakes from France to Austria and from Germany to Italy attracted many prehistoric settlers. Underwater archaeologists are now investigating sites that once were only dredged by collectors

by Aimé Bocquet

One of the most familiar pictures of how people lived in prehistoric times is that of European lake dwellers: Neolithic and Bronze Age farmers who lived in houses built on pilings over the water. This picture is false, and has been known to be so for decades. The evidence is that these villagers often built their houses at the side of a lake but rarely, if ever, built them over the water. The reason it appeared they had done so is that since that time the level of the lakes has risen, submerging the remains of the villages and leaving them some distance from the shore.

Archaeologists may regret the perpetuation of the lake-dweller myth, but they can be grateful for the numerous submergences that gave rise to it. Drowning has preserved many of the remains that would long ago have been destroyed by natural causes at any dry-land site. For example, wood objects and textiles that would weather and decay at dry-land sites can be almost perfectly preserved in wetland and underwater locations. Equally important, the same is true of both wild and domesticated plant materials. Only a decade ago few investigators paid much attention to the clues such materials offer to the varied environments of prehistoric times. Today, however, archaeologists set great store by such information. It can reveal the extent of prehistoric man's environmental adaptation, and the more (or less) efficient prehistoric man's adaptation to his environment was, the more (or less) assured was his survival.

A knowledge of prehistoric environments is also a prerequisite for assessing man's technical progress and his influences on the world around him. For example, a forest environment ought to inspire the invention of the woodsman's ax. When an archaeologist excavates an early forest dweller's shelter, it is good to find the axes the woodsman wielded. It is even better, however, to discover what kind of forest the woodsman inhabited by analyzing the plant remains. It is better still if the analysis reveals the woodsman's impact on his environment, for example his selective cutting of certain species of trees. Studies of this kind are facilitated by the preservation of plant materials at wetland and submerged sites.

The eastern reaches of France, particularly those areas adjacent to the northern face of the Alps, are abundant in lakes and lakeshores that provided village sites from as long ago as Neolithic times up to the Middle Ages. Most of these villages were later drowned; many were first discovered more than a century ago during an exceptionally severe drought. In 1853 and 1854 the level of lakes fell throughout Europe. At Lake Geneva, for example, many long-submerged shore dwellings on the French side of the lake were exposed to view. Today the number of such sites discovered at the mountain lakes of France has grown to a total of 38. Similar drowned villages were exposed in western and central Switzerland, southwestern Germany, Austria and northern Italy.

In the years between the mid-1850's and 1935 rich collections of "lake dweller" artifacts were assembled. They are now to be found in private collections and in museums throughout Europe, but with a few rare exceptions the original collecting was far from scientific. The practice was merely to drag up to the surface various objects that were scattered on the lake bottom or were shallowly buried in the bottom silt. The collectors often worked with ingenious dredging tools in order to reap this special harvest.

Much of the work was done at such Swiss sites as Auvernier and Cortaillod on Lake Neuchâtel and Meilen and Wollishofen on Lake Zurich; the French contribution was mainly from Lake Bourget in the Savoy, halfway between Grenoble and Geneva. As if to foreshadow a more scholarly approach, in 1929 the drowned site of Sipplingen, on the German shore of Lake Constance, was pumped dry a little at a time, with the aid of caissons, thereby allowing the methods of dry-land archaeology to be applied during excavation.

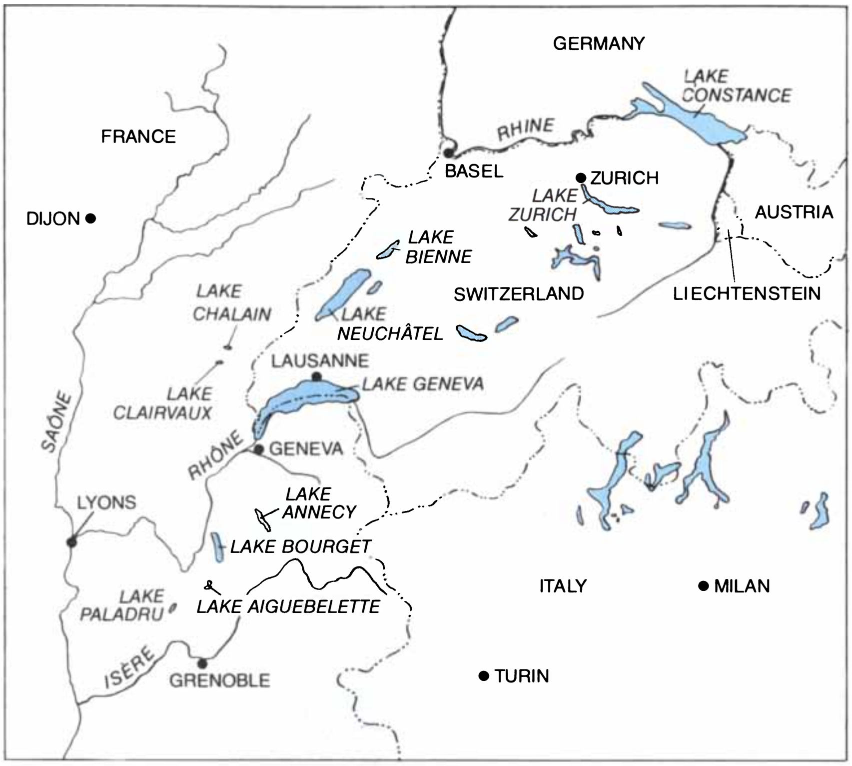
With the invention of practical free-diving gear in the 1940's an entirely new means of underwater investigation became available to archaeologists. A number of amateur divers began to collect "lake dweller" antiquities with the aid of the new apparatus. A. Favre, who worked at Lake Annecy, a little to the northeast of Lake Bourget, was one such early worker. These divers' work was not, strictly speaking, excavation, but it was a considerable advance over the earlier dragging and dredging.

True lake-bottom archaeology began in 1952, and its pioneer was Raymond Laurent of the Centre de recherches archéologiques lacustres du Dauphiné-Savoie. Influenced by the teachings of an eminent French prehistorian, André Leroi-Gourhan of the Collège de France, Laurent brought to his underwater work a comprehensive grasp of how to organize a lake-bottom site for the precise recording of the horizontal and vertical distribution of artifacts. He too worked at Lake Bourget and Lake Annecy and he also worked at Lake Aiguebelette, a little to the southwest of Lake Bourget; his underwater excavations between 1960 and 1969 paved the way for today's fully disciplined lake-bottom work. By 1963 Ulrich Ruoff was conducting equally sophisticated investigations in Switzerland at the Lake Zurich site of Kleiner Hafner. At the same

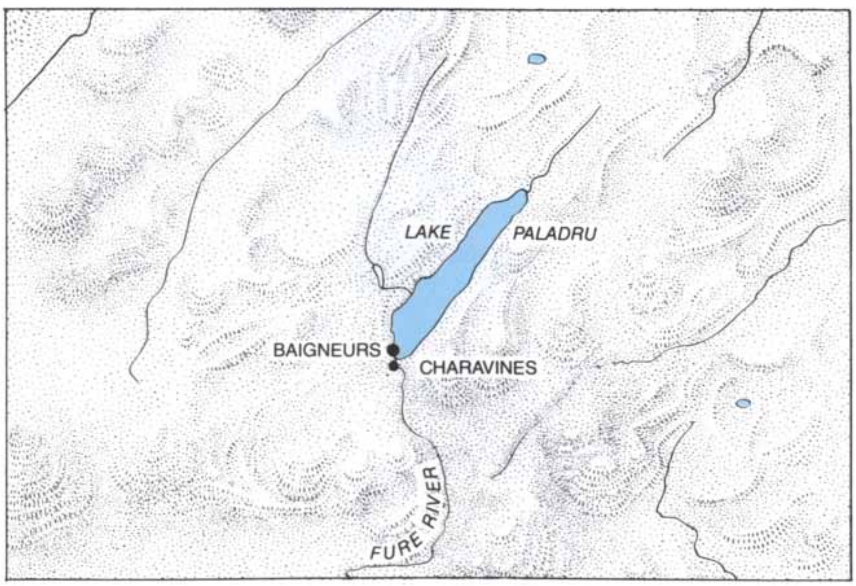
UNDERWATER EXCAVATOR at Baigneurs, a drowned Neolithic village in eastern France, is seen at work in the photograph on the opposite page near one apex of a triangular duralumin frame. The equilateral triangle has sides five meters long. Red and white ribbons subdivide the area into 25 smaller equilateral triangles, each enclosing an area of 430 square centimeters. In front of the excavator is what remains of an upright timber that was once a corner post in a Neolithic farmhouse. To the excavator's left is the collecting bucket that will carry to the surface the house-floor debris he has gathered after noting its horizontal and vertical coordinates.

time work of a more traditional kind continued. For example, between 1969 and 1972 a number of sites in the Auvernier Bay section of Lake Neuchâtel were excavated by divers, and other bay sites were diked and pumped dry to allow conventional excavation.

It is my intention here to describe lake-bottom archaeology as it stands today, taking as my example a Neolithic lake-bottom site in the northern Dauphiné, southwest of the Savoy. The site was singled out for salvage archaeology in 1971. The lake whose shores once attracted Neolithic settlers is called Paladru; it is a little less than 45 kilometers northwest of Grenoble on the route to Lyons, and the specific drowned shoreline site we investigated, some two to three meters underwater, is known as Baigneurs. The lake-bottom strata in which the Neolithic remains are found were deposited some 5,000 years ago. They are thin and readily broken up, so that they call for the most painstaking excavation.



SIX-NATION SECTOR of alpine and subalpine Europe contains the many large and small lakes that were selected by Neolithic and Bronze Age farmers for lakeshore settlements. A rise in lake levels concealed the deserted villages until a drought in the mid-19th century exposed them, bringing about the myth of the Alpine lake dwellers. The lakes named here contain many of the principal Neolithic and Bronze Age sites; only a few have been scientifically excavated.



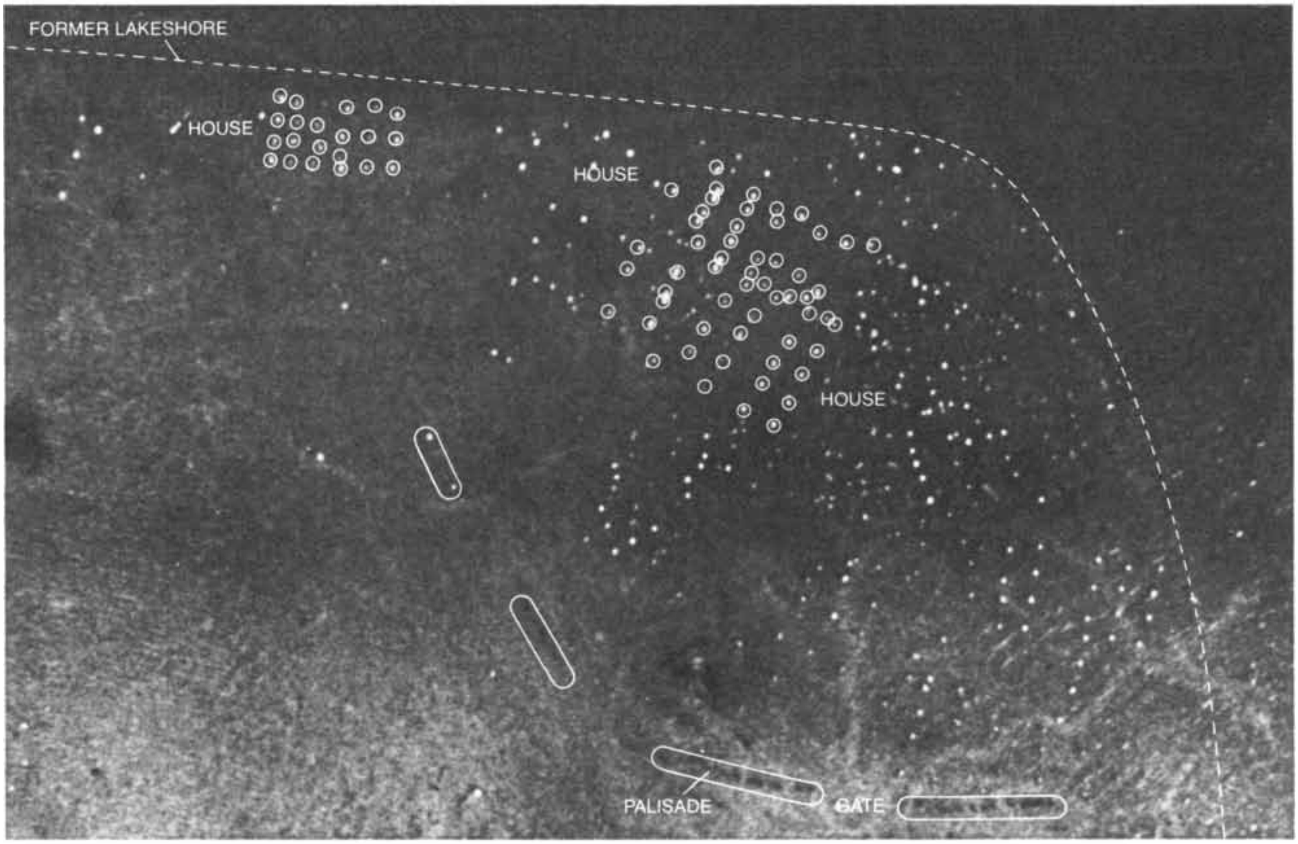
LAKE PALADRU, in the Dauphiné southeast of Lyons, gives rise to a tributary of the Isère River. Near Charavines at the outlet of the lake is the drowned Neolithic settlement of Baigneurs. The author and his colleagues have undertaken rescue archaeology there since 1972 while at the same time conducting a workshop in the techniques of underwater archaeology.

Our salvage operation at Baigneurs was based on the Laurent system of lake-bottom topographic analysis, the key requirement of which is that the location of all finds should be recorded quite precisely. Also implicit in the Laurent system is that all the artifacts in each lake-bottom stratum, no matter how small or fragile, are salvaged. Our own contribution to the Laurent system was to make sure that literally all the objects—not just the man-made ones—were collected. Ever since my colleagues (Françoise Ballet, Patrick Grandjean, Christian Orcel and Alain Cura) and I began our work some seven years ago that has been a prime objective. Only this kind of collecting makes it possible to reconstruct the environmental shifts of Neolithic times.

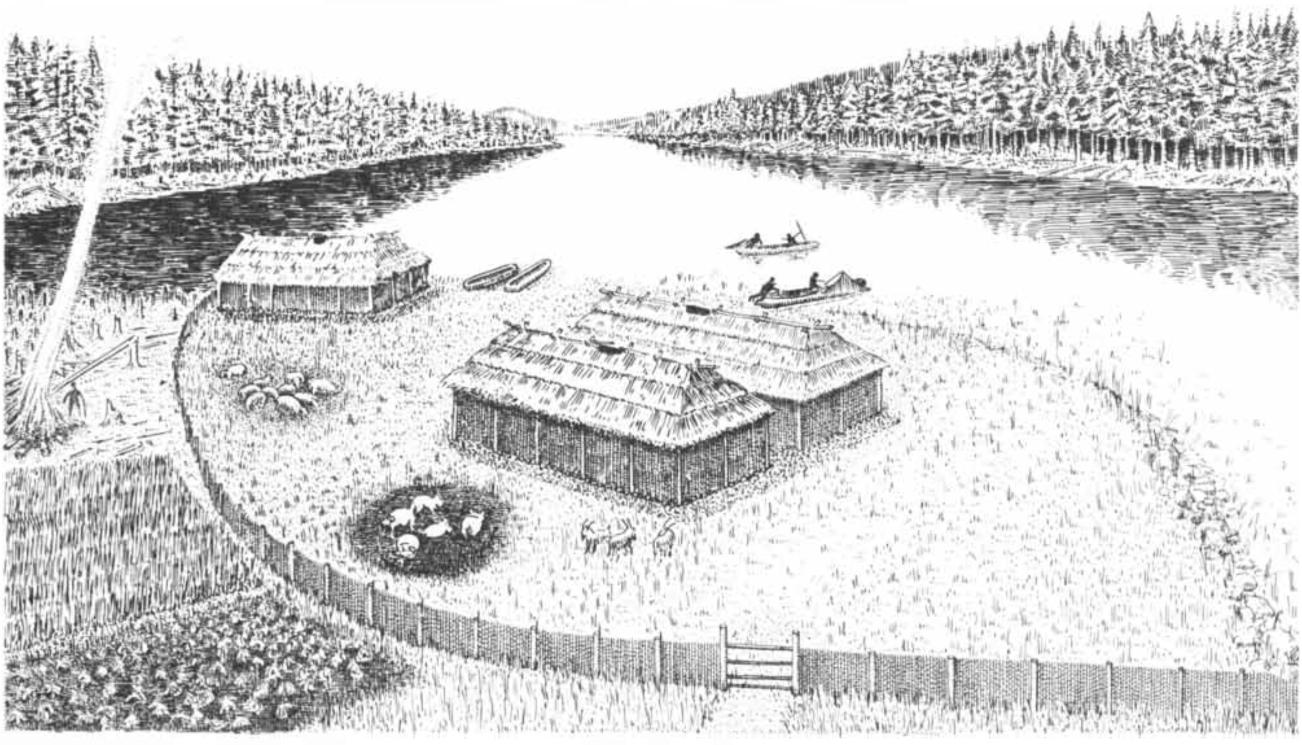
The first problem confronting any lake-bottom archaeologist is limitations in visibility. Even without taking into account the clouds of particles that rise from the mere act of digging in the lake bottom, the range of the diver's vision can be only a few centimeters, depending on the surface winds, the lake currents and particularly the season. During much of the year lake waters are naturally clouded with seasonal growths of algae and plankton.

That is one reason the grid used to control dry-land excavations (a checkerboard of one-meter squares) is virtually useless underwater. It is difficult enough under conditions of limited visibility to lay out such a grid with geometric precision, and it is even more difficult to measure the actual location of an object within a particular square. One must run two perpendiculars from the object to two adjacent sides of the square and record both the precise intersection point and the precise length of each perpendicular.

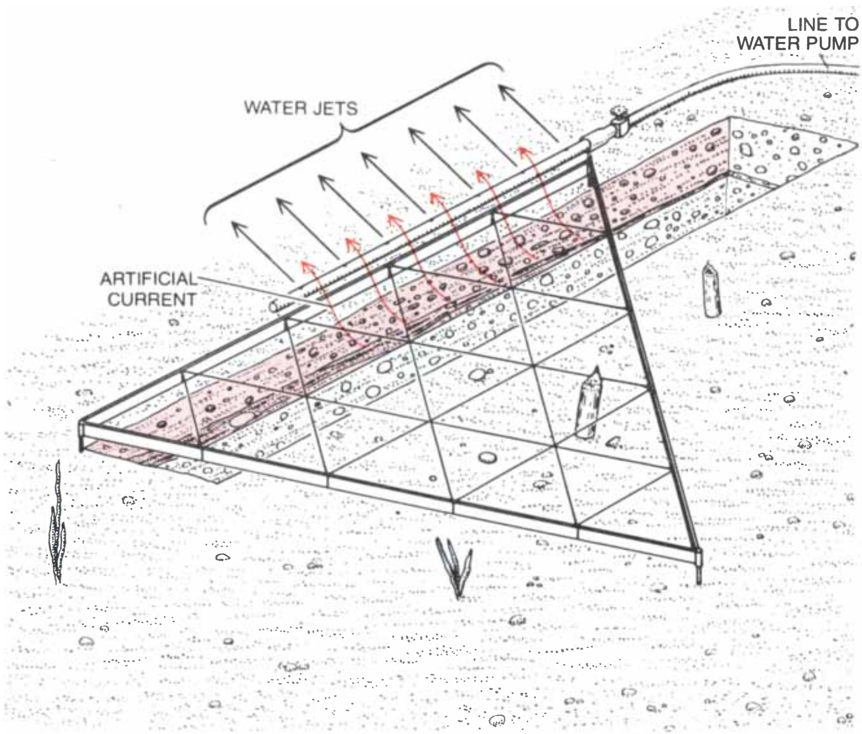
Laurent's solution for the grid problem was simple and inspired: instead of squares he used equilateral triangles. To locate an object found within one of the triangles in such a grid it is necessary only to measure the distance to the object from all three of the triangle's apexes. This is done by means of tapes attached to each apex in the grid. To build the grid itself one begins with a single known point and three small girders of equal length. At each apex the girders are interlocked by a collar that fits onto a vertical tube sunk into the lake bot-



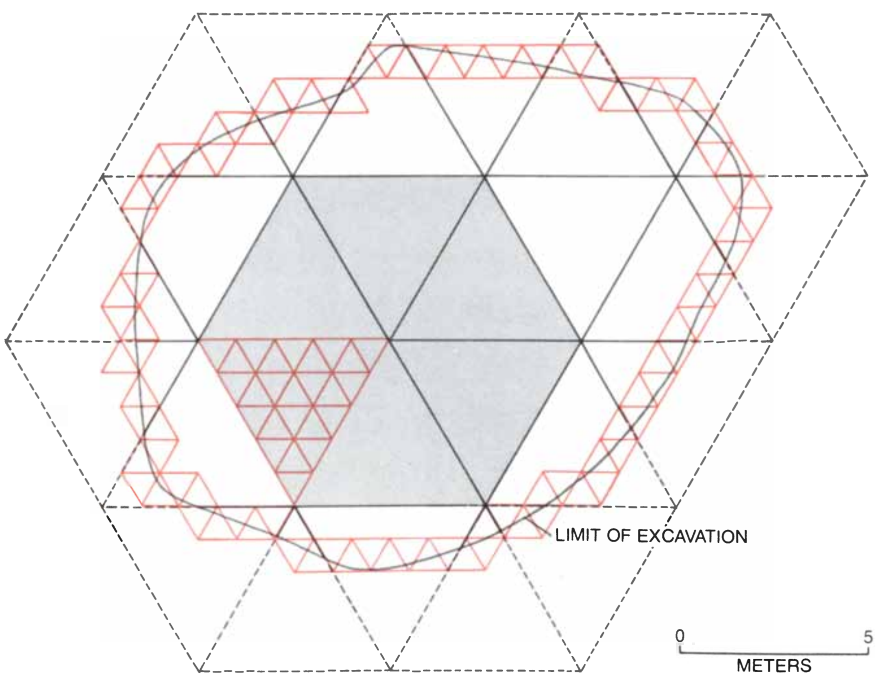
LOW-LEVEL AERIAL PHOTOGRAPH of Lake Paladru has been annotated by the excavators to indicate the extent of the drowned Neolithic village. In preparation for the photograph the divers set white metal plates on top of each upright house timber they had located on the lake bottom and set red plates on the smaller timber uprights that formed a palisade on the inland perimeter of the village. Annotation (*white*) locates one large house (*circled uprights*) at the northwest corner of the village near the ancient lakeshore (*broken line*) and two other houses nearby. The palisade and its gate are also annotated. Additional uprights and debris are still being investigated.



RECONSTRUCTION suggests the probable appearance of the first settlement at Baigneurs after the third farmhouse was built there. The palisade was evidently not defensive but was a means of confining the village livestock when the animals were not herded into nearby meadow and forest to forage. Fishing on the lake was a regular activity, as is shown by the recovery of net weights and a fishhook from the ruins.



"WATER CURTAIN," a device developed by the Swiss diver Ulrich Ruoff in 1963, is used at Baigneurs to improve lake-bottom visibility. In the diagram the water-curtain tube has been secured along one side of a five-meter triangle; the lake bottom covered by the six small triangles nearest the tube (*color*) is being excavated. A submerged pump supplies 60 cubic meters of water per hour to the tube; the water emerges through a series of bullet-size holes spaced 20 centimeters apart. The flow of water (*black arrows*) creates a local current (*colored arrows*) that serves to carry away the particles of silt that cloud the water in the diver's work area.



UNDERWATER GRID, used to determine the exact position of excavated materials, consists of interlocking equilateral triangles. The diagram shows the part of the site excavated so far (*gray line*) and the inner set of six triangles that the divers were able to bring together to form a hexagon with a 30-meter circumference (*gray tone*). The cumulative error in closing the hexagon was only seven centimeters. Additional triangles, erected on the core of the central six, embrace the entire excavation area; all but one of them extend beyond its perimeter. The subdivision of a five-meter triangle into 25 smaller triangles (*color*) is shown here in one location.

tom. Once the first triangle has been set up the grid can be extended indefinitely by the addition of girders that are interlocked in the same way. The accuracy of the system is well within the necessary limits. Using five-meter duralumin girders to form successive triangles we have been able to "close" a 30-meter hexagon made up of six contiguous triangles with a cumulative error of only seven centimeters.

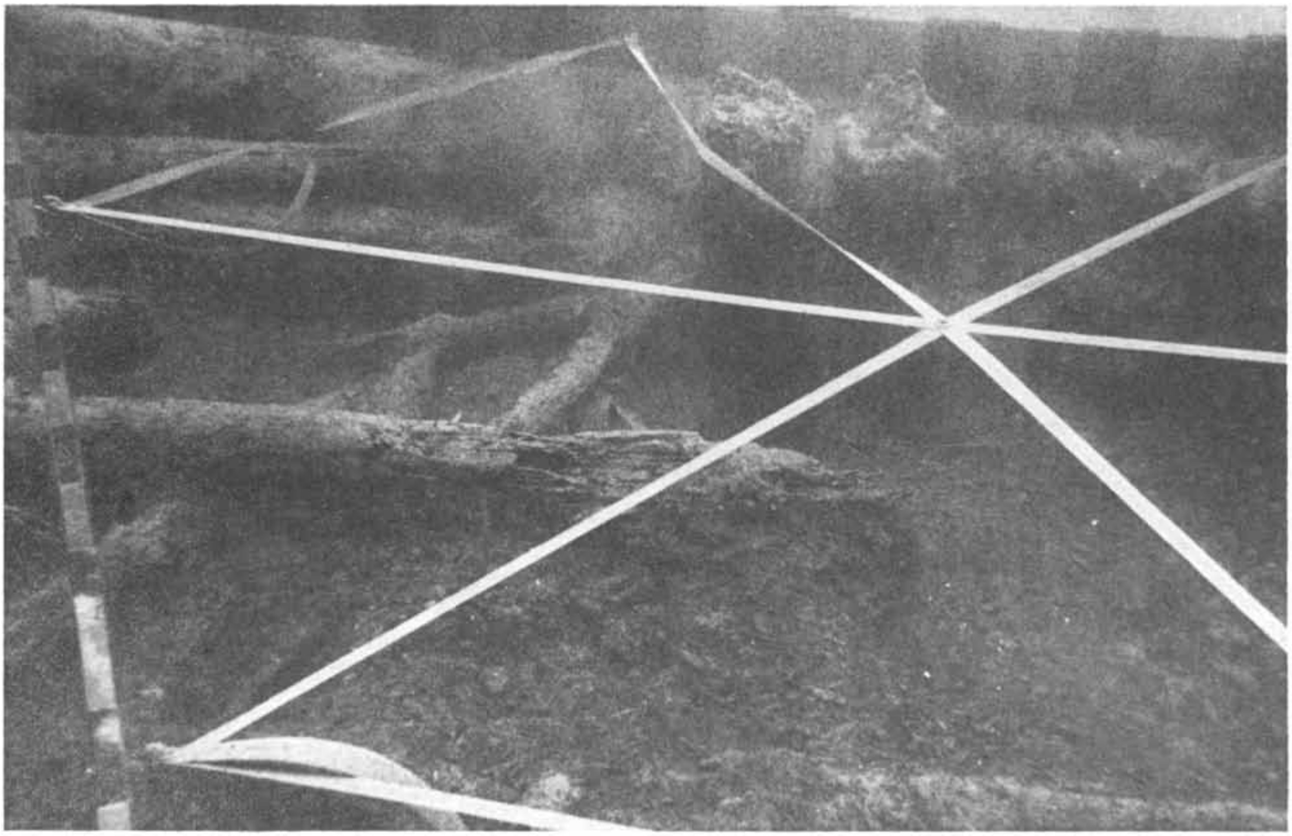
Although the interlocking metal triangles provide an excellent reference grid for both horizontal and vertical measurements, the area within each triangle (nearly 11 square meters) is much too large for a single mapping unit. We therefore use ribbon or string to subdivide each large triangle into 25 small triangles that measure one meter on a side. The area within each small triangle is much more manageable: it is less than half the area of the square-meter unit of the conventional dry-land grid.

Once the reference points are established underwater, excavation can begin. The divers may, depending on seasonal factors, be handicapped by the naturally poor visibility, but they no longer need to work in a cloud of silt of their own making. The reason is that an apparatus developed by Ruoff in the early 1960's establishes artificial water currents to carry away the fine particles that are stirred up by the digging, much as a breeze carries away smoke. We call the current created by the Ruoff apparatus a water curtain.

The excavators working on the lake bottom bear the primary collecting responsibility, but those working on shore to process the sediments collected by the divers make their own vital contribution. Their sieving extracts from the sediments the largest possible quantity of material for analysis. It is not for nothing that we refer to our kind of archaeology as two-story excavation.

The diver's work is far more demanding than might be supposed by someone who has dived recreationally in clear, warm water. Even dry-land excavation calls for dexterity and an ability to recognize the changes in soil texture and color that mark individual strata. Exercising these same capacities when one is submerged in cold, murky water is possible for only limited periods of time. Just as one's body grows increasingly numb, so do one's wits lose their quickness. To make even a small mistake is to risk the loss of irreplaceable information. The ideal underwater excavator combines physical and mental endurance with technical competence and what might be considered a moral qualification: scientific ethics.

The top layer of lake-bottom sediment at Baigneurs, from 10 centimeters to a meter and a half thick, contains nothing of archaeological interest. This "sterile" sediment, however, has helped



CLUSTER OF TIMBERS, some upright and some flat on the lake bottom, appears in this underwater photograph. The horizontal and vertical positions of each timber will be recorded with reference to the small triangles of ribbon that subdivide the corner of the larger triangle partly visible here. Analysis of the wood still protected by bark at Baigneurs reveals the time of year when the trees were felled.

to preserve the archaeological materials lying under it. Before beginning the painstaking work of excavation the diver removes the overlying sediment from his section of the bottom with a suction pump. Then, using only his bare hands, he gently lifts from the exposed surface one thin horizontal section of material after another. Only in this way can he feel any pieces of wood, bone, cloth or pottery the uncompacted debris may contain and take steps to protect them against destruction.

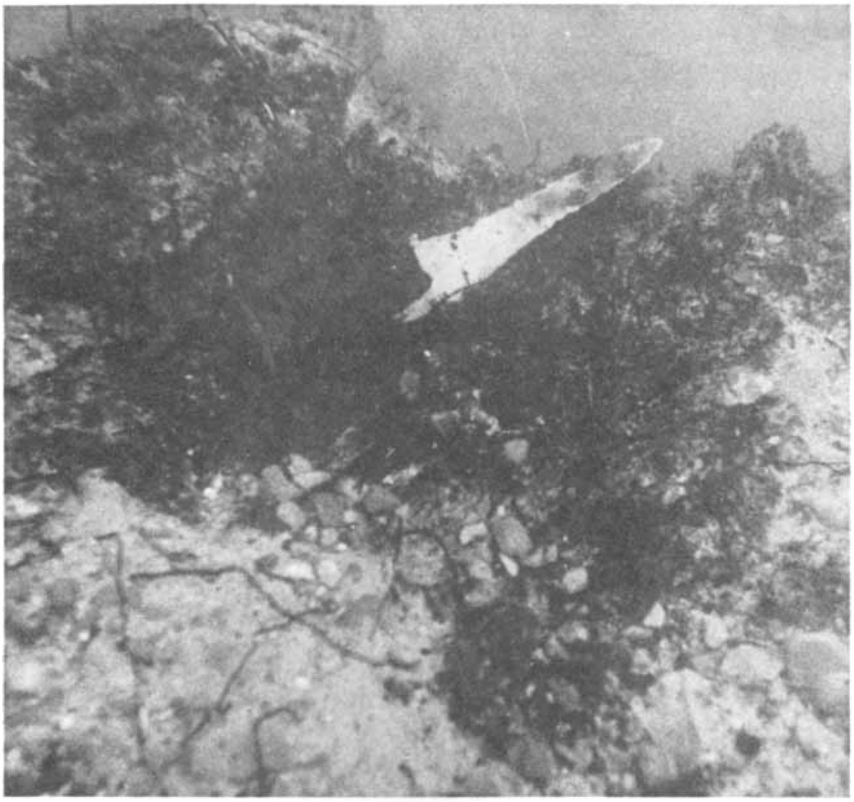
If the diver's handling indicates that a horizontal section holds no sizable artifacts, he places the material in a bucket for transfer to the sievers on land. If he touches an artifact, he carefully uncovers it and makes a sketch of it in situ before removing it from the section. If the water is sufficiently clear and there is enough light, he also makes a photograph of the object in situ. Of the objects we uncovered at Baigneurs we recorded in this manner most of the pottery, all the flint tools except the smallest ones, all the wood objects (such as spoons, combs, ax handles and small planks), animal remains (such as deer antlers) and two flint daggers with their wood handles still in place.

When even more fragile materials, such as textiles, were uncovered, the divers were able to avoid disturbing them further by cutting out an entire block of the bottom and raising the block with the material to the surface in one piece. We used the same technique on a larger scale, removing blocks of lake bottom weighing from 10 to 100 kilograms, when we wanted to preserve entire stratigraphic sections or to obtain quantities of material for paleobotanical analysis. Our excavators also took from the lake bottom numerous deep cylindrical "cores" to get samples of pollen for analysis of the ancient plant communities and samples of mollusk shell for analysis of the prevailing ancient temperatures. These cores were obtained from a float by plunging into the bottom a plastic tube eight centimeters in diameter, from two to three meters long and heavily weighted at the upper end.

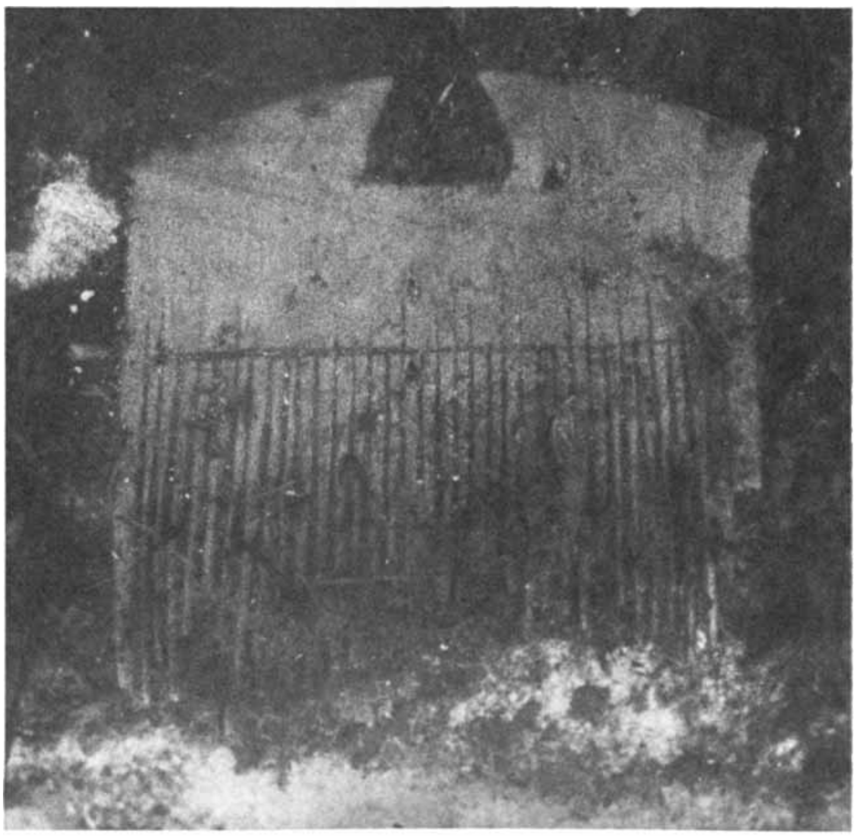
The upper half of our two-story excavation differs in several ways from dry-land archaeology. The kind of sieving we do is conventional, but many of the objects our sievers retrieve are organic, for example wood artifacts and plant materials such as seeds. Submersion has preserved these objects; to let them dry out would be to destroy them. After the objects are numbered, weighed and recorded on a chart they must be sealed in plastic bags to keep them wet. Even fragments of pottery, which are normally indestructible, may need to be bathed in polyvinyl acetate to keep them from crumbling when they dry out.

When samples of wood are present in an archaeological deposit, as they are at Baigneurs, there is the possibility of establishing the deposit's absolute age. In recent years much progress has been made in correlating actual calendar dates with dates determined by sequences of tree rings and the decay of carbon 14. In the New World the tree-ring year count has now been reliably extended back beyond 5000 B.C. In the Old World the best such record, the Chronology of Treves, goes back only to about 800 B.C. Tree-ring sequences also exist for earlier periods, such as the Neolithic, but they are not yet linked up with the more recent chronologies. Hence in general the only estimates of age for these "floating" chronologies are carbon-14 ones.

If several floating counts can be correlated, however, the absolute chronological differences between them can be expressed in actual solar years rather than carbon-14 years B.P. (before the present). This kind of work is in progress at the Dendrochronology Laboratory of Neuchâtel under the direction of my colleagues Orcel and Lambert. The tree-ring data from Baigneurs wood are now



FLINT DAGGER, deftly pressure-flaked and fitted with a handle made of wood, was one of two such wood-hafted artifacts found in the compacted plant litter that covered house floors.



WOOD LOOM COMB, one of several artifacts indicating that the farmers of Baigneurs made their own cloth, is a further example of how objects that might soon have disintegrated at an open-air site may be preserved in wetland and underwater locations. Once uncovered, however, wood objects must be kept wet until laboratory techniques for preservation are applied.

being integrated into the expanding tree-ring chronology for western Switzerland, even though the work faces such obstacles as the fact that much of the Swiss material is based on samples of oak and ours is based on samples of fir. The overall result is that we can talk about the Baigneurs settlements with a chronological precision that would have been impossible a few decades ago.

Among the samples of wood at Baigneurs are house posts and beams with the bark still on them, which makes possible even greater precision. The sequence of tree rings establishes the year a tree was felled. Bark, when it is present, protects the latest ring, a fragile outer growth. Viewed under the microscope, this developing growth provides evidence on whether the tree was felled in winter or summer, spring or fall. Ideally all the house timbers that are more than seven or eight centimeters in diameter should be sampled for their ring sequence and all the sequences should be interrelated. We have done this since 1974, and by the end of the 1978 season our 150 square meters of excavation had yielded the material for 810 individual tree-ring samples.

In addition to the contribution to dating made by these ancient posts and beams other plant remains provide clues to the life of the Neolithic farmers who cleared the forest here. They felled the fir and ash trees but left the oaks standing, presumably because they valued the harvest of acorns. They also protected other nut trees, clearing away the nearby growth so that the sun could reach them. As a result the nuts that were harvested from the protected trees were twice the size of those of untended neighbors.

Our intensive collection of plant remains has also enabled us to demonstrate different uses of different materials. For example, the inhabitants made beds of fir boughs and filled the chinks in their house walls with moss. We can even trace the lines of the house walls by the accumulation of hazelnut shells along the inside of the wall. Analysis of the charcoal from hearths indicates that for firewood the settlers preferred beech and oak.

The Neolithic inhabitants of Baigneurs selected as their building site a headland near the outlet of Lake Paladru. The settlement occupied an area of some 1,500 square meters running down to the lakeshore. On the land side the village was enclosed by a small palisade fence that included a gateway. A virtual forest of posts at the center of the enclosure reveals a pattern of adjacent small rectangular houses aligned along narrow alleys.

The village site was occupied not once but twice. The first occupation level is separated from the second by a layer of silt that was deposited during a temporary rise in the level of the lake. The first settlers arrived on the scene in about 2900 B.C. They were farmers who had apparently lived in a nearby village; the grassy headland with its thick forest of fir trees adjacent evidently caught their fancy. They came in winter, and in preparation for their return they felled several small firs with trunks ranging from 10 to 14 centimeters in diameter. These they stripped of their branches and stacked.

The visitors came back the following winter. They chopped down more fir trees and built their first house, using these trees and the ones they had felled the preceding year. The house, 12 meters long and four meters wide, was oriented with its long axis running east-west. It consisted of a single room with a clay hearth in the center and a door in the north wall. The fir-tree trunks formed the frame of the house; they were set in holes from three to four meters deep. Trunks smaller in diameter formed the horizontal beams; they were lashed to the top of the uprights with ropes and vines. The roof covering this sturdy frame was probably thatched with reeds; the walls were made of slender vertical poles, from two to three centimeters in diameter, their butts set in a shallow trench. Although fir was chosen for the uprights, many of the wall poles were hazel branches. The chinks between the poles were filled with moss and reeds. One can assume that the housebuilders burned over the area they had cleared to prepare the ground for cultivation, but there is no way of knowing whether they farmed the area the following spring.

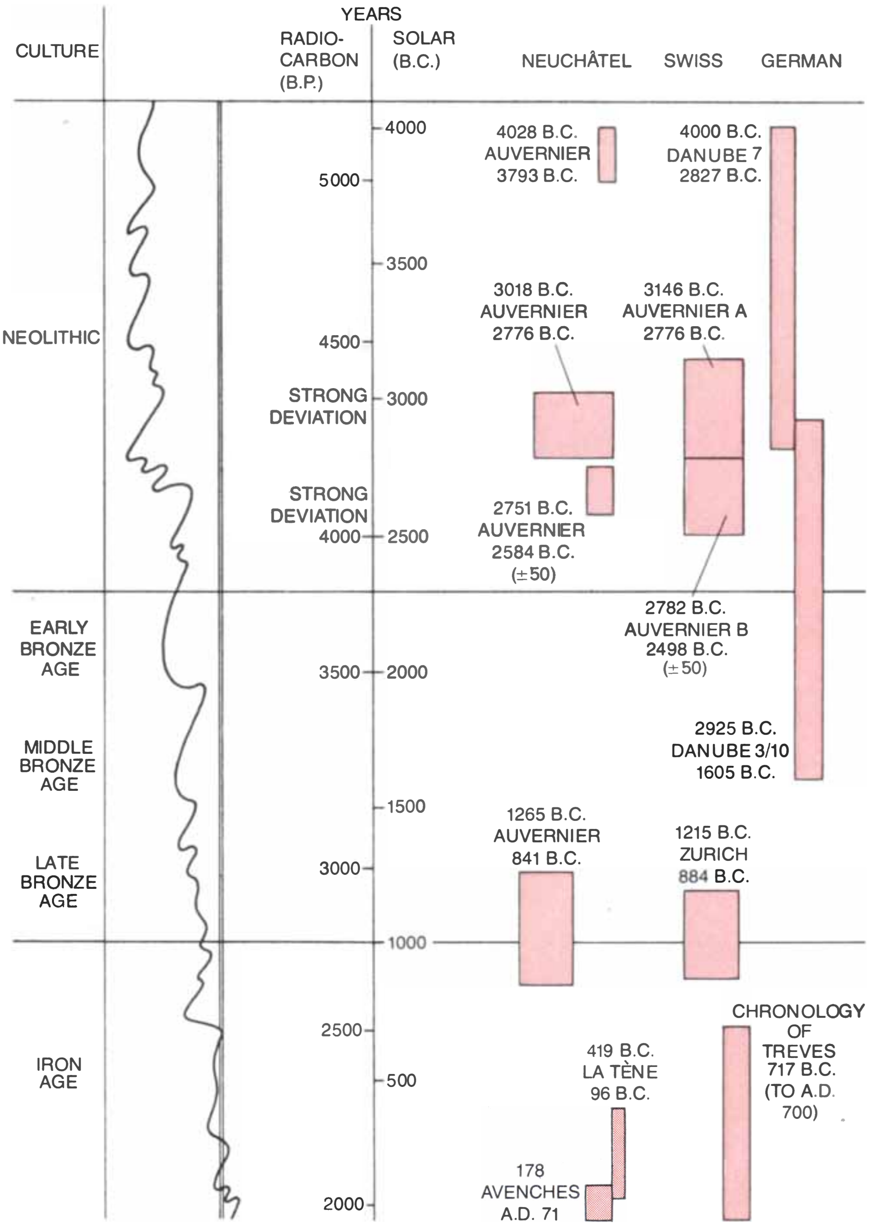
A second house, identical with the first, was built the following winter. It was south of the first house and was separated from it by an alley only 1.2 meters wide. The door of the house, in the north wall, opened on this alley. Today the prevailing wind at Lake Paladru is from the north; the east-west orientation of the Neolithic house suggests that 5,000 years ago the prevailing wind was from the east.

With the building of the second house the Neolithic settlement seems to have been firmly established. It was now destined to be occupied for some 30 years. The numerous artifacts we have brought up from the lake bottom enable us to reconstruct with some confidence the villagers' daily lives. Stone tools are less common than a 30-year occupation would suggest. Made from locally available nodules of flint, they are for the most part shaped roughly, although a few of them, including blades and scrapers, are elegantly pressure-flaked; so are the two flint daggers we recovered with their wood handles intact. Certain of the other flint tools also showed traces of wood handles.

Most of the containers that have survived are clay pots, simple in shape and not too well fired. Bits of woven basket indicate that the villagers did not depend for storage on pottery alone. The fragments of textiles provide evidence that weaving was practiced, as do needles, many balls of thread, wood loom combs and wood spindles suitable for the spinning of wool and flax. The variety of these artifacts makes it clear that the villagers did their own weaving rather than importing cloth.

Among the other artifacts made of wood are spoons that could easily be mistaken for the wood spoons of today. The villagers' capability as woodsmen, clearly apparent in the construction of their houses, is further indicated by a number of long ax handles. One of these still held a blade of polished stone. Other handles were evidently fitted with sleeves made of antler to help cushion the shock of chopping.

What did these people eat? Although they were farmers, they were by no means entirely dependent on



TREE-RING CHRONOLOGIES for Europe extend from about 4000 B.C. to the present but do not form an unbroken sequence. The three longest sequences are the Chronology of Treves, which extends approximately from 750 B.C. to A.D. 700, and two overlapping Danube chronologies that run from 4000 B.C. to about 1500 B.C. The tree-ring dates at Baigneurs correlate with a Swiss chronology, Auvernier, that extends for some 250 years starting in about 3000 B.C. For much of the period between 5000 and 1000 B.C. age estimates based on carbon 14 are later than tree-ring dates; for example, samples of 3000 B.C. (some 5,000 years ago) yield carbon-14 dates of about 4400 years B.P. (before the present). The erratic line at left shows carbon-14 discrepancy calculated by Hans E. Suess of University of California at San Diego.

domesticated foodstuffs. The animal bones we have recovered indicate that they regularly hunted deer and on occasion even bears. Little flint projectile points suggest they also hunted small game, but the bones of such animals have not survived. The villagers fished in the lake; we have found stone net weights and bits of netting (and a single copper fishhook, even though 3000 B.C. is long before the age of metal began in this part of the world). From their domestic animals, and perhaps those of their neighbors, the villagers supplemented their meat diet with beef, mutton, pork and goat meat.

From the villagers' domesticated plants—wheat and barley—they milled a coarse flour with which they baked flat cakes; the milling was done with millstones of granite. They also cultivated flax, and although they did so mainly to get fiber, they may have prized the oily flax seeds. They gathered numerous wild plant foods. In addition to the acorns and hazelnuts I have mentioned they gathered beechnuts, blackberries, wild plums and apples. They must surely have baked their cakes on the clay hearths and broiled their meat over open flames, but they also knew the art of stone boiling. For this kind of cooking one partly fills a pot with water and heats a number of stones in the fire until they are almost red-hot. The hot stones and the food are then put in the pot together and the stones bring the water to a boil. The sudden quenching often shatters the stones. These people used quartzite pebbles as their boiling stones, and their hearths became littered with bits of quartzite.

In the ninth winter of their occupation the villagers completely rebuilt the first house. Nine years later they rebuilt both houses. The house plan remained the same: an oblong with a central hearth. Over those 18 years the level of Lake Paladru may have fluctuated slightly in response to variations in the climate and on occasion could even have wet the floors of the houses. There was no major flooding; the settlement was eventually destroyed not by water but by fire.

We cannot tell exactly when the houses burned down. The fire could scarcely have been later than some 30 years after the founding of the settlement. By that time there would have been little left to burn: the maximum lifetime of this kind of wood house at a shoreline site is only 15 years. At the same time our tree-ring records indicate that the disaster came sometime after the 19th year of the settlement; we have recovered a house post with growth rings extending to that year, evidently a late addition to one of the houses.

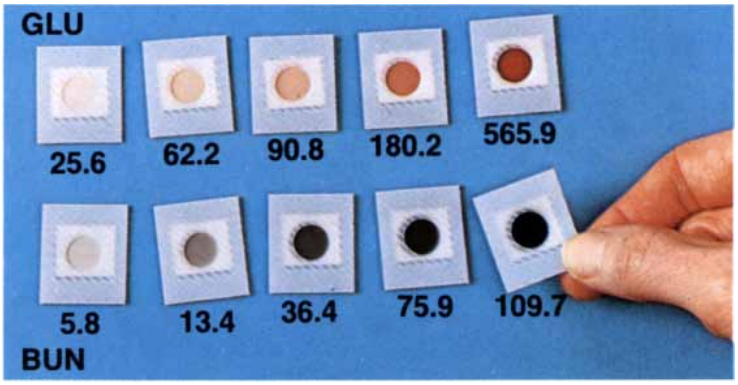
The character of the plant community that reinvaded the deserted settlement was different from that of the primeval stand of fir. The forest was opener to sunlight; among the firs grew alders, elms and ashes. As the forest advanced the lake level rose. Eventually the water stood perhaps 1.5 meters above its former level, and the burned ruins were covered by a thin layer of silt. The silt layer was then colonized by beds of reeds growing along the new shoreline.

Some 60 years after the founding of the first settlement, at a time when Lake Paladru had dropped back to its earlier level, a second Neolithic farming group settled in exactly the same place near the lake outlet. The style of their pottery and their flint artifacts was the same as that of the earlier group; the newcomers may even have been descendants of the original settlers. They did, however, build houses that were smaller and more nearly square. Surviving clay hearths were used by the newcomers, but in general their houses on the headland were more haphazardly located. If there is any difference between the two groups other than the size of their houses, it is that the early settlers floored their houses with fir boughs and ferns and their successors sometimes floored them with bark. When it came to means of subsistence and way of life, the settlements were the same.

What subdivision of European Neolithic culture is represented at Baigneurs? Studies of the pottery and the flint artifacts indicate that these lakeshore people were members of a widely distributed population whose cultural tradition flourished in western Switzerland, in the Jura Mountains, in the valley of the Saône, in the Savoy and in northern Dauphiné. To European prehistorians the culture is known as the Saône-Rhône civilization. The western members of this population had close relations with the Neolithic farmers to the south in France; those closer to Switzerland and within it had similar ties with the Neolithic farmers of central Europe. Carbon-14 analysis places the Baigneurs settlements in an early Saône-Rhône phase, dated at about 2900 B.C. Established by chance at the end of Lake Paladru, the two settlements spanned some 90 years before the second settlement was abandoned for no apparent reason. Thereafter a rise in water level transformed the lakeshore into lake bottom and the ruins of Baigneurs remained untroubled for five millenniums.

STRATIGRAPHIC SECTION of lake bottom at Baigneurs shows, from top to bottom, an upper layer of sediments that proved to be sterile, or entirely lacking in artifacts (*light area*), a layer containing the debris of the second Neolithic occupation (*thick dark layer*), a second sterile layer of sediments separating the second Neolithic occupation from the first (*light*), a very thin first stratum of occupational debris (*dark*) and finally the sterile lake bottom below it. Wedged between the lake bottom and the first stratum is a flat milling stone made of granite.

Out of photographic technology but not at all photographic

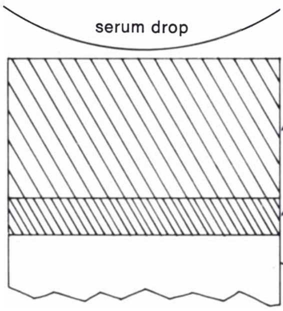


KODAK EKTACHEM Clinical Chemistry Slides.
Actually, nobody handles them or looks at them.

At several hospitals in recent months, laboratory tests for blood glucose (GLU) and blood urea nitrogen (BUN) have been done by a new method based on little squares of film, 16 mm x 16 mm, in plastic mounts as though for a projector. But they bear no images.

In health care, society can ill afford novelty for the sake of novelty. To justify our massive multidisciplinary effort, we must first tell you that the popular image of the clinical laboratory, source of the numbers with which today's quantitatively thinking physicians do their thinking, is out of date. The bubbling glassware representing science at work can't handle the volume of routine tests routinely ordered. Nowadays little test tubes move through very complex machines as liquids are added, withdrawn, stirred, filtered, heated, cooled, bubbled, scanned, etc. in mechanical imitation of the chemists who were called in when the doctors stopped relying on just the gross appearance of the patient and the patient's contents.

Our approach restores simplicity. No pipes, no squirts. The only liquid is the serum under assay. It works because we know how to coat layer upon tissue-thin layer of compositions within which chemical and physical events and interactions can be controlled with a nicety that seemed preposterous when first proposed for color photography long ago. In those days color photography meant three-in-one cameras and infinite patience. Photography has become what it is today because the more that's packed into the layers of the film, the simpler the manipulation outside the factory. So it is to be in clinical chemistry, we intend. To wit:

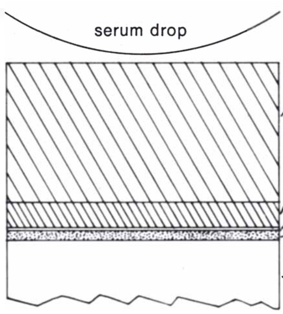


GLUCOSE ASSAY

80% air **SPREADING LAYER** whitened with TiO_2
Precision metering by isotropic porosity instead of fancy plumbing. A variation of 10% in drop volume would make only 1% difference in serum per unit area reaching reagent layer.

REAGENT LAYER
Contains
1. glucose oxidase, enzyme that catalyzes oxidation of glucose to gluconic acid, releasing H_2O_2;
2. peroxidase, enzyme that catalyzes coupling, in proportion to H_2O_2 present, of
3. 7-hydroxy-1-naphthol to
4. 4-aminoantipyrine to make dye;
5. buffer to maintain pH near optimum for both reactions

TRANSPARENT SUPPORT through which density of dye to be formed will be measured against white reference



BUN ASSAY

SPREADING LAYER
Same as for GLU

REAGENT LAYER
Loaded with urease, enzyme which releases NH_3 and NH_4^+ from urea. Buffered to pH 8 to push equilibrium toward NH_4^+.

BARRIER LAYER
Cellulose acetate butyrate passes only nonionic substances, bars buffer and OH^- which would cause dye to form if they penetrated.

INDICATOR LAYER
Cellulose acetate carrying an indicator that gives color weakly with NH_3 and not at all with NH_4^+. (Otherwise the range of color density would be too wide for convenient reading because the corresponding BUN range of medical interest—4 to 130 mg%—is so wide.)

TRANSPARENT SUPPORT through which dye density will be measured

For answers to the kind of questions that are prompted by more than just passing curiosity, write to R. Barnes Parsons, Kodak, Rochester, N.Y. 14650.

The KODAK EKTACHEM *GLU/BUN* Analyzer is our own version of an instrument which accepts undiluted serum samples and blank KODAK EKTACHEM Clinical Chemistry Slides, measures concentration of GLU or BUN or both, and accumulates the used slides for disposal. It receives push-button instructions to set itself for GLU or BUN, applies a 10-μl drop to the appropriate type of slide with no carryover from one sample to the next, incubates for the proper number of minutes at 37°C, and uses the TiO_2-impregnated spreading layer of the slide as a background for a photomultiplier to read reflection density through a narrow-band filter at 540 nm for GLU or 670 nm for BUN. The precision required of that reflectometer unit was the kind of challenge to which optical engineers in the color photography business have become accustomed. The microprocessor in the unit even corrects for nonlinearity of reflection density with dye concentration!

Now wait till you hear about the little chip of nonphotographic film for assaying triglyceride in serum. It contains ATP, the storage stuff for biological energy, and four enzymes, including peroxidase from horseradish and an oxidase we prepare from *Streptococcus faecium*. Or our film for amylase assay. Or the one for bilirubin. We could go on and on. In fact, we have.

PONTIAC GRAND PRIX

YOU'LL FIND SOMETHING NEW TO LOVE EVERY TIME YOU DRIVE IT.

At first, it may be Grand Prix's beautiful, trim styling that captivates you the most. With its rich, formal roofline. And new cross-hatch grille.

The next time, it may be Grand Prix's luxurious interior that charms you. With its soft, available bucket seats. Its impressive roominess. And impressive quiet, due in part to special body mounts and door and body seals.

It's easy to fall in love with Grand Prix's road ability, too. There's a new freer breathing standard 3.8 litre (231-CID) V-6 that will amaze you with its response. Grand Prix is equipped with GM-built engines produced by various divisions. See your dealer for details.

It all makes Pontiac's new 1979 Grand Prix one very enchanting automobile to buy or lease.

And makes owning one a love affair . . . that never seems to end.

GM

THE 1979 PONTIACS OUR BEST GET BETTER

SCIENCE AND THE CITIZEN

The World According to GARP

A description of the earth's atmosphere as a single physical system has long been sought by those who attempt to penetrate the mysteries of weather and climate. In the early 1960's the United Nations' World Meteorological Organization established the World Weather Watch, a global system for the collection, processing and dissemination of meteorological data. A short time later the World Meteorological Organization joined the International Council of Scientific Unions in launching the Global Atmospheric Research Program (GARP), a project aimed at extending the range of weather-prediction systems and obtaining a better understanding of the physical basis of climate. This year a decade of successful international cooperation in GARP culminates in what may be the most extensive international scientific undertaking in history: more than 140 nations are participating in the Global Weather Experiment, which seeks to determine an optimum system for the observation and prediction of world weather.

On December 1 of last year the Global Weather Experiment began an intensive monitoring of the earth's atmosphere and oceans that is expected to yield the most complete record of global weather ever assembled. This unprecedented data-collecting operation, touching on a vast array of meteorological phenomena, will continue for one year. More than 40,000 daily weather observations are provided by the surface and upper-air systems of the World Weather Watch, which include 3,400 land stations, 800 upper-air stations, nine weather ships, more than 7,000 merchant ships and 1,000 commercial aircraft. Data on the expanses not accounted for by the World Weather Watch observations are provided by five geosynchronous ("stationary") weather satellites, several polar-orbiting weather satellites, more than 50 research ships, some 100 aircraft and 300 high-altitude weather balloons. In addition some 300 specially equipped drifting buoys have been deployed across the oceans of the Southern Hemisphere, where until now there has been little meteorological monitoring.

Another area for which weather data are sparse is the Tropics: the broad belt around the Equator that absorbs much of the sun's energy and generates much of the world's weather. (A large fraction of the solar energy that reaches the earth is first stored in tropical clouds and later dispersed by high-altitude winds.) The Tropics are the subject of two special observation periods, chosen to coincide with the winter and summer monsoon seasons: from January 5 through March 5 and from May 1 through June 30. Among the variety of additional observations made during those periods the measurement of tropical winds plays a particularly important part.

When the Global Weather Experiment ends on November 30 of this year, the mass of collected data will be sent to processing centers in a number of countries. Larger research facilities in the U.S., the U.S.S.R., Britain and Sweden will undertake the final refinement of the data, applying powerful computers to generate the first comprehensive history of an annual cycle of world weather. With this unique record it should be possible to construct (and validate) new, highly realistic models of atmospheric circulation. These models should serve to indicate, among other things, the theoretical limits of weather prediction. It is hoped that with an improved theoretical and practical understanding of the global weather system it will be possible to develop systems for weather observation and forecasting that attain those limits.

The analysis of the data collected in the course of the Global Weather Experiment will almost certainly continue far into the future. As demands on the earth's natural resources mount, an understanding of the natural (and unnatural) fluctuations of weather and climate becomes increasingly important. A conference will be held this month in Geneva to lay the plans for the World Climate Program: an international project that seeks to answer a wide range of questions about man and his climate by the year 2000. It seems likely that this project and others similar to it will continue to exploit the results of the Global Weather Experiment for many years.

The Other Energy Crisis

In the poorer nations of the world as much as 86 percent of the wood consumed annually is burned as fuel. Rural people in particular are almost totally dependent on wood and such other noncommercial fuels as animal dung and crop residues. The pressure on wood supplies as a result of increasing population in those countries is giving rise to what has been termed "the other energy crisis," which may last longer and be more difficult to overcome than the one associated with the increase in the cost of petroleum. The wood crisis is discussed in *Unasylva,* a journal of forestry published by the Food and Agriculture Organization of the United Nations, by J. E. M. Arnold of the FAO Forestry Department and Jules Jongma, a Dutch forest economist.

With the caution that most of the data on the consumption of wood as fuel are estimates, since most of the gathering and burning of wood takes place outside of commercial channels, the authors put the annual use of wood for fuel at 1.2 billion cubic meters in the developing countries and 150 million in the developed countries. "Wood fuels account for two-thirds of all energy other than human and animal energy used in Africa, for nearly one-third in Asia, for one-fifth in Latin America and for 6 percent in the Near East," they write, whereas in the developed countries the level is .3 percent. In the developing countries "the major obstacle to the substitution of other commercial fuels, even in towns and cities, is quite simply that people are too poor to buy the necessary equipment to use them."

The pressure on supplies of fuelwood makes the crisis worse in several ways. People have to go farther to obtain wood. They burn more animal dung and crop residues, which might otherwise be employed in agriculture. ("The diversion of dung from use in agriculture is equivalent to burning food in order to cook food.") The decline in agricultural output is likely to create pressure to encroach further on the forest for farmland. The human effort required to gather fuelwood may create a shortage of agricultural labor.

Arnold and Jongma conclude that wood is likely to be the principal source of energy in the developing countries for the foreseeable future, since the poorer people are unable to afford anything else. Focusing therefore on what can be done to stretch the supply of wood, they find "a number of options." One is to burn wood more efficiently through simple improvements in the preparation of wood and in the design of cooking pots and stoves. Another is to convert more wood into charcoal, which not only is easier to transport but also can be "based on the large volumes of wood of other than commercial timber species which are otherwise destroyed in land clearing . . . or which are left unused in tropical forests after logging." A third option is to develop fuelwood plantations. Since such plantations require mainly land, labor and time, they "may enable poor rural populations to generate fuel supplies at an acceptably low cash cost."

Communication by Packet

A relatively little-known communications technology has revolutionized the transmission, exchange and sharing of data over the past 15 years, according to an issue on the subject published by *Proceedings of the Institute of Electrical* *and Electronics Engineers.* Called packet switching, the technology enables the units in a communications system (such as the central processing unit, the memories, the data channels and the terminals of a computer system) to "talk" with one another without the need of direct or circuit-switched transmission lines running between all units. Instead the units are connected in a network configuration, so that there are numerous indirect pathways linking any two of them.

The packet-switching system divides a message into packets, or small sequences, of data and separately moves each packet to a common destination. Each packet, however, may reach the destination by a different route. The key feature of the packet-switching network is that such routes are not predetermined. The system as a whole selects the route as the trip is being made on the basis of up-to-date data-traffic conditions. It is as if a human driver set out for a particular destination with no specific route in mind, preferring instead to decide at each intersection what road to take on the basis of the current traffic conditions.

In communications there have always been two competing approaches to constructing transmission pathways: preallocation and dynamic allocation. The telephone system and commercial radio fall in the category of preallocation. The telephone network is a circuit-switched system, where a fixed transmission line is preallocated for the duration of each call. Commercial radio involves the permanent preallocation of a band of wavelengths for each station. The mail system, on the other hand, falls in the category of dynamic allocation. The exact path a letter takes from its source to its destination is not determined in advance; rather, routing decisions are made along the way. Hence packet switching is not a new invention in communications but a sophisticated extension of a system that postal services have used for more than a century.

The packet-switching network is ideal for a large number of units, each of which makes a relatively small demand on the system but needs access to all of it. Under such circumstances the packet-switching network is much more efficient than preallocation systems, in which individual transmission lines are idle except for occasional moments of preallocated use. Because a decision on where to send the packet next must be made at each node, or intersection, in the network, microprocessors (or microcomputers) that can make such decisions must be stationed there. Hence packet-switching systems reduce transmission-line waste by increasing processing power. In 1969 the cost of dynamic-allocation microprocessors fell below the cost of preallocation transmission lines for many applications, and so the development of packet-switching systems was accelerated.

Modern packet-switching systems that communicate data to and from computers were developed first for military applications. In the 1960's the Advanced Research Projects Agency (ARPA) of the Department of Defense built ARPANET, the first packet-switching computer system that linked more than 100 pieces of equipment. A microprocessor located at each node in the network took blocks of data from the equipment connected to it, divided the data into 128-byte packets and added to each packet the address of its destination. On the basis of a constantly updated traffic report the microprocessor sent the packet toward its destination over the least congested route. Whenever the packet reached a node, the microprocessor there received it, sent a message back to the source acknowledging receipt and rerouted the package according to the current traffic conditions.

ARPANET demonstrated successfully that a packet-switching network could make the resources of each unit (memories, terminals and so on) available to every other unit. The success of ARPANET prompted Government agencies, businesses and laboratories to establish packet-switching networks to connect computer resources. Plans are now afoot to extend packet-switching techniques to both radio and satellite communications. The basic idea of packet radio is to have many stations, each of which transmits only in short bursts, share one relatively broad band of wavelengths. By dynamically allocating channel space in this way none is wasted. If broadcast-communications satellites are incorporated into packet-switching systems, then geographically remote units can all be linked together through a common network. This means, for example, that an isolated research center could have access to the most sophisticated computers at a considerable distance without transmission lines lying idle when access was not needed.

Missing Link

Wheat supplies most of the starch and protein in the human diet. It has therefore been the subject of intensive genetic manipulation with the goal of increasing its yield, its nutritional value and its resistance to drought and disease. A major approach has been hybridization: crossing two species of wheat in the hope that the offspring will possess more desirable qualities than either parent. Now Moshe Feldman of the Weizmann Institute of Science in Israel has identified a new wheat species that appears to be the "missing link" in the evolution of all cultivated bread wheats. Named *Triticum searsii* for Ernest R. Sears of the University of Missouri, a

MISSION IMPOSSIBLE? NOT FOR HUGHES.

The mission:

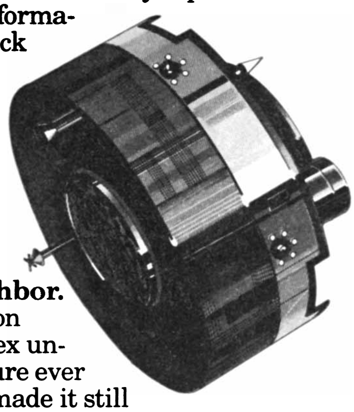
Build two different kinds of spacecraft. To take two different flight paths to Venus. And send back to Earth a stream of new information.

Orbiter arrives.

The first spaceship was Orbiter. Crammed with a dozen scientific instruments, it was launched last May by NASA. 300-million miles later, it arrived at Venus. But it's still traveling. It's now on a series of 243 one-day elliptical orbits around the planet—studying its atmosphere and mapping its terrain, close in and far away.

Multiprobe arrives.

The second spaceship was Multiprobe. Carrying 18 instruments, it was launched in August by NASA on a more direct 220-million mile trip. At a point 7.8 million miles from Venus, it divided into five fact-finding probes. And then these probes, including the parent "bus" that took them there, entered Venus' atmosphere to explore five widely separated planet areas. The information they beamed back about the planet's winds, clouds, and atmosphere will help clarify the mystery of how our own weather operates here on Earth.



A hostile neighbor.

The twin mission was the most complex unmanned space venture ever undertaken. What made it still tougher was the downright hostile nature of our nearest planet neighbor, as experienced firsthand by Multiprobe.

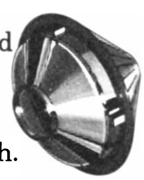
920° hot.

Venus has a surface temperature of 920°F.—hot enough to melt tin or lead. Its surface pressure is as crushing as the ocean 3,000 feet deep. Its atmosphere is almost pure carbon dioxide. And its dense clouds aren't innocent water. They're sulfuric acid.

Aluminum blankets.

But scientific ingenuity at Hughes took up the challenges. For example, Multiprobe's fragile internal electronics were guarded by blankets made of special aluminized plastic sheets with great resistance to intense heat.

Titanium shells.



Special titanium shells proved to be ideal pressure vessels. Light in weight, they still could resist corrosion and 1,400 pounds of pressure per square inch.

A diamond window

Finally, our designers needed an unusual window for an instrument that senses radiant energy. Typical window materials weren't rugged enough. Sapphire windows used for other probe instruments would block infrared wavelengths. Solution: a 13-carat diamond window the size of two pennies stacked together. It worked.

90 revealing minutes.

In 90 minutes, the twin mission managed by NASA's Ames Research Center told 115 scientific and technical investigators more about Venus than astronomers have learned in the five centuries since Galileo.

Mission impossible?
NASA didn't think so.
And neither did Hughes.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

pioneer in wheat genetics, the new species represents an entirely untapped genetic stock and promises to be a valuable source of new traits for improving existing varieties of bread wheat.

Cultivated wheats are polyploid: they have multiple sets of chromosomes derived from two or more closely related progenitor species. All bread wheats in existence today incorporate three genomes, or complete sets of chromosomes, derived from three primitive wheat species. The genomes are designated *A*, *B* and *D*. Only the primitive species carrying the *A* and the *D* genome had been identified in the wild. The progenitor plant bearing the *B* genome was unknown, and most wheat geneticists believed it had died out more than 10,000 years ago. *T. searsii*, the new species identified by Feldman, appears to be the carrier of the *B* genome.

Feldman's work on *T. searsii* began in the early 1960's, when he was a graduate student at Hebrew University in Jerusalem. One day when he was passing an empty lot near the campus, he noticed a strange-looking wild wheat plant. He tentatively identified it as a small-eared variety of the common species *T. longissimum* and put it away for further study. In 1972 he came across the atypical plant again during a field trip and speculated that it might be a previously unknown species rather than a variety of *T. longissimum*.

In collaborating with a taxonomist, Mordechai Kislev of Bar-Ilan University, Feldman compared the atypical plant (later to be named *T. searsii*) with *T. longissimum* and found enough structural differences to classify it as a separate species. A karyotype analysis (an examination of the number and shape of the chromosomes) also demonstrated the nonidentity of the two plants. Surprisingly, however, the single set of chromosomes in *T. searsii* proved to be virtually identical with the *B* genome in *T. dicocoides* (a wild wheat containing both the *A* and the *B* genome) and in cultivated bread wheats (which contain the *A*, *B* and *D* genomes). Biochemical studies at the Weizmann Institute and at other laboratories also supported the identification of *T. searsii* as the contributor of the *B* genome.

Many wheat cytogeneticists, recalling their too-quick acceptance a few years ago of the erroneous notion that *T. speltoides* was the donor of the *B* genome, are taking a "wait and see" attitude. They want to know whether or not the chromosomes of *T. searsii* will be found to pair more readily with those of genome *B* than with those of *A* or *D*.

The *B* genome is particularly significant to wheat breeders because it contains a deleterious gene that discourages the pairing and exchange of chromosomal segments if they do not have a precisely equivalent structure. The chromosomes of different varieties of bread wheat often differ slightly as a result of evolutionary changes, so that the antipairing trait has considerably complicated the transfer of genetic information by crossbreeding. Feldman is currently searching for individuals of *T. searsii* in which the antipairing trait is weakened or absent. The discovery of such individuals would aid in the production of new wheat species that are easier to breed.

Feldman is also engaged in synthesizing a polyploid bread wheat from its three primitive progenitors: *T. monococcum* (the *A* genome donor), *T. searsii* (the presumed *B* genome donor) and *T. tauschii* (the *D* genome donor). He has already succeeded in obtaining hybrid plants containing both the *A* and the *B* genome and is crossing the *AB* plants with the *D* donor. When the synthetic wheat plant has been created, it will serve as a bridge for the transfer of desirable genetic information from various strains of primitive wheat to existing commercial varieties of bread wheat.

The Handmade Six-Gun

A central concept of mass production is the interchangeability of parts, and the historical paradigm of mass production based on interchangeable parts has long been held to be the manufacture of the Colt revolver in the middle of the 19th century. It is apparently not so. Robert A. Howard of the Hagley Museum in Wilmington has looked at old records and old revolvers and has shown that the storied handguns of the American West and the Civil War were made of "very similar parts but they were not interchangeable." His account of interchangeability in U.S. arms manufacture is given in a recent issue of the journal *Technology and Culture*.

Interchangeability, Howard points out, was a function of both the need for precision and the economic realities. The muskets and the muzzle-loading rifled muskets turned out by Government armories required very little precision, and the military market imposed no cost imperatives; parts could be and were interchanged. It was different for the private sector of the arms industry, which manufactured the revolvers. Making a revolver is a tricky business; the element of timing is introduced. For example, the loaded chamber must be in absolute alignment with the barrel during discharge; the detent that locks the chamber in alignment must engage during discharge and no longer, or the chamber will not rotate to present the next round.

The aim of the revolver manufacturers was to make high-quality arms as economically as possible. They made extensive use of machinery (drop-forging, milling and grinding machines and turret lathes, for example) to produce parts that were respectably similar but not similar enough to be interchangeable. Full interchangeability was too expensive. The standard practice was to depend on "fitting": choosing from the supply of parts to be mated two that fitted well together and reducing the size of a part that was too large by filing. Fitting, according to Howard, "required greater skill in one operation but resulted in lower overall costs."

Samuel Colt did not invent the mechanical revolver (a revolver whose chamber was rotated by the action of the hammer), but the one he patented in 1836 was the first to be manufactured in large quantities; by the beginning of the Civil War he had produced some 300,000, far more than all the other manufacturers combined. If anyone had attained interchangeability by 1860, the year whose output Howard examined in detail, it should have been Colt. Moreover, the effervescent Colt repeatedly made such assertions as: "The separate parts travel independently through the manufactory [and are] assembled, from promiscuous heaps, and formed into fire-arms, requiring only the polishing and fitting demanded for ornament." Howard asked a number of small-arms experts to examine their collections of Colt revolvers and other revolvers of the period. The revolvers were disassembled and the attempt was made to assemble functioning weapons from parts chosen at random. It could not be done; the parts were not interchangeable.

Even today revolver parts are not interchangeable. Howard checked with Colt and with Smith & Wesson, which is now the major manufacturer of revolvers. A Colt spokesman said: "All the guns require . . . varying degrees of hand fitting." The man at Smith & Wesson replied that "the use of totally interchangeable parts . . . has never been totally accomplished."

Computer on the Don

For some time scholars have conducted statistical studies of literary works, a task for which large computers are notably well suited. A topical case in point is presented in the journal *Computers and the Humanities* by Geir Kjetsaa, professor of Russian literature at the University of Oslo. He describes how textual analysis by computer may have helped to solve a long-standing question about the authorship of *The Quiet Don,* a classic of modern Russian literature.

An epic four-volume work dealing with the Don Cossacks before and after the Russian Revolution, *The Quiet Don* is the most widely read novel in the U.S.S.R. Ever since Mikhail Sholokhov first published parts of the novel in a Moscow periodical in 1928, however, there have been persistent rumors that he plagiarized much of the book. Such

speculation continued even after Sholokhov received the Nobel prize for literature in 1965.

In 1974 the rumors were given more substance when an anonymous Russian critic published a study in Paris maintaining that the real author of *The Quiet Don* was the nearly forgotten Cossack writer Fyodor Kryukov. Kryukov had served in the Don army during the Cossack uprising against the Communists and died of typhus during the retreat of 1920. According to witnesses, before his death Kryukov had been working on a major novel dealing with the Don Cossacks. The Russian critic accused Sholokhov of somehow obtaining Kryukov's manuscript, rewriting about 5 percent of the first two volumes and 30 percent of the second two volumes (in order to overshadow the Cossack nationalism of the original manuscript with pro-Communist views) and publishing it as his own work.

This hypothesis predicted that the language and style of Kryukov's authenticated writings should have more in common with that of *The Quiet Don* than Sholokhov's authenticated writings do. Kjetsaa and some Swedish and Norwegian co-workers set about testing the prediction by utilizing an IBM 370/155 computer at the University of Uppsala to compare the linguistic and stylistic features of the novel with undisputed texts by Kryukov and Sholokhov. Textual analysis cannot prove definitively that a work is by a particular author, since the possibility remains that it might have been written by another person with the same stylistic traits. It can, however, exclude the possibility of authorship, much as a blood test can clear a man of the suspicion of paternity even though it cannot prove he is the real father of the child.

Kjetsaa and his colleagues analyzed passages in the original Russian from the novel and from representative texts by both authors. Paragraphs containing direct speech or a character's thoughts were omitted in order to concentrate on the author's distinctive "voice." The study measured sentence length and the distribution and combination of parts of speech at the beginning and end of sentences. Since style is a complex phenomenon, however, analysis of vocabulary structure was also necessary to arrive at an unequivocal conclusion.

The type-token ratio, a parameter that reflects the size of the vocabulary employed by the author, was determined for each text. Computer analysis showed that Sholokhov's undisputed text had the richest vocabulary (with a type-token ratio of .656), closely followed by *The Quiet Don* (.646). Kryukov's text (.589) fell markedly behind, because it had significantly more repeated words.

Another linguistic parameter employed in the analysis was the lexical spectrum: the frequency with which certain words appear in a text. Here again Sholokhov was closer to the novel than Kryukov: the 20 commonest Russian words amounted to 22.8 percent of Sholokhov's text, 23.3 percent of *The Quiet Don* and 26.2 percent of Kryukov's text. Even more striking differences were seen in the number of words appearing only once in the text, which occupied 80.9 percent of the vocabulary in Sholokhov's text, 81.9 percent of the vocabulary in *The Quiet Don* and only 76.9 percent in Kryukov's text.

Assuming that the sampled texts are representative of the two writers, the results of the analysis clearly exclude Kryukov as the possible author of the novel. Although the findings cannot prove that Sholokhov did write the book, Kjetsaa concludes: "All the parameters investigated showed a uniform tendency, viz. that Kryukov is markedly dissimilar from Sholokhov in his writing, and that Sholokhov writes remarkably like the author of *The Quiet Don.*"

When you travel as far as the high Andes,

you stay for 3000 years.

NOVA travels to the isolated mountain world of the Q'eros Indians ... where time has stood still for three thousand years ... where the seeming simplicity of their lives conceals a complex pattern of survival. Watch NOVA's *Patterns from the Past.*

Photos: John Cohen

Past. Present. Future.

This month, NOVA looks at the ancient world of Peru's Q'eros Indians ... the man-made wilds of the new super-zoos, where endangered species are getting a second chance ... plus two important programs on possible solutions to the world's devastating energy shortage.

NOVA takes you to worlds you've never seen.

Series of Surprises

Exciting journeys to unusual places. Rare encounters with fascinating people. Meetings with remarkable ideas.

Take NOVA to a new world every week.

NOVA

Thursday, February 8 at 8:00 pm*

* Check local listings

PBS

A COMPANY CALLED TRW

NOVA is made possible in part by a grant from TRW.

Silviculture

The management of a forest by such techniques as the selective harvesting of trees, prudent thinning and planned reforestation can double and perhaps even triple the productivity of the forest

by Stephen H. Spurr

Just as agriculture is the cultivation of fields, so silviculture (from the Latin *silva* or *sylva,* meaning forest) is the cultivation of forests. The former is far more widely practiced than the latter. If the forests of the world are to be counted among the world's renewable resources, however, silviculture will have to be much more widely and intensively practiced than it is now. The rewards could be substantial. Allowing for the realities of ecology, economics and politics, one can still calculate that the application of demonstrated silvicultural principles could double and possibly triple the productivity of the commercial forests in the U.S. within 50 years.

The land area of the world amounts to some 13.3 billion hectares, or 32.8 billion acres. (A hectare is 2.47 acres.) About a third of the area, or 4.1 billion hectares, is classified by the Food and Agriculture Organization of the United Nations as being forestland. Most of the major regions, except for Asia and the Pacific, have at least a fourth of their land under forest; in North America, South America and the U.S.S.R. the total is more than a third. In the U.S. 305 million hectares (a third of the land area) is classified as forestland. Some 202 million of those hectares can be characterized as commercial timberland under the definition established by the U.S. Forest Service: land capable of growing at least 1.4 cubic meters of timber per hectare per year (20 cubic feet per acre) and not legally withdrawn from timber harvesting, as the wilderness areas designated by Congress are. The fact that a given area is classified as commercial timberland does not necessarily mean that timber is harvested on it or even that it has a stock of trees, only that it is capable of producing wood at the defined rate.

Another standard employed by the Forest Service defines wood as the merchantable boles, or trunks, of forest trees that are at least five inches (12.7 centimeters) in diameter at a height of 4.5 feet (1.37 meters) above the ground; the definition includes all wood above a one-foot stump and extending upward to the point where the trunk becomes less than four inches in diameter. By this definition about half of the total biomass produced by a forest is usable wood. One should bear in mind that the forest has many other values, most of which cannot be priced. The respiratory processes of forest trees, taken together, play a major role in removing carbon dioxide from the atmosphere and discharging oxygen into it. Forests are the principal source of fresh water, since they occupy most of the land area of the world that is not frozen and where precipitation exceeds evaporation. Forests everywhere are an important source of browse for both livestock and wildlife. They also provide much space and opportunity for human recreational activity.

Types of Forest

One tends to think of a forest as consisting of a variety of trees that have established themselves through the more or less random effects of natural forces. Actually a forest is more likely to consist of a fairly limited variety of trees, mainly because the conditions of temperature, rainfall and sunlight are favorable for certain species and unfavorable for others. Where human managers have taken a strong hand, a forest may consist of a single species of tree.

Because of environmental effects it is possible to classify the major forest types of the world into a fairly small number of groups. The forests of the high-latitude regions of the Northern Hemisphere known as boreal forest are dominated by stands of such conifers (needle-leaved trees, or softwoods) as spruce, fir and larch. In the middle latitudes of the Northern Hemisphere the forests tend to consist both of conifers such as pine and Douglas fir and of broad-leaved deciduous trees (hardwoods), notably oak, beech and maple. Close to the Equator one finds the tropical rain forests, dominated by broad-leaved evergreen trees (a wide variety of species) that are adapted to hot and humid conditions.

In the U.S. the Great Plains divide the predominantly coniferous forests of the West from the mixed coniferous and broad-leaved forests of the East. Along the Pacific Coast, Sitka spruce, western hemlock, Douglas fir and, in northern California, redwood are the principal tree species. Inland to the crest of the Cascades and the Sierra Nevada, Douglas fir is the most abundant, with cedar, true fir, hemlock and other conifers growing in various mixtures depending on soil, elevation and precipitation.

From the east-facing slopes of these mountains eastward at higher elevations in the Rocky Mountains ponderosa pine occupies the lower and drier slopes, with lodgepole pine growing on cooler and wetter heights and Douglas fir, Engelmann spruce and true fir forming the highest continuous forest. In the northern Rocky Mountains near the Canadian border western white pine and western larch are locally important. In the Southwest piñon and juniper constitute widespread low-elevation scrub woodland between the ponderosa-pine region and the desert.

In the East a mixed hardwood forest is found everywhere (characterized by oak and hickory mixed with other broad-leaved trees) except in the northern tier of states, where maple, beech and birch constitute the Northern hardwood forest. In the drier sandy and rocky areas, however, pines grow well, particularly in the broad belt of the coastal plain and the Piedmont region extending from Virginia to Texas, where pines (loblolly, shortleaf, slash and longleaf) make up the Southern pine forest, the most economically important forest in the East. In the Northern states other pines (white, red and jack) are locally important. The same is true in the lake states and northern New England of species that extend out from the Canadian boreal forest: spruce, fir, aspen and birch. In the Southeast bottomland forests of gum, bald cypress and oak are characteristic.

The experience of people living in industrialized countries does not suggest that nearly half of the wood felled in the

MAN-MADE FOREST in New Zealand is managed by the New Zealand Forest Service. It is Kaingaroa State Forest, near Rotorua. Planted some 40 years ago on approximately 300,000 acres of open land, it is one of the largest man-made forests in the world. The major species that was planted in it is the Monterey pine (*Pinus radiata*), which occurs naturally only in California. The trees are harvested at between 20 and 30 years of age for a variety of uses. The recently clear-cut area at the left center is scheduled to be replanted soon.

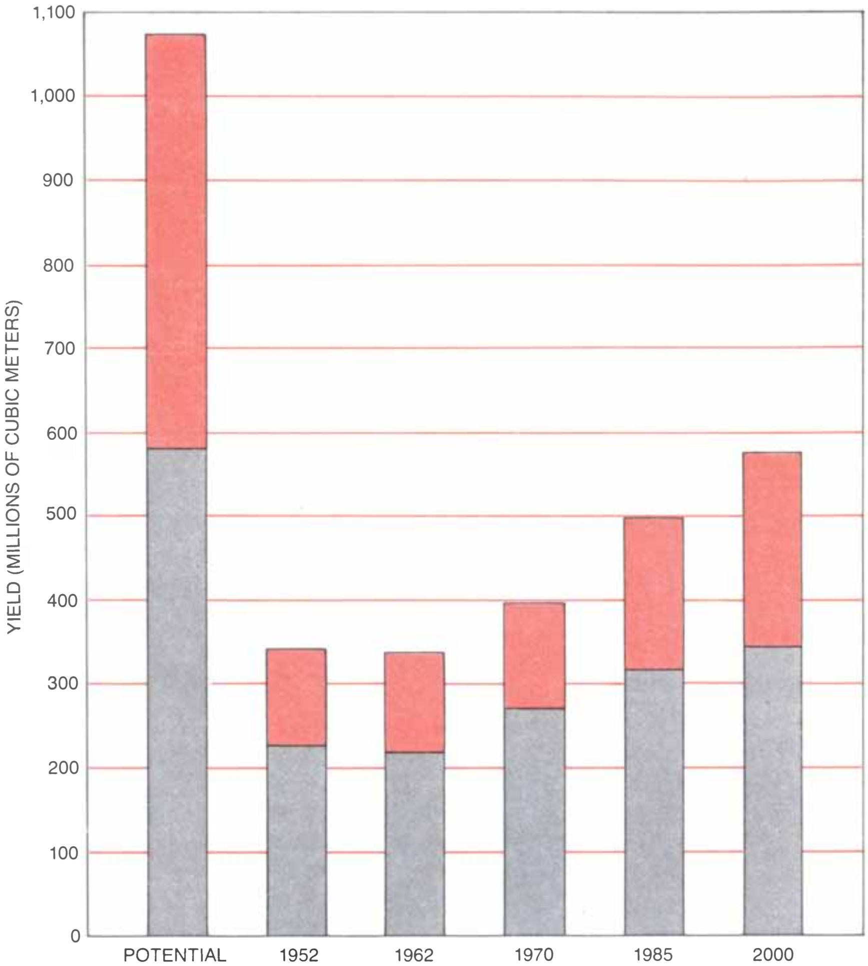
SILVICULTURAL MANAGEMENT is evident in this forest near Umtali, Rhodesia, operated by the Rhodesia Forestry Commission. The trees are pines (*Pinus patula*) that were originally planted at a fairly heavy density of 680 trees per acre. They were pruned at the ages of four and five years and thinned at the age of four years, which was about two years before the photograph was made. Additional thinnings will be made at ages 12, 18 and 25 years. The trees will be harvested, mainly for timber, when they are from 30 to 35 years old.

PINE SEEDLINGS are grown in a nursery in Thailand for transplantation into a forest. The seedlings are *Pinus kesiya* pines, a species native to Thailand. They are being grown as part of a program administered by the Royal Forest Department to develop large forest plantations on agricultural land that has been abandoned and on forestland that has been cut over. Age of the seedlings is about six months, meaning that they are almost ready for planting. The government's silvicultural objective is managed forests that will yield pulpwood.

EFFECT OF INTENSIVE SILVICULTURE is apparent in these cross sections of two western hemlocks (*Tsuga heterophylla*). The trees were approximately 20 years old at the time of cutting. For all but three years they received little attention; then the stand was thinned and fertilizer (urea) was applied by hand. The resulting rapid increase in growth is shown by the three outer rings. The trees grew in the state of Washington in a naturally seeded wild stand. The mark in the cross section at the right was caused by a branch that grew at that point.

world is consumed as fuel, but that is the case. Of the total wood harvest of slightly more than two billion cubic meters per year, 990 million goes for fuel; the rest is put to use in industry. At present wood serves as fuel mainly in the less industrialized countries, where it is burned directly. It is quite possible that the industrialized countries will turn increasingly to wood as fuel, not so much by burning it directly as by making more use of bark and industrial by-products and by converting wood or wood residues into methanol, gas and charcoal.

In industry wood serves not only as a source of structural products (lumber, plywood, particleboard, flakeboard, fiberboard, insulating board) and fiber (the basis of paper, cardboard and many other pulp products) but also as a source of extractive materials (oils, tannin, naval stores, drugs) and chemicals (sugars, alcohols, phenols). In the U.S. about 63 percent of the 250 million tons (226 million metric tons) of raw wood processed annually by the forest-products industries goes into primary structural materials and 35 percent into fiber-based products. The Forest Service has predicted that demand in the year 2000 will be double what it was in 1970, with about half of it going into structural products and 46 percent into fiber-based products.



FOREST PRODUCTION in the U.S. as achieved in 1952, 1962 and 1970 and as projected by the U.S. Forest Service for 1985 and 2000 is compared with potential growth, which is what would occur if all commercial forestland were fully covered with trees and the forests were managed by such conventional extensive silvicultural techniques as systematic harvesting and thinning of trees. Chart shows coniferous trees, or softwoods (*gray*), and hardwood trees (*color*).

Forest Management

All forests are affected to some degree by human actions. In the broadest sense all such effects could be termed forest management. Even in the national wilderness areas of the U.S., where cutting, road construction and the use of motorized vehicles are prohibited by law, human activity touches the forest in many ways. The global burning of fossil fuels increases the amount of carbon dioxide in the air. People introduce exotic plants and animals to the region and change the balance of wildlife by killing predators. Even inactivity can influence the forest, as when a decision is made not to fight a particular forest fire or not to deal with a particular epidemic caused by insects or disease.

Until about 1950 the history of the American forest had been largely one of exploitation, that is, the forest had been treated as a nonrenewable resource. The American Indian used fire to drive game into the range of his bow and arrow. The European colonist depended on the forest for fuel and lumber but at the same time cleared much of the eastern U.S. for farmland. The land that was too infertile or too steep and rocky for farming was logged repeatedly for lumber and for fuel to heat homes, smelt iron and run locomotives. Only with the opening of the coal deposits in the Appalachians did coal supplant wood as the basic source of energy. By then, however, little of the precolonial forest remained in the East. Today it has virtually disappeared except for remnants that are protected in state parks and other reserves. About 15 million acres (5 percent) of the 359 million classified as commercial forestland in the East (east of the Great Plains) do not have enough trees to be regarded as forest.

As the West was opened early in the 19th century many Eastern farms were abandoned. The unused fields and pastures seeded into pines and other pioneer tree species, giving rise to a second forest that reached commercial size early in this century. Hardwoods that had seeded in under the conifers took over when the conifers were harvested. As time passed shade-tolerant hardwoods gradually crowded out the less tolerant species, thereby continuously changing the composition of the forest. Many variations of this old-field succession have resulted from infinite combinations of climate, site and the forest's history of fire, wind, logging and seed years (the intermittent times, usually about every third, fourth or fifth year, when this or that resident species of tree puts out a large number of seeds). As a result of these ecological forces, including human activity and inactivity, the present Eastern forest is quite different from the precolonial one. For the same reasons it will be quite different again several decades from now.

In the West the harvest of the old-growth, precolonial forests has proceeded apace. Extensive logging began on private lands—usually the lower, flatter and most productive ones—in about 1900 and in the national forests, which cover most of the mountains, after World War II. Today little of the old-growth forest remains on private land. Much of the second-growth forest on land owned by timber and paper companies has been put under management.

In the Western national forests, which constitute 55 percent of the commercial forestland west of the Great Plains and 50 percent of the nation's entire supply of standing sawtimber, the remaining old-growth stands are interspersed among other areas that have been burned or clear-cut and reforested. Middle-aged stands are less plentiful, since little logging was done before World War II; they owe their existence mainly to fires that occurred early in the century. In the ponderosa-pine national forests of the interior the tendency is to harvest trees quite selectively, since the species grows on relatively dry land that is likely to support only grass or sagebrush if the forest is clear-cut. In the fast-growing national forests of Douglas fir and associated species west of the Cascade-Sierra summits the practice is basically to clear-cut 40 acres or more in a patch. Some six million acres (5 percent) of the 129 million in the West that is classified as commercial forestland have no significant stock of trees.

Strategies of Silviculture

Until the 18th century silviculture was rudimentary and rarely practiced. The foresters of western Europe during the Middle Ages were mainly game wardens who patrolled hunting reserves for feudal landlords. Trees were treated as a readily available natural resource to be exploited without much regard for their replacement. Exploitation is still the rule in much of the world, particularly in the Tropics and in the boreal forests.

By the 18th century the need for fuel wood and lumber in western Europe gave rise to silvicultural practices aimed at the reforestation of cutover areas and the management of forests for the sustained production of wood and other values. This approach, still widely followed today, is termed conventional extensive silviculture. It emphasizes systems of harvesting mature forests so as to encourage natural regeneration. The technique also includes the thinning of growing stands to increase the growth in diameter of trees that will eventually be harvested.

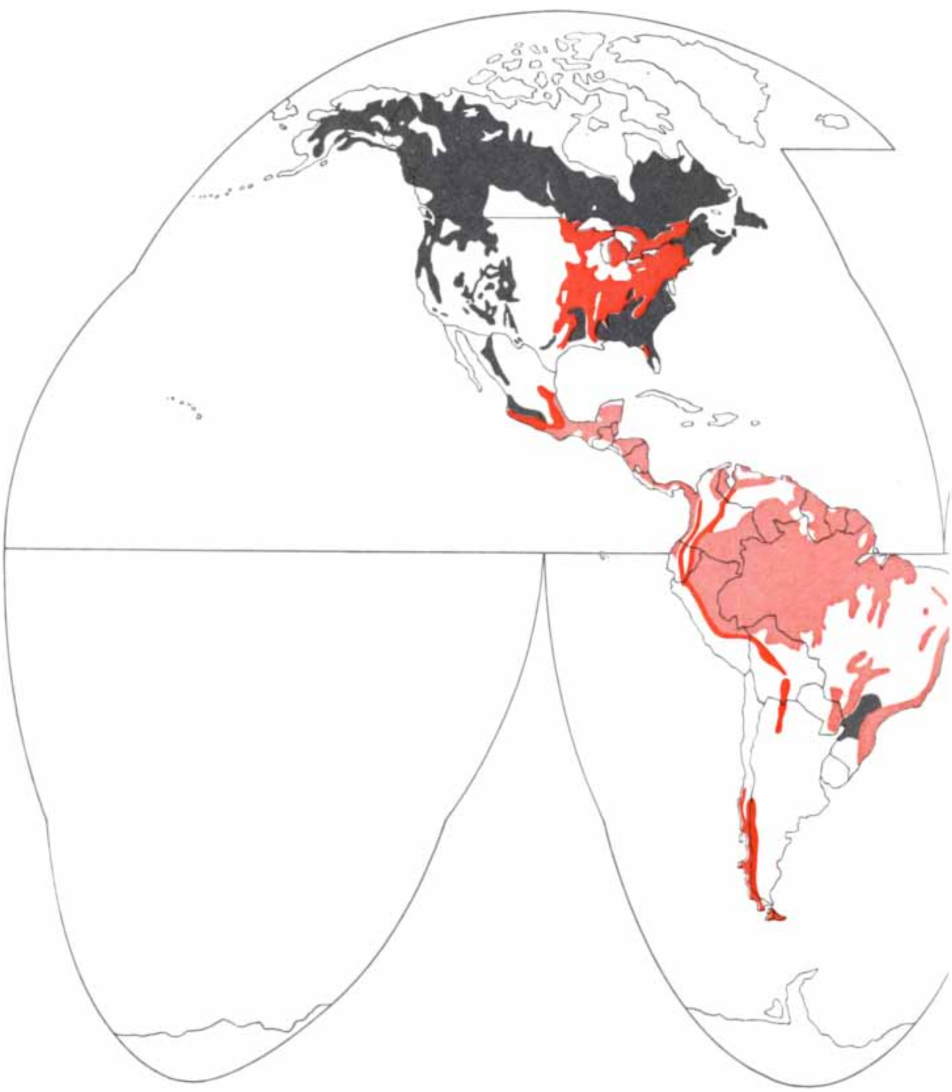
During the same period many of the hardwood forests in western Europe were converted into stands of softwood by the clear-cutting of the existing woodland and the planting of pure stands of Scotch pine and Norway spruce. Most such conversions succeeded, and so the growing of valuable commercial tree species in more or less pure stands became commonplace throughout much of the developed world. This technique, which can be called conventional intensive silviculture, is highly developed in the Douglas-fir forests of the Pacific Northwest, in the southern-pine region of the U.S. and in pine plantations of western Europe, South Africa, New Zealand, Australia and Chile. It is, in effect, tree farming.

By the middle of the 19th century the second and third crops of conifers were being grown on plantations in Germany, Switzerland and France. Professional foresters (by then the product of forestry training in universities) began to notice that yields were often lower than they had been for the first crop, particularly for spruce grown outside its normal habitat. The foresters attributed this decline to the formation of acid soil under forests consisting of only spruce or pine. (The needles of such trees have a low nutrient content and on falling to the ground form a raw humus that is quite acidic.) Among these foresters another concept of silviculture developed. It can be termed the naturalistic approach, in which attention is paid to the matching of trees to the ecology of the site. The concept also calls for the maintenance of a mixture of trees, including hardwoods because the leaves that fall from them will attract soil-building organisms and so help to maintain the fertility of the soil.

Finally, within the past decade the rising interest in the use of plant materials as a source of energy and of chemical feedstocks has led to much discussion of and some experimentation with the idea of short-rotation silviculture. In this concept trees that grow quickly are harvested as often as every two years.

The reader who has been counting my enumeration of silvicultural strategies will be aware that I have named five: exploitive, conventional extensive, conventional intensive, naturalistic and short-rotation. A number of intermediate approaches also exist. In each of the five basic strategies many fundamental questions remain unanswered.

For example, it can be argued that the exploitive approach is not really silviculture, since it involves only the cutting of trees without regard to the regeneration of the forest. Still, it is a form of forest management. From the standpoint of maximizing the economic return to the owner from the trees, such a harvest may well be justified on the grounds that the cost of reestablishing the forest may exceed the expected re-



WORLD'S FOREST RESOURCES are indicated on this map. The regions where such conifers as pine, spruce, hemlock, fir, redwood and cedar predominate are shown in gray; darkly

turn from the new stand of trees. (The U.S. national forests are operated under a Congressional mandate that they must be managed for a sustained yield; in them the exploitive approach is therefore illegal regardless of its economic merits.)

Under a policy of exploitation the harvest of trees is termed a commercial clear-cut if most of a stand is removed and a selective cut or a high-grading cut if only the most valuable trees are cut and a substantial number of less valuable ones are left. That number will depend on the age and structure of the forest as well as on economic considerations. Many of the trees that are left will fall because of wind or the weight of ice and snow once they have been deprived of support from the harvested trees and exposed to the elements.

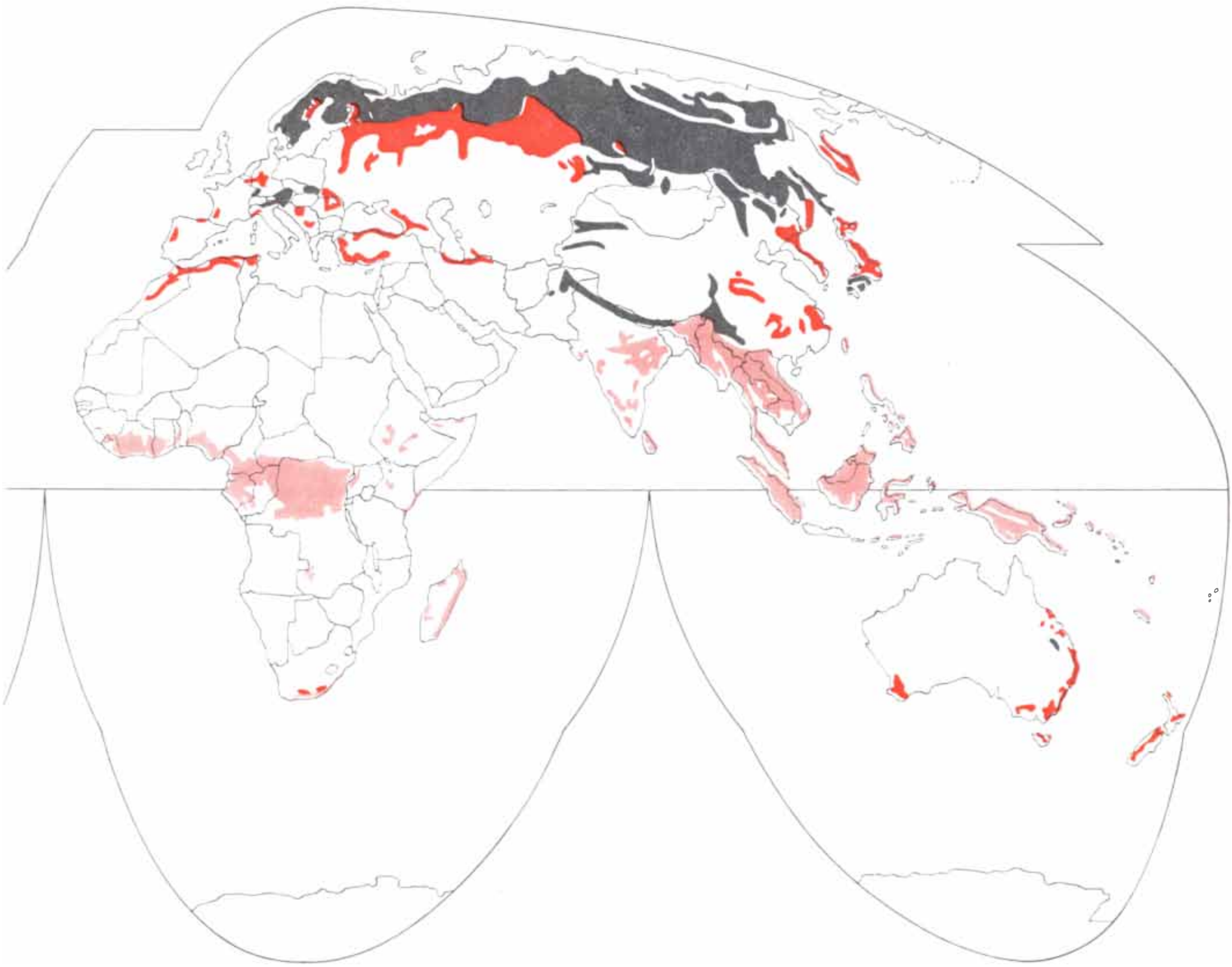
The disadvantages of the exploitive approach are well known and frequently stated. First, a large commercial clear-cut is an ugly wound in the landscape and will remain that way for at least several years. Second, the careless cutting of roads through the area and the indiscriminate use of heavy logging equipment may result in severe erosion. Third, much wood may be wasted through the destruction of trees that have grown for years but have not attained marketable size. Fourth, the impact on wildlife, water supply, recreation and other values is often severe. Finally, the regeneration of the forest is left to chance and may or may not occur.

An exploited forest can be left as it is in the hope that the surviving trees will be supplemented by natural regeneration, so that a more or less fully stocked second-growth stand eventually develops. More often than not, however, the best course will be to bulldoze the pitiful remains or spray them with herbicide, burn the site to reduce the organic remains to wood ash and replant with conifer seedlings big and hardy enough to form a new stand.

Conventional Silviculture

Two main practices have come to exemplify conventional extensive silviculture as it has evolved over the past two centuries from its beginnings in western Europe. One is the harvest of mature trees through silvicultural systems, which are cutting patterns designed to ensure the regeneration of trees in the cutover areas. The other is the technique of thinning and making other cuttings in order to free the potential crop trees from undesirable competition.

The emphasis in conventional extensive silviculture is on maintaining a sustained yield of wood. Under this strategy trees are grown until their boles are large enough to be sawed into lumber. In the national forests of the U.S., under



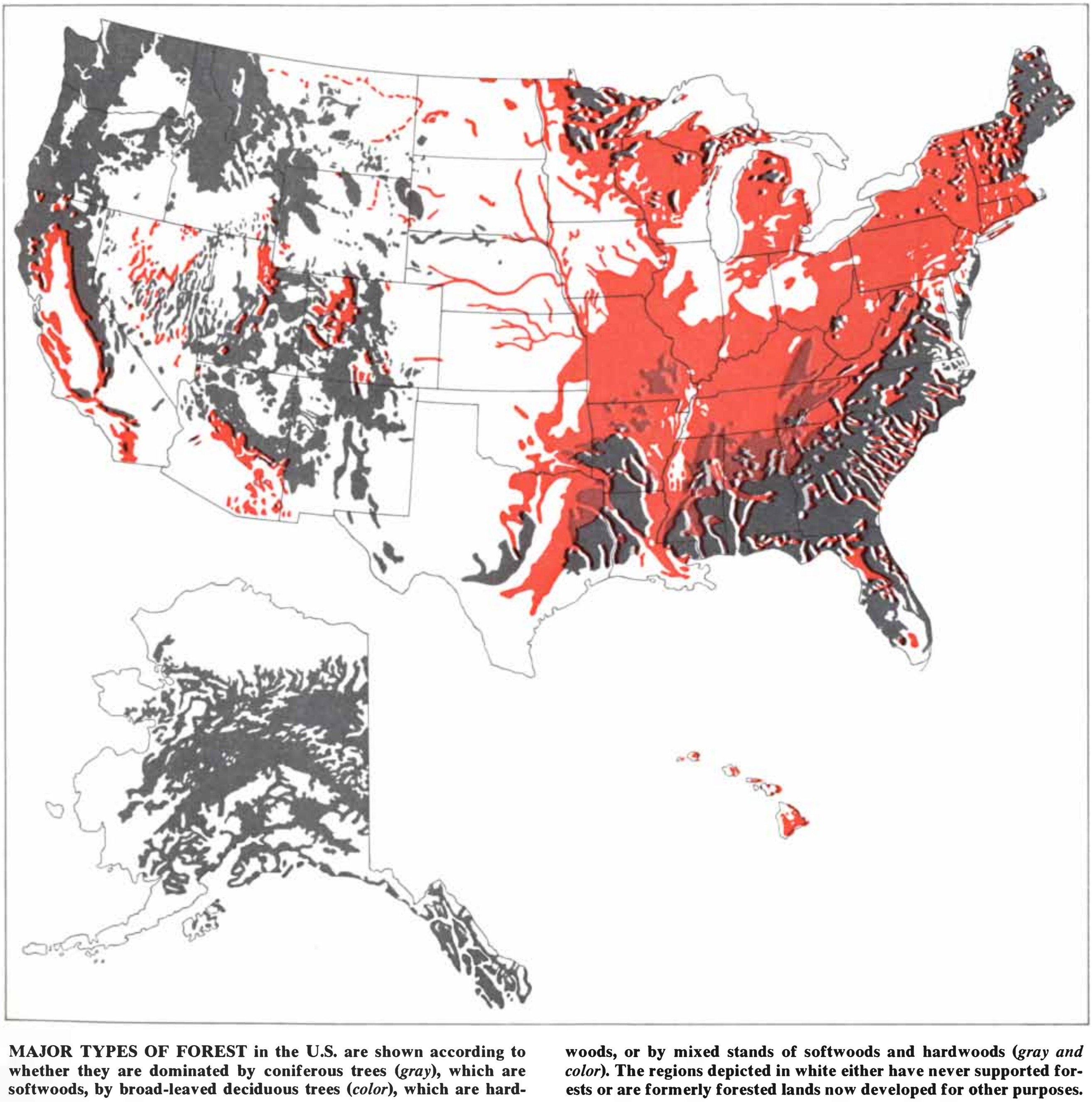
colored areas indicate the regions where temperate hardwoods including oak, hickory, maple, poplar and walnut predominate, and the lightly colored areas show regions where such tropical hardwoods as teak, mahogany, rosewood and ebony are among many species found.

policies adopted by Congress, the rotation age is set at the culmination of mean annual increment, that is, the age at which the trees of the forest are producing their largest average cumulative growth. In practice this policy means that trees are grown larger on public lands than on private ones, where economic pressures usually favor earlier harvesting. The rotation age for most commercial tree species in the U.S. tends to fall between 40 years (pulpwood harvested on private land) and 100 years (timber cut for lumber in national forests).

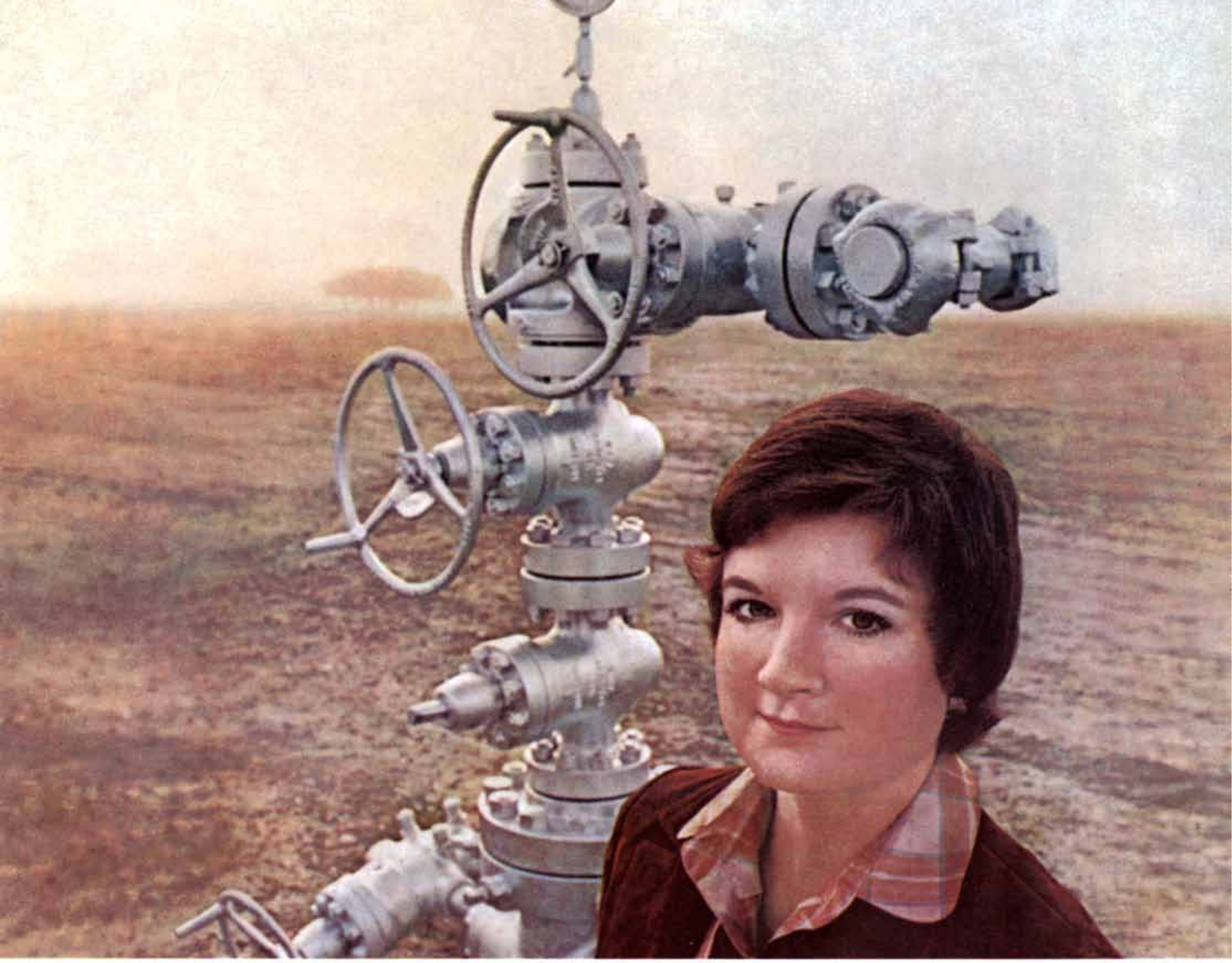
The silvicultural systems employed for harvesting mature trees can be grouped in four categories. One is the selection system, which consists of light partial cuts repeated every few years to create a forest of uneven age. The technique is suitable only for the relatively few commercial tree species that can thrive for most of their life in shade.

Most commercial tree species do best under full sunlight and must be grown in stands of even age. The other three silvicultural systems are designed to create such stands. One is clear-cutting, with or without subsequent planting or seeding but at least with thought given to natural regeneration and with care taken to minimize the damaging effects of the harvest. A second technique is to leave seed trees standing; it is not much employed because the trees left are likely to be brought down by wind or ice, they may not produce enough seeds to restock the forest and they may be so few in number that it is not worthwhile to return later and harvest them. The final system is called shelterwood cutting, because the mature stand is removed in two or more partial cuts so that the new stand can become established under the shelter of a partial canopy of remaining trees.

As for thinning, it is usually done in pole-size stands. From 25 to 40 percent of the stand may be cut, partly to harvest early the trees that appear likely to die and partly to increase the ultimate



MAJOR TYPES OF FOREST in the U.S. are shown according to whether they are dominated by coniferous trees (*gray*), which are softwoods, by broad-leaved deciduous trees (*color*), which are hardwoods, or by mixed stands of softwoods and hardwoods (*gray and color*). The regions depicted in white either have never supported forests or are formerly forested lands now developed for other purposes.



Twenty million years ago, the site of Marsha's discovery was a seashore. Now it's 40 miles from the sea, and half a mile down.

Marsha Green took a closer look at a 20-million-year-old seashore and found this new gas field.

For geologist Marsha Green, it was a case of reviewing the evidence just one more time and coming up a winner. After only two years with Gulf Oil, Marsha discovered a new and untapped gas field.

Spotted a trend

Her find was near Victoria, Texas, which 20 million years ago was a seashore, rich in the marine life that creates natural gas.

There were plenty of producing wells in this general area. The drilling record of each one tells a geologist something new about the earth's structure.

Marsha pored over these records. And she began to spot a trend.

Textbook curve

Marsha persuaded Gulf management to drill five exploratory wells. Four were dry. But on one, the instrument that detects natural gas went right off the scale. "A real textbook curve," Marsha says.

Marsha persuaded Gulf management to drill a series of exploratory wells.

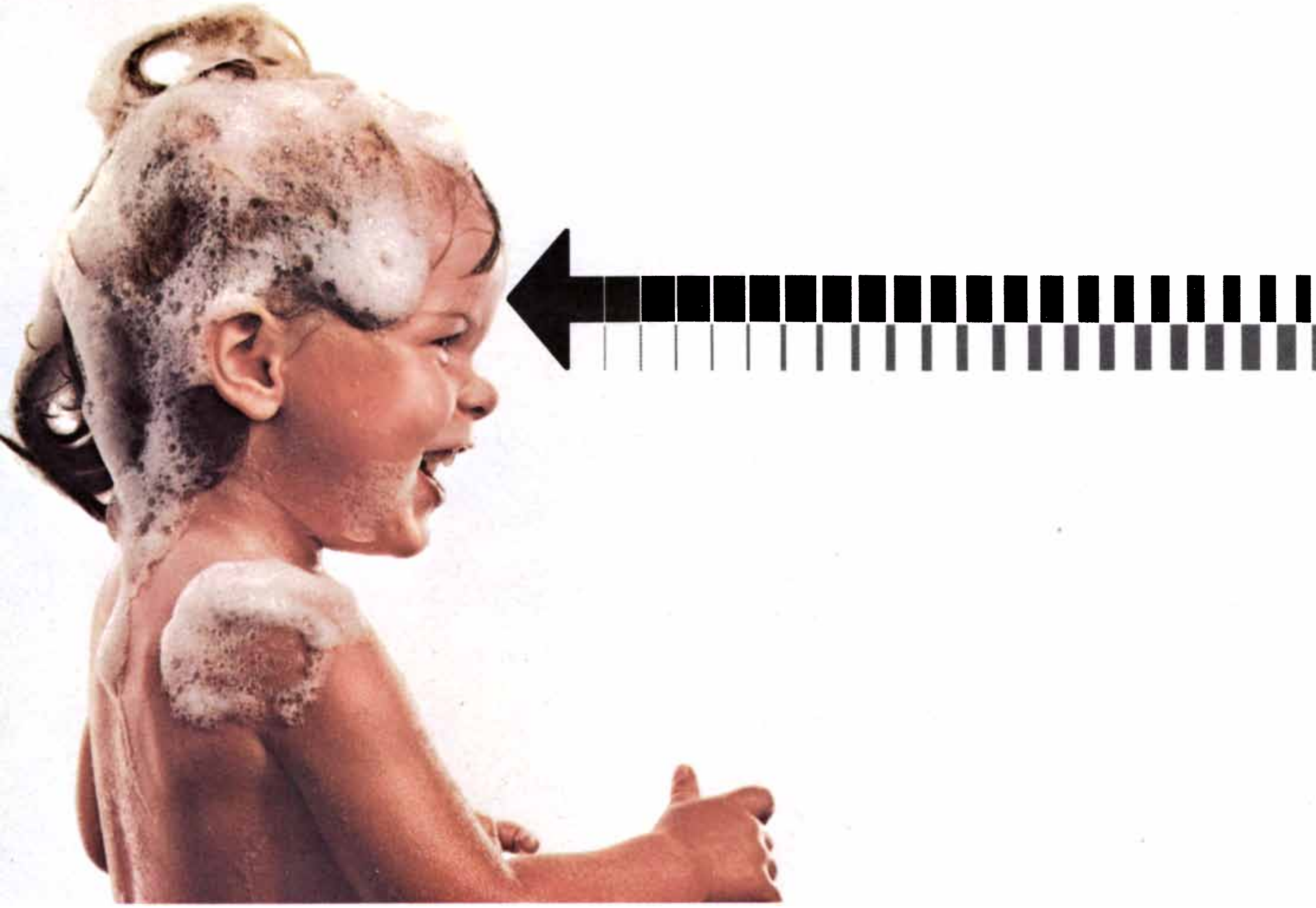
Such moments are rare. But Marsha is a dedicated professional. "I like to try to figure out what happened millions of years ago," she says, "and put the whole picture together."

Marsha Green. One of 52,000 Gulf people, meeting the challenge.

Gulf

Gulf people: meeting the challenge.

Gulf Oil Corporation.



At the press of one button, sound waves bounce off your subject and return to

Polaroid introduces SONAR automatic focusing.

Imagine taking pictures with a camera that does all the focusing for you automatically, instantly and perfectly–with *sound waves.*

A new era in photography is here.

Polaroid's revolutionary Sonar OneSteps are the first cameras ever introduced that focus by themselves using completely automatic ultrasonic technology.

At the press of one button, the new Sonar OneStep cameras automatically focus, set the shutter speed, the lens aperture, and make the exposure. In as little as a second and a half, you have the developing picture in your hand.

No other camera has ever given photog-

HOW THE BELL SYSTEM OVERSEES 40 MILLION LONG DISTANCE CALLS A DAY.

ON AN EASY DAY.

The Network Operations Center, Bedminster, New Jersey.

Part of the network status board, where NOC personnel watch for possible jam-ups.

You are looking at the Bell System's Network Operations Center. Here, our technology and people work 24 hours a day to help your long distance calls go through quickly, effortlessly.

When you make a long distance call, it has several different routes it can take, automatically.

But sometimes traffic gets particularly heavy. We can get a bottleneck.

That's when the people of the Network Operations Center move in. Using the most advanced Bell System computer technology, they re-route the traffic to get your call through

In round numbers, the Network Operations Center helps manage nearly 40 million calls, on a normal day. At busy times on busy days, the volume surges even higher.

So come Christmas or Mother's Day, hurricane or high water, virtually every long distance call you make goes through quickly and easily.

Thanks to all the people of the Bell System.

No matter when you make your long distance call, the NOC stands ready to help it get through without a hitch.

Bell System

Keeping your communications system the best in the world.

diameter of the remaining trees. Except for a few species that tend to stagnate under crowded conditions, thinning will not normally produce higher yields per unit of area but will improve the growth of the bigger and more valuable trees. A commercial thinning is economic when the trees that are removed can be sold for pulp, fuel or other purposes for which wood of small dimensions is appropriate.

Conventional extensive silviculture is widely practiced in western Europe, the U.S. and many other temperate regions. Since most of the silvicultural operations carried out under this heading pay for themselves in terms of wood products extracted, no substantial capital investment is required. The main cost is in foregone profits.

On the other hand, such operations may be insufficient to maintain a continuous yield. Natural regeneration may fail, so that the site has to be planted or seeded. Young trees of desired species may have to be freed from competition with unwanted species. Other investments may be needed to achieve a fully stocked forest and maximum growth. In this way silviculture becomes more intensive.

The growth and yield of trees can be substantially increased by the application of intensive silvicultural techniques. Even so, since it takes decades to grow a crop of trees, the economic incentives for applying the techniques may be less than enticing. Indeed, the investment made to establish a stand or to increase its growth may exceed the anticipated increase in the value of the final crop. Nevertheless, intensive silviculture, involving the investment of considerable amounts of capital in the forest, is becoming increasingly feasible and is being more widely practiced, particularly by industrial landowners seeking a continuous supply of wood for their lumber, pulp and veneer mills.

In intensive silviculture the landowner employs at least one and probably more of the following techniques: (1) Work is done on a cutover site to prepare the land for planting or to create surface conditions favorable to the natural regeneration of the stand. (2) Forests consisting of species that grow slowly or are of less value than another species that could grow on the site are removed by clear-cutting, and trees of the more desirable or faster-growing species are planted. (3) Stock that has demonstrated improved genetic characteristics is planted. (4) Herbicides are employed to reduce unwanted competition, mainly from hardwoods of limited commercial value. (5) The stand is thinned before maturity to counter stagnation and to stimulate the growth of potential crop trees. (6) Fertilizer is applied to the stand to compensate for deficiencies of nitrogen and phosphorus. (7) Pesticides are applied on prescription to reduce serious threats from insects and fungi.

Consider the potential of just one of these treatments. The Forest Service estimates that gains in yield of from 8 to 15 percent in one generation could be achieved through the genetic selection of Douglas fir in western Oregon and Washington. On the better sites the yield could be doubled by the third generation. Already gains of from 10 to 20 percent in volume and in quality are being realized in the first generation from the genetic improvement of Southern pines.

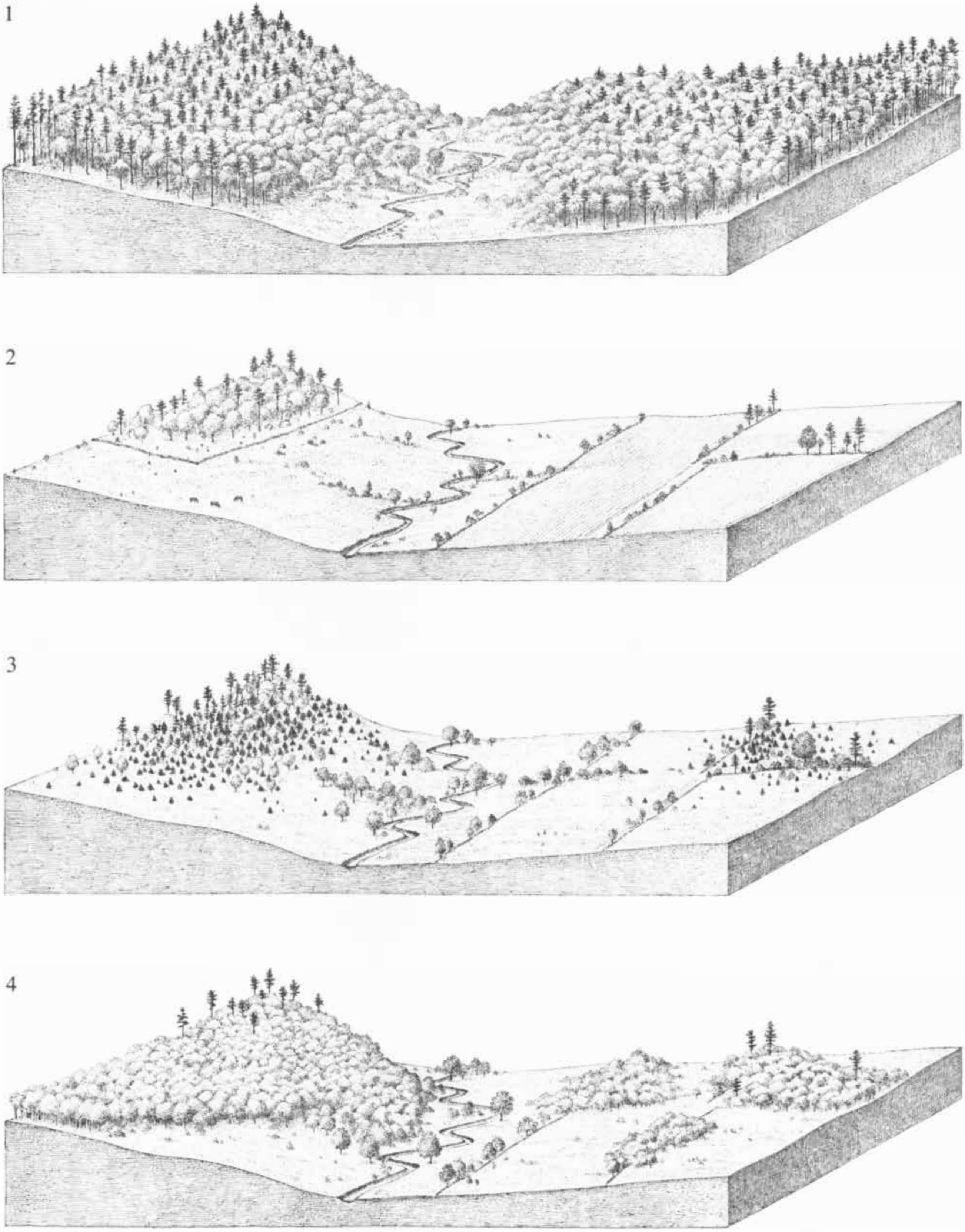
Other Techniques

The naturalistic approach to silviculture is typified by practices in Switzerland. Moving away from the concept of pure stands of even age and from strong-arm methods of cultivation, the Swiss have come increasingly to rely on natural tree types and natural successional trends as a basis for their silviculture. The mixed stand of uneven age has become the ideal.

The main arguments for this technique are that it requires little capital and entails few risks. Once a forest is established along naturalistic lines it is maintained by periodic light partial cuts designed to produce the highest economic return consistent with the maintenance of a mixed, continuous forest of uneven age. The site does not deteriorate, and the generally quick establishment of acceptable species reduces silvicultural investments to a minimum. A naturalistic forest is less likely than a pure stand to be susceptible to attack by disease and insects.

Certainly such a forest can be productive and profitable for some species of tree on some sites. It is equally certain that an even-age forest of the same species can be grown and that it will often produce more biomass and higher economic returns than the naturalistically managed forest. One simply cannot generalize. Both naturalistic silviculture and conventional silviculture have a place in forest management. What to grow and how to grow it are decisions that must be made for a particular site according to the conditions of climate and soil that exist there.

The short-rotation approach to silviculture consists in producing large quantities of tree material by selecting highly efficient photosynthesizers and growing them under intensive cultivation with frequent harvesting. The proposal to develop "biomass plantations" is essentially a return to a form of the coppice system that was employed extensively in Europe and the eastern U.S. during the 19th century to produce fuel



OLD-FIELD SUCCESSION indicates what happens when a forest is essentially unmanaged. The succession is portrayed here for a generalized forest in the East, beginning (*1*) with the mixed conifer-hardwood forest that existed when the colonists arrived. Gradually the colonists cut down nearly all the trees, at first for fuel and lumber but later to clear the land for farming (*2*). With the opening of the West many Eastern farms were abandoned; by natural regeneration they grew a second forest that consisted mainly of pines or cedars but included a few hardwoods (*3*). When the conifers were harvested (*4*), hardwoods took over. In the present stage shade-tolerant hardwoods are crowding out the less tolerant species, giving rise to an oak-hickory climax forest that will be quite different from the precolonial one.

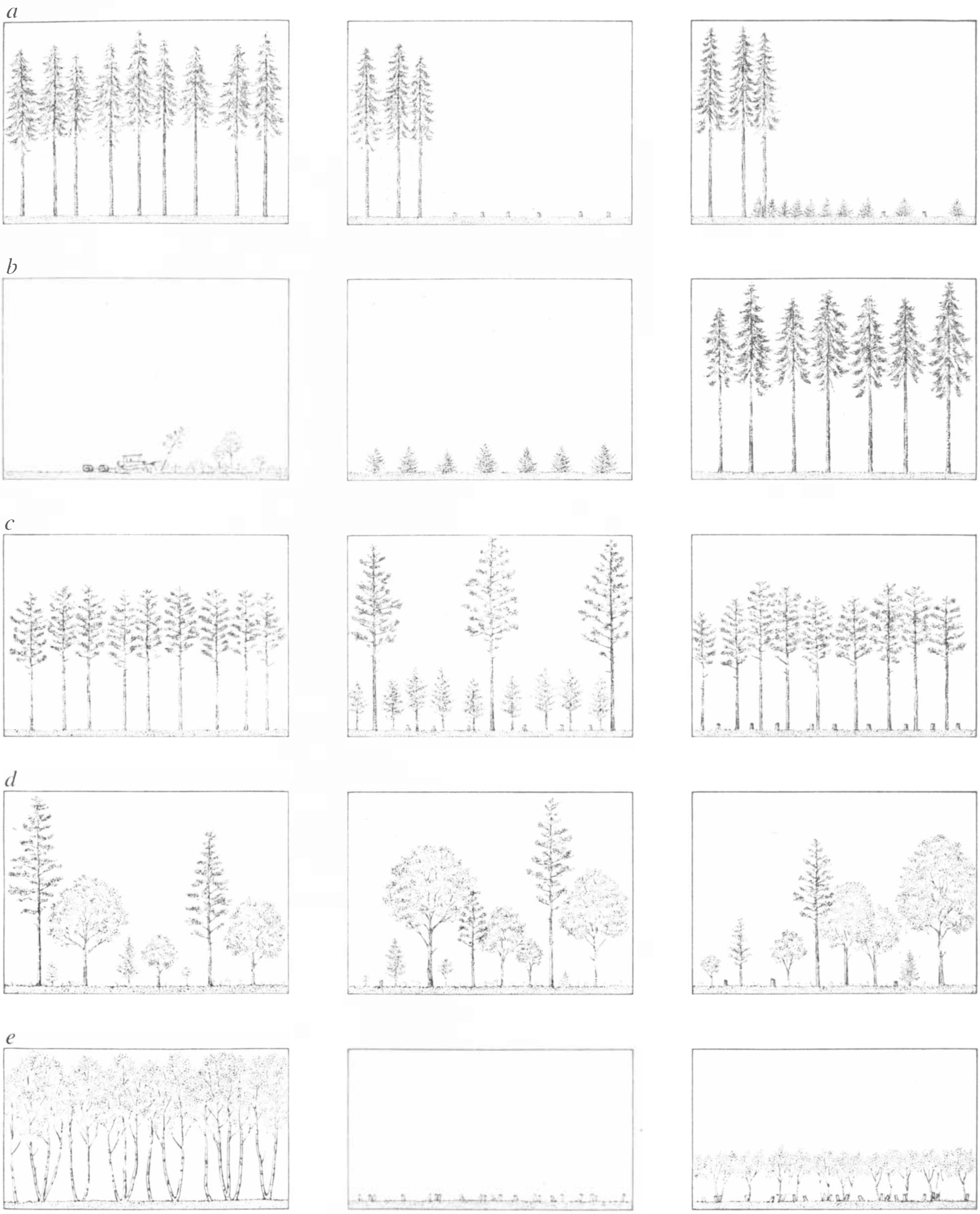
from stands of hardwood that were harvested on short rotation and regenerated by sprouting from stems. In the current manifestation the strategy is to rely on such fast-growing species as cottonwood, sycamore, red alder and aspen, to harvest the stands mechanically on rotations of from two to four years (when the trees are from 10 to 20 feet high and from one to two inches in diameter and can be converted into chips for fuel or chemical feedstock), to rely on sprouts from the remaining stems for reproduction (or on root suckers with aspen) and to cultivate and fertilize the sprouts intensively. Yields of several tens of tons of oven-dry plant material per hectare per year are predicted on the basis of the physiological potential of the trees and of studies made on small plots.

If such yields could actually be attained over broad areas without heavy investments of energy in fertilization, irrigation, cultivation and harvesting, short-rotation silviculture would indeed have considerable promise. Unfortunately it has not yet been possible to grow hardwoods in plantations with any great success, let alone to grow them at the rates achieved in trials of genetically superior stock under intensive care in the nursery or some other carefully monitored condition. This is not to say that the experimental work has been unsound but rather that it has not been possible to transfer the results to operations of commercial size.

The Silvicultural Outlook

Forests have a considerable potential for producing energy and chemical feedstocks. For example, if the energy value of wood is taken to be 7,500 British thermal units per pound, the annual net primary production of the world's forests is roughly 1,200 quads (a quad is one quadrillion B.t.u.) or five times the world's use of 230 quads of energy in 1975. Robert L. Burgess of the Oak Ridge National Laboratory has estimated (on the assumption that the annual aboveground growth of the world's forests is 13 billion metric tons, of which half is economically accessible) that one million megawatts of electricity could be produced from wood on a sustained-yield basis—about twice the present generating capacity in the U.S.

In 1970 the annual growth rate of the commercial forests in the U.S. was estimated at 2.7 cubic meters per hectare (38 cubic feet per acre). If all the commercial forest areas were fully stocked with trees and balanced in age classes, the potential growth would have been 5.2 cubic meters per hectare per year. The biological productivity of the commercial forestlands of the U.S. is such that the net realizable growth of the forests could be doubled within half a cen-



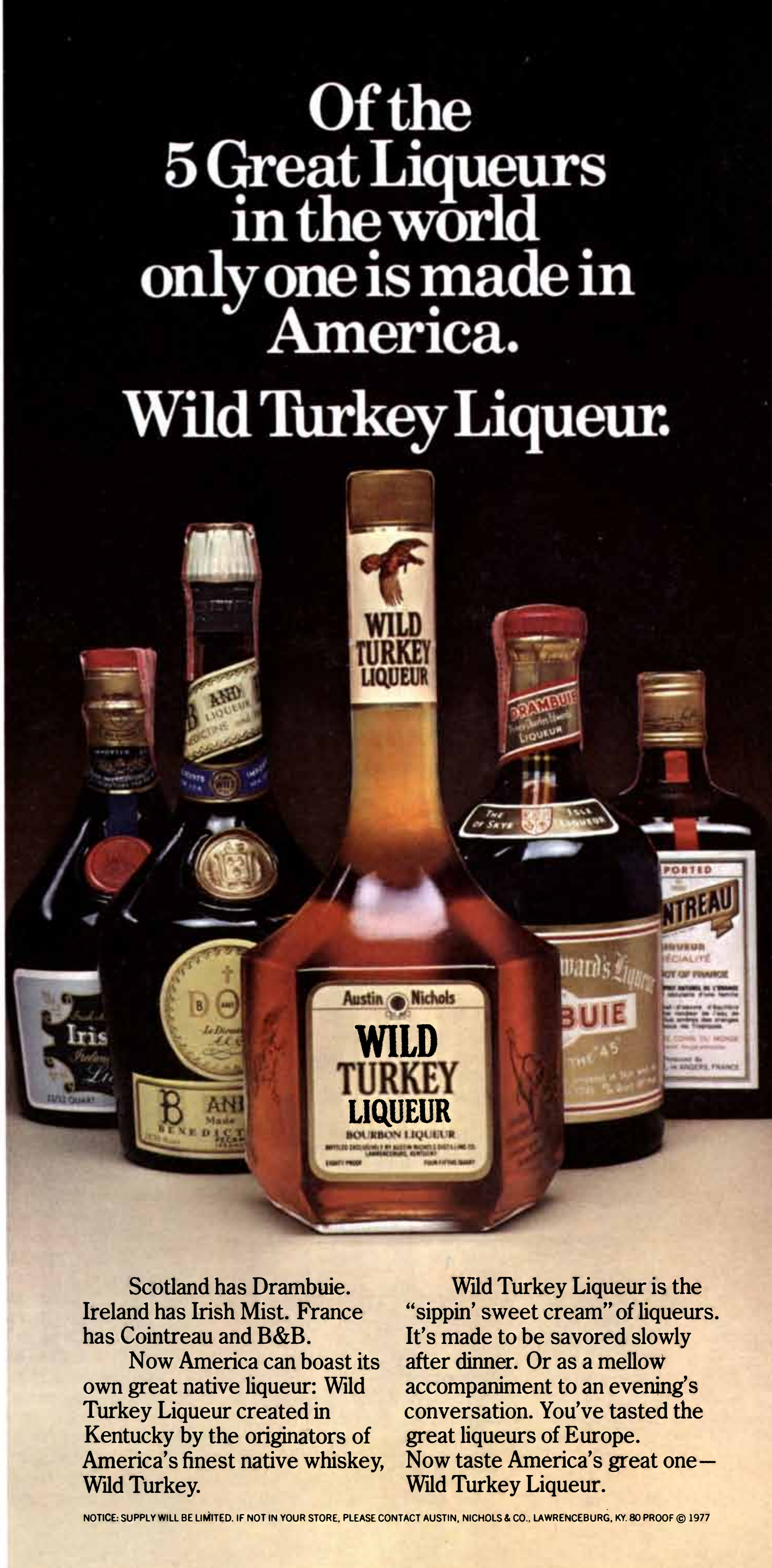
SYSTEMS OF SILVICULTURE are depicted. In clear-cutting with natural regeneration (*a*) a stand of trees is entirely cut down in a given area. Regeneration of the forest may occur through the natural fall of seeds from trees that remain standing near the clear-cut area. In clear-cutting with planting (*b*) a cleared site may be prepared for the planting of genetically improved seedlings, which grow into a pure stand of even age. The forest is thinned at intervals and treated with fire, fertilizers, herbicides or pesticides as needed. In the shelterwood system (*c*) a stand is cut in such a way that trees remain in the area to provide seed and shelter for a period of years and thereby to make regeneration fairly certain. Once the new growth is established the overstory is removed to leave a young stand of even age. The selection system (*d*) relies on the maintenance of a mixed stand of uneven age and trees that are more or less shade-tolerant. In the proposed short-rotation method, a variety of the coppice system (*e*), trees that grow rapidly and regenerate by sprouting would be intensively cultivated and harvested by machines on rotations as short as two years. The harvested trees would be converted into wood chips for fuel or chemicals.

tury by the widespread application of proved silvicultural practices. With a widespread application of intensive silviculture, including the complete use of hardwoods and the utilization of the entire tree of every harvested species, the potential productivity of the forests in the U.S. would be closer to three times the present level.

What are the possibilities of improving on demonstrated silvicultural practices through research? In the U.S. research on silviculture and on related aspects of forestry has largely paralleled work in agriculture. It is carried out by the Forest Service at 82 laboratories, by workers at some 60 universities (mostly with Federal funding) and by a limited number of wood-using industries. The present level of activity is about 400 scientist-years devoted annually to research relating to the biology, cultivation and management of forests, to the genetics and breeding of forest trees and to the economics of timber production. For more than 50 years extensive studies in the laboratory and in the field have dealt with the ecological and physiological requirements of the principal tree species. The point has been reached where silviculturists understand the interaction of these trees with their environment, including the major insects and diseases that form part of the forest ecosystem.

Several accomplishments can be cited. Two of them have been described here: the genetic improvement of the Southern pines and the flow of nutrients, water and other components of a forest ecosystem in monitored forested or clear-cut watersheds. Useful work has also been done on the atmospheric physics of weather as it affects the behavior of forest fires and as it relates to the modification of weather and the control of lightning; on the analysis of aerial photographs and other forms of remote sensing as a means of obtaining inventories of large forest areas (including the forest resources of the nation as a whole), and on the use of linear-programming models to determine and schedule tree harvesting and other forest-management operations.

On the other hand, not enough work has been done on a number of problems that have become major issues in the encounters between environmentalists and forest managers. How do forests of even age and uneven age (forests composed of different tree species and growing on different sites) compare in productivity and loss of nutrients? How do pure and mixed stands compare on these matters and on relative vulnerability to insects, diseases and other hazards? The U.S. is sadly lacking in definitive research studies that could have a broad application in the shaping of national policies on forests.



Solitons

They are waves that do not disperse or dissipate but instead maintain their size and shape indefinitely. A recent finding is that solitons may appear as massive elementary particles

by Claudio Rebbi

Waves and particles have been intimately related in physical theory since the formulation of quantum mechanics in the 1920's. In the past few years another connection between them has emerged. It comes from a surprising source: the analysis of certain wave equations that are not part of quantum mechanics but instead derive from classical physics. Solutions to these equations describe waves that do not spread or disperse like all familiar waves but retain their size and shape indefinitely. The wave can be regarded as a quantity of energy that is permanently confined to a definite region of space. It can be set in motion but it cannot dissipate by spreading out. When two such waves collide, each comes away from the encounter with its identity intact. If a wave meets an "antiwave," both can be annihilated. Behavior of this kind is extraordinary in waves, but it is familiar in another context. Given a description of an object with these properties, a physicist would call it a particle.

Waves that propagate without dispersing have been known for some time in hydrodynamics, where they are called solitary waves or solitons. What has been discovered recently is that nondissipative waves also arise from some of the equations formulated to describe elementary particles. The name soliton has been borrowed for these objects. The solitons I shall discuss here are those that share a particular mechanism of confinement: they are prevented from dispersing by a topological constraint. They cannot decay by spreading out for the same reason a knot tied in an endless rope cannot be removed without cutting the rope.

For now the solitons of particle physics are entirely the creation of theorists, and it may yet turn out they do not exist in nature. On the other hand, if the equations that describe elementary particles are found to be among those that admit soliton solutions, then the solitons should appear as new particles. They would be very massive, perhaps thousands of times heavier than the proton. A soliton particle would also have certain distinctive properties; for example, one theory predicts that each soliton would be a magnetic monopole, an isolated north or south magnetic pole.

Even if such particles do not exist, solitons may enter the realm of elementary-particle physics in another role, as objects confined not only to a definite region of space but also to a moment in time. Such evanescent solitons have been named instantons. Like the soliton, the instanton is a classical object with a quantum-mechanical interpretation. It is viewed not as a particle but as a transition between two states of a system, a manifestation of the phenomenon called tunneling. Transitions facilitated by instantons have already been invoked to explain a pattern of particle masses that had been a long-standing puzzle to theorists.

In order to understand what is remarkable about a wave that does not disperse, one need only consider an ordinary, dissipative wave such as the one generated when a pebble is dropped into a still pond. Such a wave travels over the water surface as an expanding ring. Careful observation reveals that the disturbance becomes less pronounced as it moves away from the source, and eventually it dies out completely.

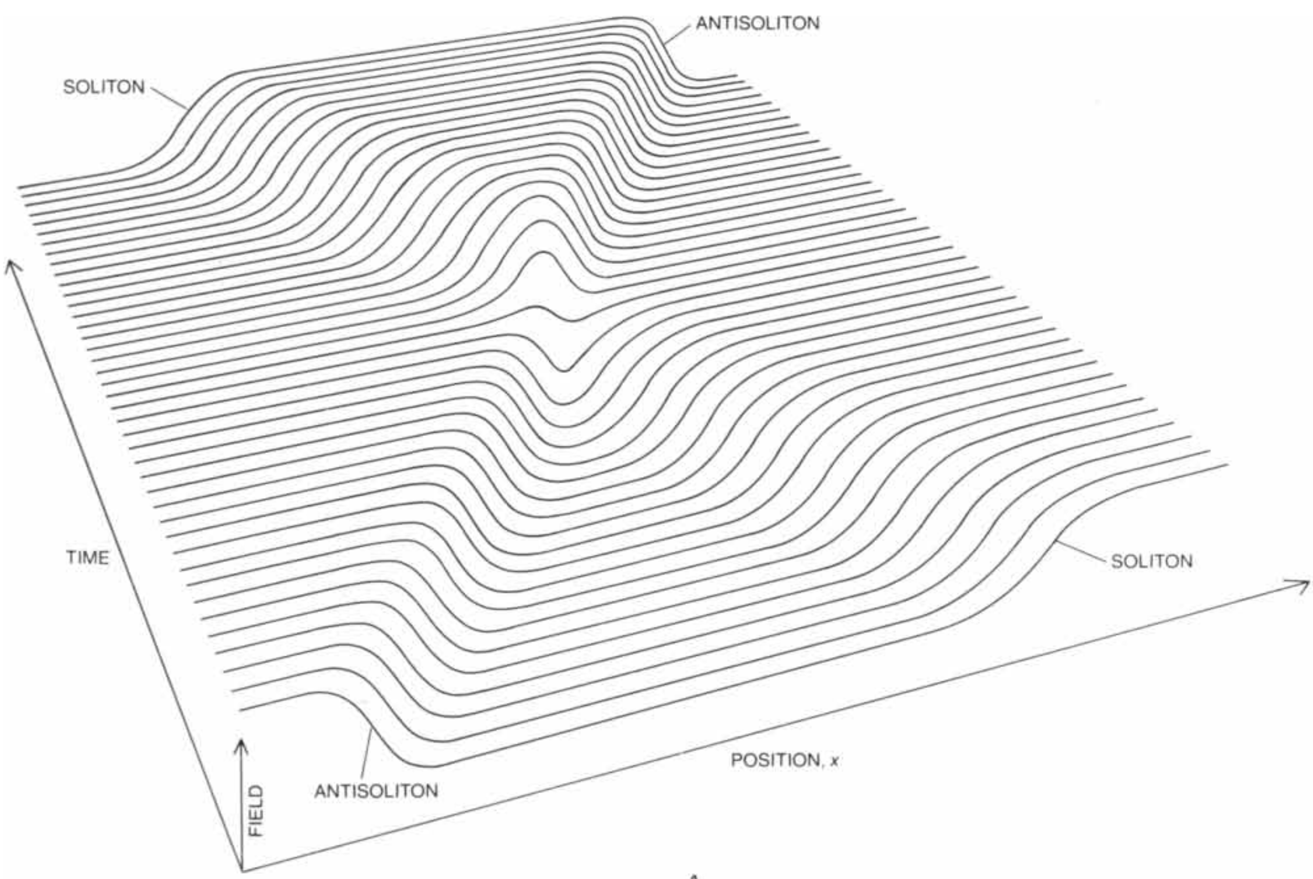
An important factor in the decay of water waves is the viscosity of the medium, a manifestation of friction. In some mediums, however, waves propagate without friction but nonetheless decay, and even waves in a pond filled with a frictionless fluid would die out. The reason is that components of the wave that have different wavelengths propagate with different speeds, spreading the energy of the wave over a larger area. When the wave is viewed in cross section, it becomes continually broader but of smaller amplitude. If it is allowed to continue to infinity, it disperses over an infinite area and thereby disappears.

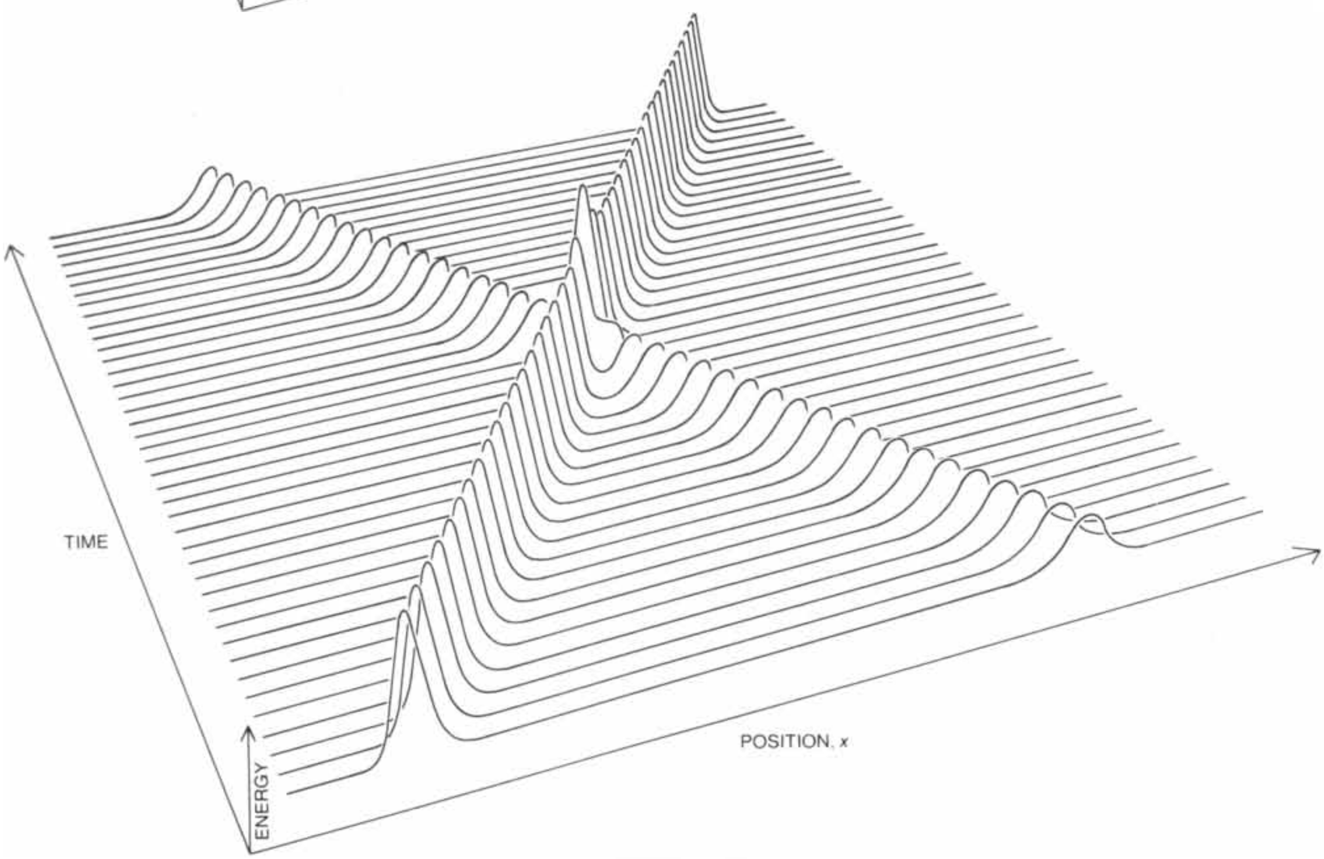
Solitons are not, strictly speaking, immune from dispersion; rather, they are waves in which the effects of dispersion are exactly canceled by some compensating phenomenon. The compensation is possible only in a certain class of waves, those whose equation of motion is of the kind said to be nonlinear. The propagation of such a wave is influenced not only by the shape of the disturbance but also by its magnitude.

The first recorded observation of a soliton was made almost 150 years ago by John Scott Russell, an engineer and naval architect. He reported to the British Association for the Advancement of Science: "I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped—not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large, solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles per hour, preserving its original figure some 30 feet long and a foot to a foot and a half in height. Its height gradually diminished, and after a chase of one or two miles I lost it in the windings of the channel."

Scott Russell proposed that the stability of the wave he had observed resulted from intrinsic properties of the wave's motion rather than from the circumstances of its generation. This view was not immediately accepted. In 1895, however, D. J. Korteweg and Hendrik de Vries gave a complete analytic treatment of a nonlinear equation in hydrodynamics and showed that localized, nondissipative waves could exist. Solitons have since become a well-recognized phenomenon in various fields of engineering and applied mathematics.

The solitons of interest in particle physics arise from equations that describe fields, or physical systems that are extended in space. A field assigns to every point in space a value of some speci-

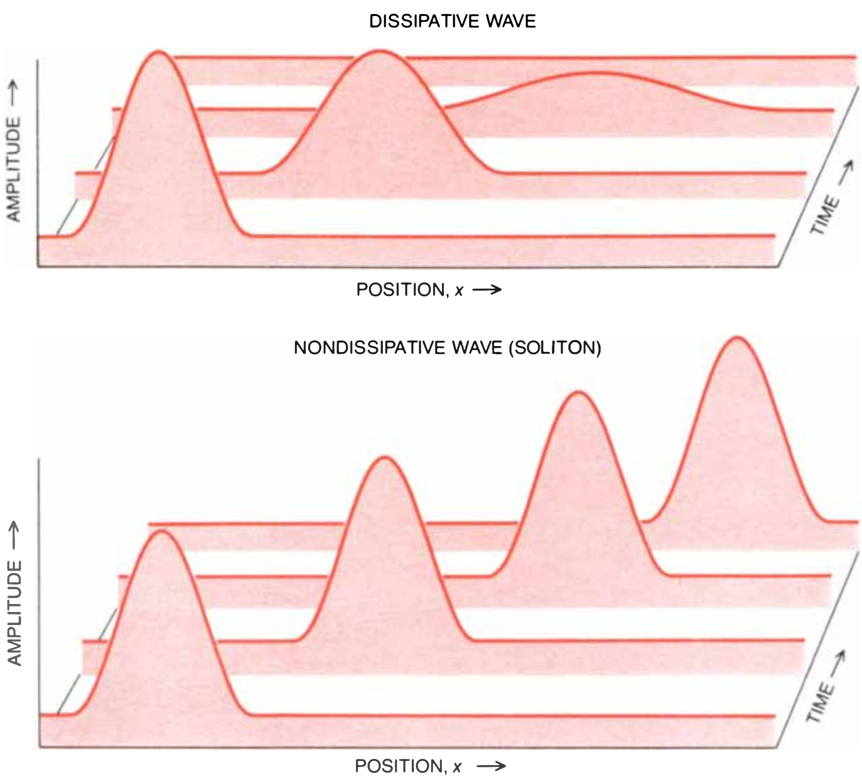




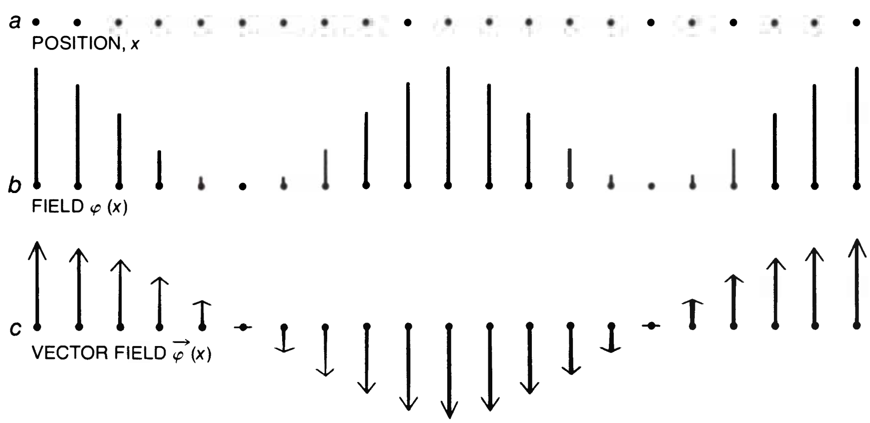
SOLITON AND ANTISOLITON appear as kinks of opposite direction in the structure of a field. In the upper illustration the evolution of the field itself is depicted; proceeding from front to back, each line represents a successive configuration of the field. The lower illustration shows the energy distribution corresponding to each configuration. At the beginning of the sequence the soliton and the antisoliton are approaching each other, the antisoliton moving faster and therefore having a higher energy. When the two waves meet, they are both annihilated and from their energy another soliton-antisoliton pair is immediately created. The one-dimensional field in which the waves propagate is called the sine-Gordon field. Everywhere outside the soliton and the antisoliton the field takes on values with zero energy.

fied quantity, such as electric potential. Often more than one quantity is defined at each point. The values can change from point to point and from moment to moment, but they must do so smoothly and continuously.

The most familiar field is the electromagnetic one described by the field equations of James Clerk Maxwell. In Maxwell's theory six values are assigned at each point in space; they represent the components of the electric and the magnetic fields along any three orthogonal axes. Gravitation is also described by field equations, those of Einstein's general theory of relativity. The surface of a pond can be regarded as a model of a two-dimensional field. The quantity to be specified at each point is vertical position, or height above some reference level. A wave passing through the water then becomes a perturbation of this field.



PERSISTENCE OF A WAVE is limited mainly by dispersion, the process whereby components of a wave that have different wavelengths propagate at different speeds. Because of dispersion, an ordinary wave (*upper graphs*) tends to flatten and spread as it moves and must eventually die out entirely. A soliton (*lower graphs*) is a wave that does not dissipate because the effects of dispersion are canceled by other features of the wave's motion. The soliton represents a quantity of energy that can move from point to point but cannot spread out in space.



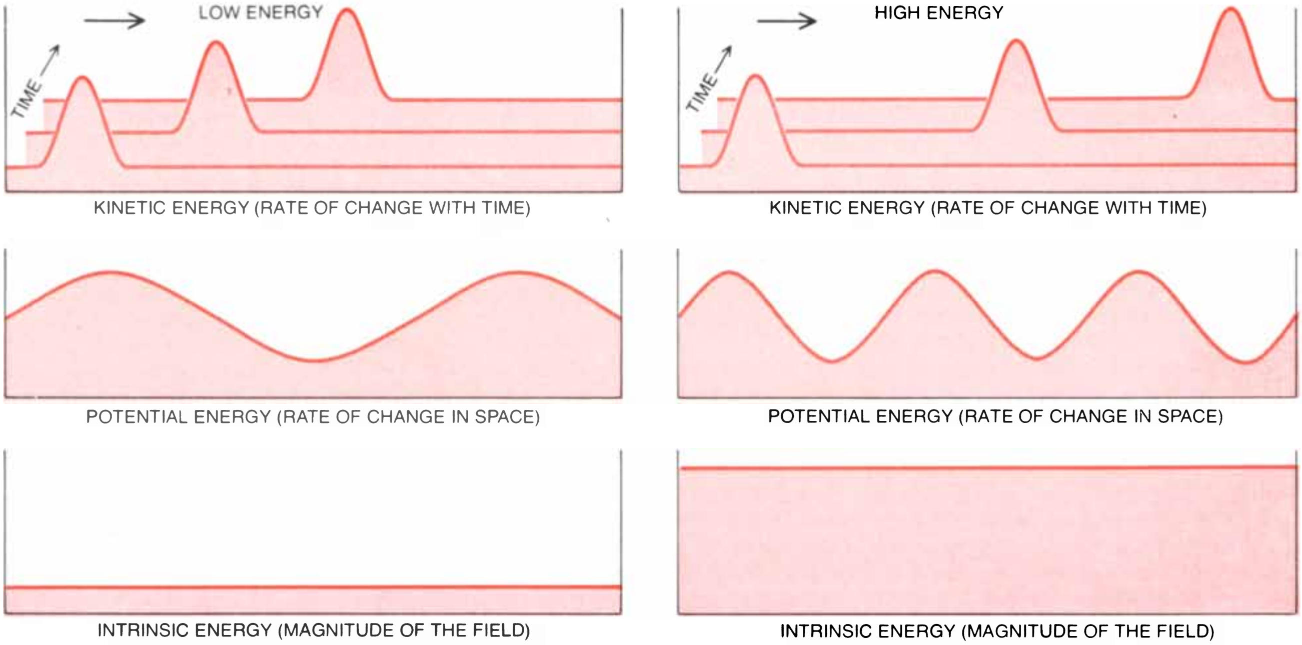
FIELD IN SPACE is the medium in which solitons propagate. A field assigns to each point in space some definite value of a specified quantity. Here the space (*a*) is one-dimensional, and the points shown are merely a few selected from an infinite continuum. In the simplest field (*b*) only a magnitude is measured at each point, represented here by the length of a line segment. A more complicated field, a vector field (*c*), has both a magnitude and a direction, shown here by arrows. For a field to be physically possible it must be defined everywhere (although its value can be zero), and if it changes from point to point, variation must be smooth and continuous.

An essential property of a field is that it can carry energy, just as a particle can. The energy of the field per unit volume is expressed mathematically as the sum of three quantities. One of these quantities is proportional to the square of the rate at which the field varies in time. The second term has a similar form but is proportional to the square of the rate at which the field varies in space. The third quantity is determined not by a rate of variation but by the actual magnitude of the field at each point. It is customary to call the first term kinetic energy and the sum of the other two terms potential energy. In this discussion, however, it will be convenient to have a name for each of the three quantities, and so I shall give the name intrinsic energy to the component that depends on the actual magnitude of the field, leaving the name potential energy for the component that is proportional to the square of the spatial rate of variation.

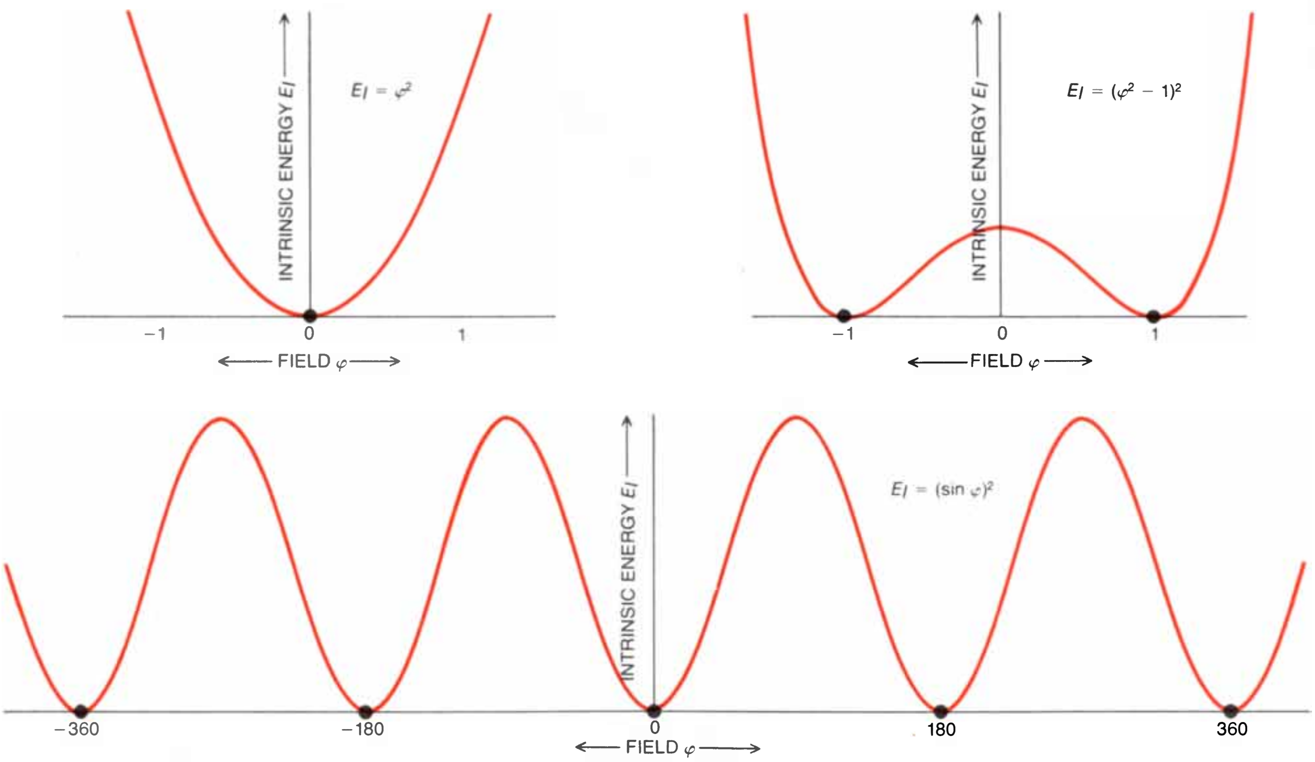
An intuitive rationale can be provided for this analysis of the energy of a field. The kinetic-energy term states that the total energy of the field rises when the value of the field (at each point) changes more quickly. In the water surface considered as a model of a field this relation implies that the kinetic energy is proportional to the square of a rate of change in position, or in other words to the square of a velocity. It is a matter of everyday experience that when something moves faster, it carries more energy.

The potential energy, as I have defined it, rises when a change in the state of the field is compressed into a smaller space. Intuition again confirms that the energy needed to bend or deform an object rises when the bend is made sharper. The interpretation of the intrinsic component of the energy is even more straightforward: it should hardly seem surprising that the energy of a field depends on the magnitude of the field. In the pond model the intrinsic energy is related to the overall surface level of the water. A higher waterline corresponds to a higher energy.

Because the kinetic energy and the potential energy of a field are both squared quantities they can never take a negative value. In evaluating the intrinsic energy of a field only energy differences need be considered, such as the difference between two water levels in a pond. The smallest observed intrinsic energy can therefore be set equal to zero, and so the intrinsic energy can also be defined in such a way that negative



ENERGY OF A FIELD is the sum of three components, all of them illustrated here on a water surface, which can be regarded as a model of a field. The kinetic energy is proportional to the square of the rate at which the field varies in time; a field with faster-moving waves embodies more energy. The potential energy is proportional to the square of the spatial rate of variation; a field whose fluctuations have a shorter wavelength has a higher energy. The intrinsic energy is determined by the magnitude of the field, although the rule relating these two quantities can take many forms. In the example of the water surface a higher water level corresponds to higher intrinsic energy. The energy of the field is zero when the field is static and uniform and in a configuration that yields the minimum intrinsic energy.



VACUUM STATES of a field are the zero-energy configurations. Because the kinetic and the potential energy can always be reduced to zero by making the field static and uniform, only the intrinsic energy need be considered in determining the vacuum states. Here the intrinsic energy is plotted as a function of the magnitude of the field for fields described by three simple equations. If the energy increases as the square of the magnitude of the field (*a*), there is only one vacuum state: the energy is zero when the field itself is zero. A variant of this equation (*b*) yields a field with two vacuum states. The energy is zero when the field has values of +1 or −1 but, significantly, the energy is greater than zero when the field itself vanishes. The energy of the sine-Gordon field (*c*) is a periodic function and has an infinite number of vacuum states. If the magnitude of the sine-Gordon field is interpreted as an angle in degrees, then the energy is zero whenever the magnitude is equal to zero, 180, 360 and so on. Topological solitons can exist only in fields that have multiple vacuum states.

values never appear. Thus the minimum total energy of a field is zero.

The state of a field that has minimum or zero energy is called the vacuum state. It is clear that in the vacuum state the field must be constant throughout both space and time, since any variation would give the kinetic or the potential energy a value greater than zero. To achieve the vacuum state it is thus necessary only to minimize the intrinsic energy. The vacuum is the state of a uniform field that has zero intrinsic energy.

Since the intrinsic energy is determined by the magnitude of the field, the most obvious configuration is one where the intrinsic energy is at its minimum when the field itself is zero everywhere. The vacuum is then the state with no field at all, which corresponds to the intuitive notion of a vacuum as being empty space. Surprisingly, however, that is not the only possibility. There are a number of field equations for which the intrinsic energy vanishes at some nonzero value of the field. Hence the state of the system with no field, which might at first seem to represent the vacuum, can actually have a higher energy than some alternative state with a nonzero field. That alternative state is the true vacuum, the point of zero energy, even though it describes a space permeated by a uniform field. An example of a field of this kind is the magnetic field of a ferromagnet. The state of minimum energy for a ferromagnet is not the demagnetized condition; on the contrary, a ferromagnet spontaneously generates and maintains a magnetic field because it can reduce its energy by doing so.

If the intrinsic energy can fall to zero for some value of the field other than zero, then it might be zero at more than one value. Indeed, certain field equations give rise to a multiplicity of vacuum states, each state associated with a different value of the field. All the vacuum states are equivalent (they have the same zero energy), but they are distinct. The existence of at least two vacuum states is a necessary condition for the creation of the solitons I shall describe here.

Suppose a wave, an excitation of some field, is observed at a given moment to occupy a finite volume of space. Outside this region there are no other excitations and the value of the field is such that the energy is zero everywhere. The overall state of the system cannot be the vacuum because there is energy in the wave. As the wave disperses, however, its energy will be distributed through ever larger volumes, and when it spreads to infinity, the energy per unit volume will have fallen to zero. The wave has then disappeared and the field has a constant vacuum value throughout space.

If the field is one that can have multiple vacuum states, there is another possible configuration of the system. In the regions surrounding an isolated wave there may be different values of the field, all of them corresponding to the vacuum but nonetheless distinct. If the topological arrangement of the vacuum states is such that the field cannot be extended to a consistent vacuum value everywhere in space, then the wave will be unable to expand and disperse. The result is a stable perturbation of the field: a soliton.

The topology of a soliton can be made clearer by considering a few examples of particular field equations. The simplest of these equations describes a field that exists in a one-dimensional space, a line of infinite length where each point can be specified by a single coordinate, x. A field in this space assigns to each point on the line some value, which I shall designate by the Greek letter φ (phi). A more formal description of the field is given by the mathematical expression $\varphi(x)$, which is read "φ is a function of x," or more briefly "φ of x," and which means that for every point on the line there is a unique value of the field φ.

In the same way that φ depends on x the intrinsic energy depends on φ. This dependence is recorded explicitly by the notation $E_I(\varphi)$: the intrinsic energy is a function of φ, and thus each possible value of the field has some definite intrinsic energy. The physical meaning of these mathematical formulas is easily deciphered. Given any point, x, in the one-dimensional space, the value of the field is specified at that point without ambiguity. Given the value of the field at each point, one can then calculate its intrinsic energy. Solitons appear when several values of the field have the same minimum intrinsic energy.

One expression for the intrinsic energy that can give rise to solitons is the equation $E_I = (\sin\ \varphi)^2$. The equation states that for any value, φ, of the field, the intrinsic energy can be found by taking the sine of φ and then squaring the result. The equation of motion for waves propagating in a field that has this property is called the sine-Gordon equation. It is a modification of another equation first discussed in 1926 by Oskar Klein and Walter Gordon.

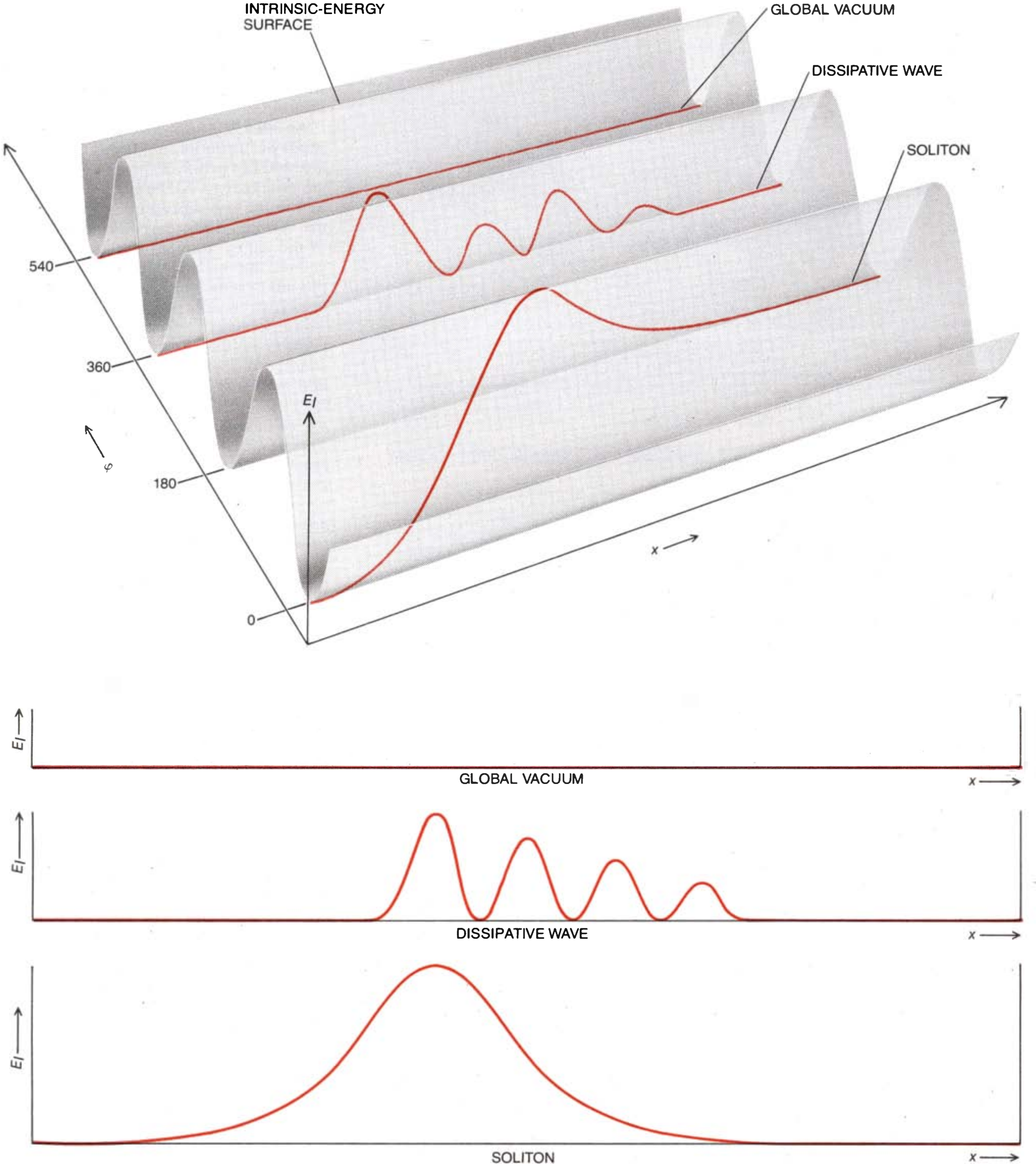
The properties of the sine-Gordon field can be explored by substituting numerical values for φ in the equation that defines the intrinsic energy. The values need not be assigned any dimensions. For the sake of simplicity I shall assume here that the units are chosen so that the intrinsic energy varies between zero and 1 and that φ is measured in degrees.

The sine is a trigonometric function whose value ranges from zero to $+1$ and from zero to -1, and so its square must be confined to the range of values between zero and $+1$. If φ is equal to zero, then the sine of φ and the square of the sine are also zero: hence the sine-Gordon equation is one in which the state with no field is a vacuum state. As φ increases from zero to 90 the sine and the square of the sine rise smoothly, reaching a value of 1 when φ is equal to 90. As φ continues to increase, however, the square of the sine decreases and returns to zero when φ is equal to 180. That value of φ therefore specifies another vacuum state of the system. With a continuing increase in φ the sine becomes negative, but the square of the sine, of course, remains positive and again has a value of 1 when φ is equal to 270. When φ is equal to 360, the square of the sine is zero again, representing another vacuum state. Additional states of zero intrinsic energy are found when φ is set equal to 540, 720, 900, 1,080 and so on. As φ increases there are infinitely many possible vacuum states, each separated from the neighboring vacuum by a hump where the intrinsic energy rises smoothly to a value of 1.

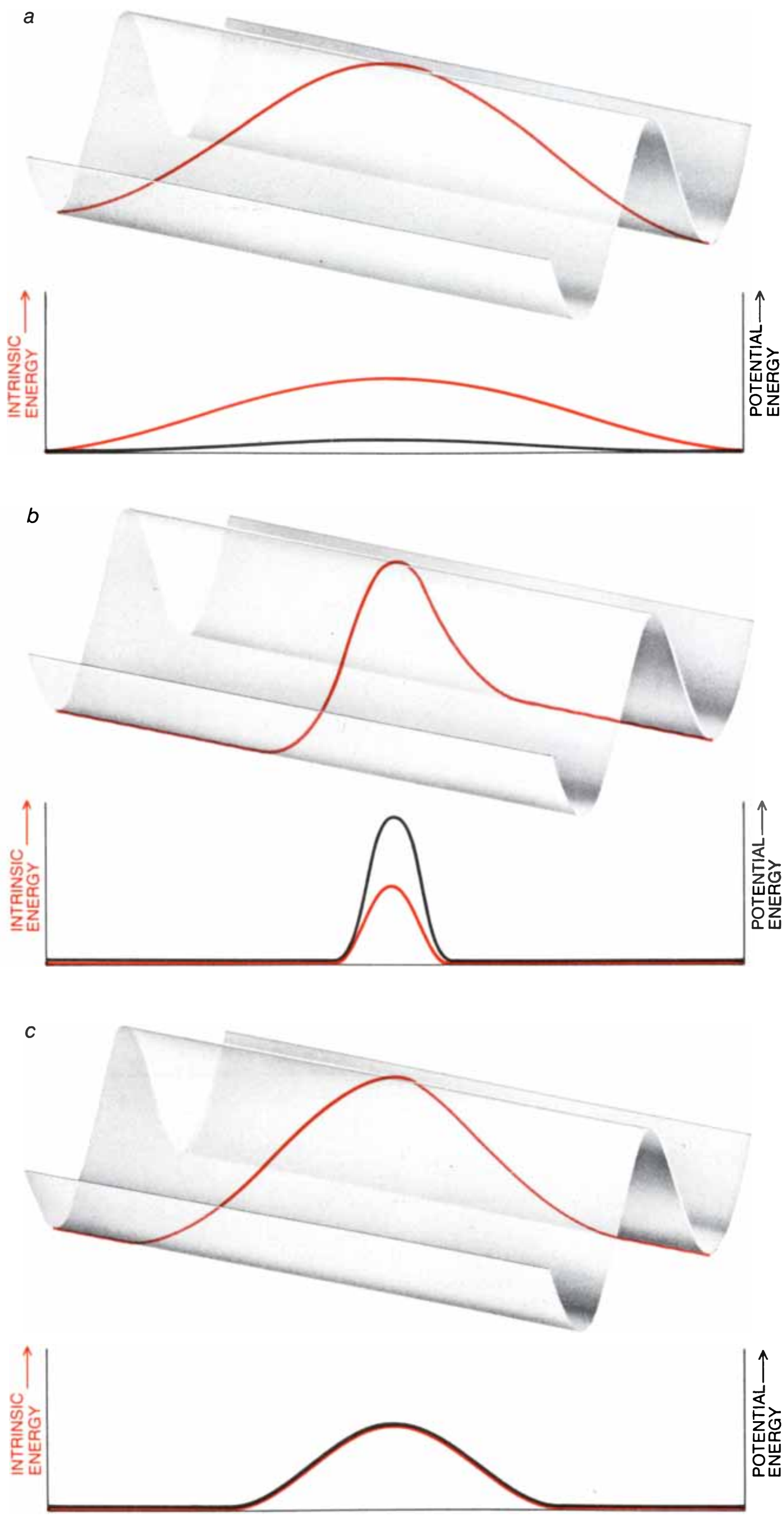
The configuration of a one-dimensional field can be represented by a two-dimensional graph, that is, by a graph on a plane. One axis is labeled with all the points, x, in the one-dimensional space; the other axis gives all possible values of the field, φ. For each point x a mark is made on the plane at a position corresponding to the value of the field at that point. Because the field must be defined at every point and because the values must form a continuum, the marks can always be connected to make a line or curve. If the line is straight and parallel to the x axis, it represents a state in which the field is constant at all points.

In order to display the intrinsic energy of the field a third dimension is added to the graph: intrinsic energy is made proportional to height above the plane. For the sine-Gordon field the points representing the intrinsic energy of every possible value of the field form a sinusoidally undulating surface. In this energy surface the undulations are parallel to the x axis and have troughs of zero energy where φ is equal to zero, 180, 360 and so on [*see illustration on opposite page*].

In this three-dimensional graph the configuration of the field is represented by a line laid on the undulating energy surface. One possible configuration is represented by a straight line that remains at the bottom of one of the troughs, say the trough at $\varphi = 180$. The meaning of the graph is that the field has a magnitude of 180 at every point on the line x. It does not vary in space and unless it is perturbed by some outside influence it cannot vary in time. Moreover, because $\varphi = 180$ is one of the values of



ENERGY SURFACE of the sine-Gordon field illustrates the topological constraint that confines a soliton. Distance along the x axis, which extends to infinity in both directions, gives the position of a point in a one-dimensional space. Distance along the axis labeled φ gives the magnitude of the field at the point x. The height of the energy surface above the plane (along the axis labeled E_I) gives the intrinsic energy of the field at this point. The sine-Gordon field has multiple vacuum states, namely the troughs in the energy surface. A line that lies in one of these troughs describes a global vacuum state: the field has the same magnitude everywhere, and it is a magnitude that corresponds to zero intrinsic energy. A line that oscillates within a valley represents a local excitation of the field, or an ordinary wave. Because the line wanders above the floor of the valley this field carries energy, but the wave will eventually disperse. A soliton forms when the field assumes different vacuum values along the two directions that lead to infinity. For the soliton shown here the field to the left is in the vacuum state at $\varphi = 0$, but to the right it has the alternative vacuum value of $\varphi = 180$. The field is required to be continuous, and so at some point along the x axis it must cross the hump in the energy surface that separates the two vacuum states. The kink at the transition zone is the soliton. It can be pushed from side to side, but if the surface is infinite, the kink can never be removed and the soliton can never disperse. At the bottom the intrinsic energy of the three field configurations is projected onto a two-dimensional graph. The dissipative wave and soliton also have potential energy, and if they are moving, they have kinetic energy, but these quantities are not shown.



OPTIMUM SHAPE for a soliton is the shape that yields the smallest sum of potential energy and intrinsic energy. The potential energy, being determined by the spatial rate of variation, is smallest when the transition is gradual and the line crosses the ridge at a shallow angle (*a*). A long segment of the line, however, is then near the peak in intrinsic energy. The intrinsic energy is minimized when the line crosses the ridge almost perpendicularly (*b*), but such an abrupt transition makes the potential energy large. The best compromise is achieved when the line crosses the ridge at an intermediate angle (*c*) and the potential and intrinsic energy are equal.

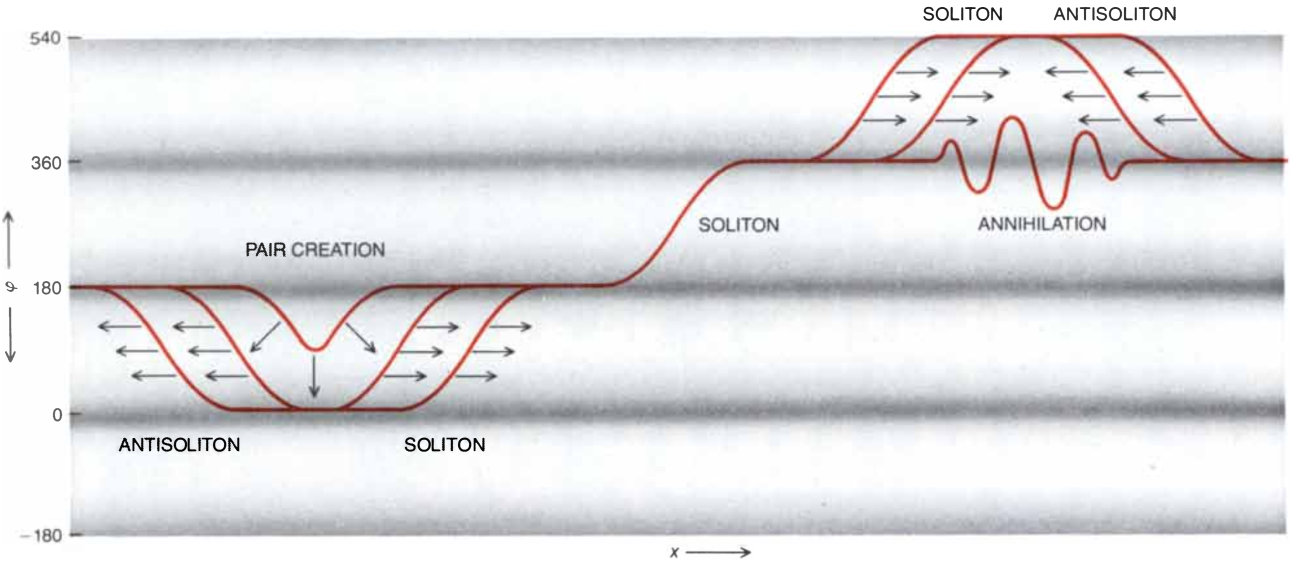
the field for which the intrinsic energy is zero, the total energy of the field is zero. The field is in one of its many equivalent vacuum states; because it is the same state everywhere it is called a global vacuum.

Another configuration results when a small perturbation is introduced into a line that lies in one of the troughs of zero intrinsic energy. At the site of the perturbation the line representing the state of the system climbs partway up one wall of the valley, plunges back to the floor and up the other wall, then continues this sinuous pattern in a series of oscillations that become progressively smaller. The graph corresponds to a wavelike excitation of the field, but one that will eventually disperse. Where the line wanders above the floor of the valley the field has intrinsic energy, and because it varies in space it has potential energy. The process of dispersion itself adds kinetic energy. As the wave spreads out over the infinite length of the line, however, the energy per unit length approaches zero and the field returns to a stable state of global vacuum.

A soliton of the sine-Gordon field appears when the line representing the state of the system climbs from the bottom of one valley in the energy surface, crosses a ridge and descends to the bottom of an adjacent valley. Extending in either direction from this point of transition the field is in a stable vacuum state but not the same vacuum state. For example, the magnitude of the field at all points to the left of the transition might be 180 and at all points to the right 360. In the transition region the field does not have a vacuum configuration. It has intrinsic energy because the line must surmount the hill in the energy surface, and it has potential energy because the value of the field changes from one point to the next.

The soliton represents a kink in the configuration of the field, and if the line x extends to infinity in both directions, the kink can never be removed. One end of the line is anchored in the vacuum state at $\varphi = 180$ and the other end at $\varphi = 360$. Since the line must be continuous, it must cross the energy maximum between these two vacuum states at some point along its length. The soliton can never disperse and its energy can never be dissipated.

The energy of a soliton in the sine-Gordon field depends on the geometry of the transition between vacuum states. The potential energy is minimized when the field varies as gradually as possible, and therefore when the line crosses the ridge at a shallow angle, nearly parallel to the x axis. In that configuration, however, the line is elevated above the bottom of the valley for a substantial distance and the intrinsic energy is high. The intrinsic energy assumes its minimum value when the line crosses



CREATION AND ANNIHILATION of solitons resemble the corresponding operations among quantum-mechanical particles. In the sine-Gordon field a section of the line that describes the state of the system can be lifted from the bottom of one valley and draped over a ridge into a neighboring valley. The result is the simultaneous creation of a soliton and an antisoliton, which can move away from the site. In the reverse process a soliton and an antisoliton meet and annihilate each other. Equations that govern the sine-Gordon field always create another pair immediately after the annihilation, but in other soliton systems energy of the waves can be dissipated in other ways.

the hill perpendicular to the x axis, making a very abrupt transition between vacuum states, but in that case the rate of spatial variation is high and the potential energy reaches an extreme value. The minimum total energy of the soliton is achieved when the crossing is smooth and at an intermediate angle, so that the intrinsic energy and the potential energy are equal.

Although the sine-Gordon field has a remarkably simple structure, the solitons that appear in it have several important properties in common with material particles. The stability of the soliton at rest has already been discussed. The soliton can also be set in motion without altering its form: the transition region between the two vacuum states rolls along the hill at a constant velocity. Even when the soliton is in motion, the width of the transition region remains constant, and the only change in the energy is the addition of a term for the kinetic energy.

There is nothing in the sine-Gordon equation that restricts the field to configurations with only one soliton, and in principle there could be an unlimited number of both solitons and antisolitons. By convention a soliton is a kink where the field increases (from one vacuum value to the next) as x increases and an antisoliton is a kink where the field decreases with increasing x. It is a simple matter to create a soliton-antisoliton pair from a state of global vacuum: simply pick the line up from the bottom of one valley and drape a loop over the ridge into an adjacent valley. The process corresponds to the creation of a particle and its antiparticle in a quantum field theory.

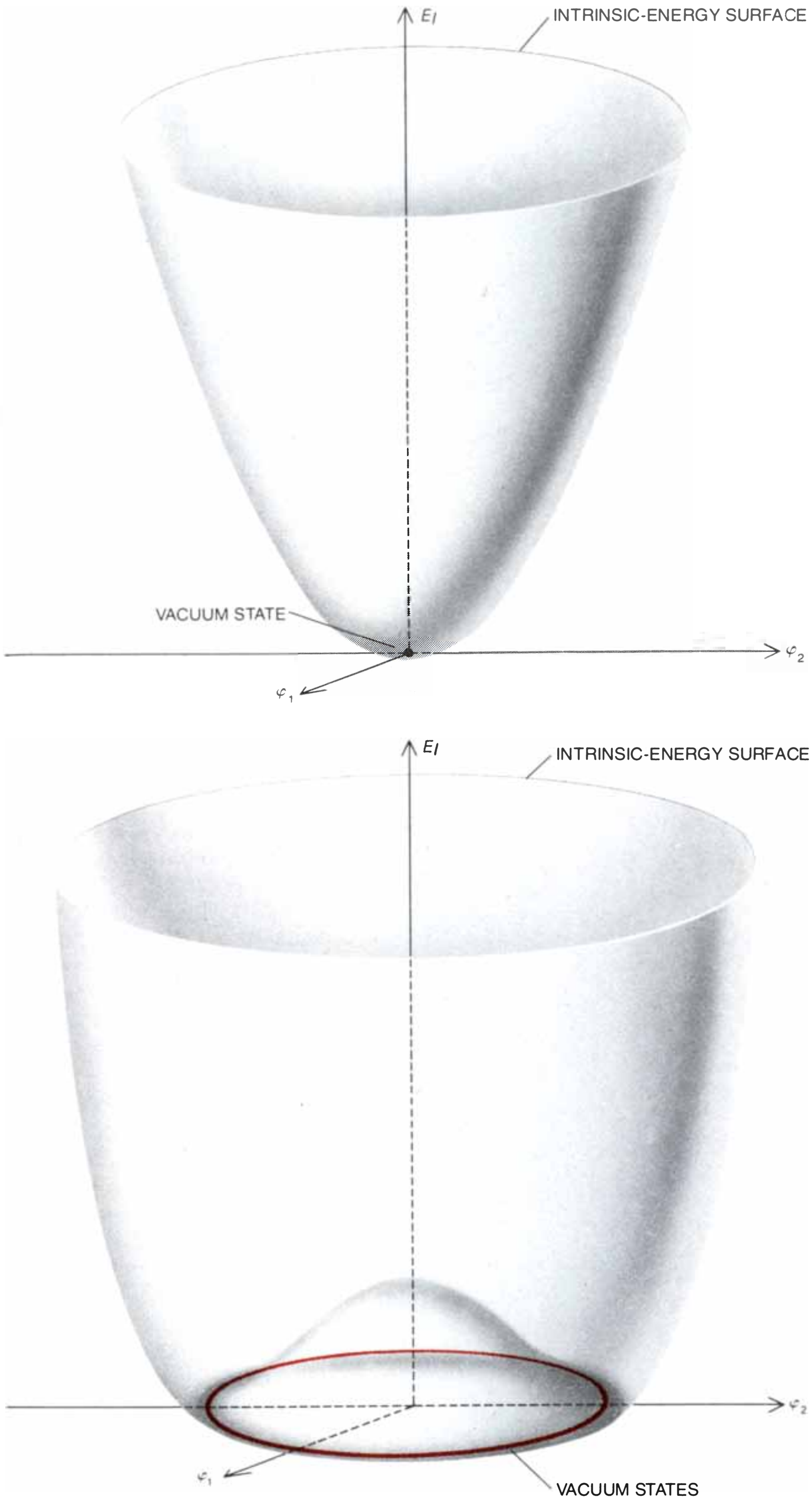
In the reverse process a soliton and an antisoliton collide. The sine-Gordon equation has the special property that both the soliton and the antisoliton emerge from the collision unchanged, but it is easy to modify the equation so that the waves are annihilated. Both kinks then disappear and their energy is dissipated in a wave that disperses into the global vacuum. Mutual annihilation is also observed among particles and antiparticles. Because any soliton-antisoliton pair can readily be created or annihilated, it is only the difference between the total numbers of these objects that is conserved. For example, a given field might go off to infinity in one direction at a value of $\varphi = 180$ and in the other direction at $\varphi = 360$, but there could be seven solitons and six antisolitons in the transition region between these two values. Six solitons and six antisolitons could be made to annihilate one another, but there is no way for the last soliton to be eliminated.

It is easy to make a working model of the energy surface of the sine-Gordon field. Fold a piece of heavy paper into several accordion pleats, or if possible into smooth undulations, then place a length of flexible key chain across the surface. If the chain begins in the bottom of one pleat and ends in another, a soliton must exist where the chain crosses from one vacuum state to the other. The absolute conservation of the soliton cannot be proved unless the surface and the chain are made infinitely long, but certain other properties can be investigated. By tilting the surface, for example, one can make the soliton move, and with careful manipulation it may even be possible to observe the creation and annihilation of soliton-antisoliton pairs. I first saw such a model demonstrated by Holger Nielsen, a Danish physicist who has done pioneering work in the study of solitons.

It may appear that the sine-Gordon field is an arbitrary construction that has no point of contact with the real world. As a one-dimensional field its descriptive power is necessarily limited. What is more, the multiple vacuum states that are essential for the existence of solitons are introduced by what may seem to be a rather artificial assumption, namely the assumption that the intrinsic energy is a periodic function of the field. There are physical phenomena in the three-dimensional world, however, that are effectively confined to one dimension. An example is the motion of electrons along a stack of molecules, and the sine-Gordon equation has been applied to the analysis of this system and similar ones. The periodic variation of the intrinsic energy is also not too remote from experience. Consider the energy of a pendulum expressed as a function of the angle that measures deviation from the vertical. As the angle increases from zero degrees the energy rises to a maximum when the angle measures half a turn, then returns to zero after a whole turn, and so on, passing through many equivalent maximums and minimums as the angle continues to increase.

The qualitative similarity between interacting solitons of the sine-Gordon field and certain interacting particles has been extended to a formal equivalence. In 1958 Walter E. Thirring of the University of Vienna formulated a quantum-mechanical model of particles and antiparticles moving in a one-dimensional space. Sidney R. Coleman of Harvard University has recently shown that the Thirring model and the solitons


AMBULANCE



TWO-DIMENSIONAL FIELDS can incorporate solitons if there are multiple vacuum states with an appropriate topology. The fields are defined by two quantities at each point in space, designated φ_1 and φ_2. Each combination of φ_1 and φ_2 implies a definite intrinsic energy, which is given by the energy surface for the field. In one possible field theory (*upper diagram*) the energy is zero only when both components of the field are zero; no solitons are possible in this field. Solitons can appear when the energy surface has a somewhat more complicated structure (*lower diagram*). Here a field whose components are both zero has a definite, nonzero energy. The vacuum states of the field are the combinations of φ_1 and φ_2 that describe the circumference of a circle with unit radius. If φ_1 and φ_2 are the components of a vector, then the field has zero energy when the magnitude of the vector is 1, regardless of its direction.

of the sine-Gordon equation describe the same phenomena. In the Thirring model the particles are assumed to exist and are then made to obey a postulated scheme of interactions; the solitons, on the other hand, are not introduced a priori but arise naturally from the equations of motion. Nevertheless, the two kinds of objects propagate and interact in the same way.

The notable successes of the sine-Gordon theory aside, there is no question that the real world has three spatial dimensions rather than one dimension. If solitons exist as real particles, they must be found in a three-dimensional theory. Two-dimensional waves with the properties of solitons have been known for a few years; the discovery of three-dimensional solitons is more recent. The latter may exist in nature as particles; the two-dimensional solitons are of interest in other branches of physics and are useful models for illustrating the properties of the three-dimensional ones. Indeed, once the step from the line to the plane is made, no further conceptual barriers are encountered in the passage to spaces of higher dimensionality.

The stability of many solitons in higher-dimensional spaces can be proved by topological arguments similar to those employed in the sine-Gordon field, although the geometric configuration of the fields becomes somewhat more difficult to visualize. In two dimensions the value of a field is a continuous function of position on a plane. A graph of the field can be constructed by drawing the plane and letting height above each point represent the value of the field at that point. The set of all values forms a surface erected over the plane; it is a flat surface parallel to the plane if the field is constant or a more complicated surface for a field that changes from point to point. In three dimensions the field must be defined at every point in ordinary space; it is not possible, however, to draw a graph of a three-dimensional field except in a hypothetical four-dimensional space. On the other hand, the properties of the two- and three-dimensional fields are very similar, so that for most purposes the simpler field can serve to illustrate both.

In one dimension the property that was found to be essential for the existence of solitons was the presence of multiple vacuum states, where the intrinsic energy vanishes at various values of the field. A soliton appears when the field approaches different vacuum values as one proceeds toward infinity in opposite directions on the line. On a plane, infinity can be reached by proceeding in any direction from a chosen origin; the number of directions that lead to infinity is itself infinite. The directions can be put into one-to-one correspondence with the infinity of points

PENTAX SYSTEM 10.
THE FIRST 110 SLR AS PRECISE AS A FINE WATCH AND ALMOST AS SMALL.

It might make you feel good to know that if you sometimes think you are hard to satisfy, we believe *we* are almost impossible to satisfy. Ever.

This attitude was the reason we spent years and millions of dollars seeking a way to drastically improve the 110 camera.

And we have succeeded. Almost beyond our own expectations. And certainly far beyond the expectations of the photographic press.

The new Pentax System 10 is the first 110 SLR (single lens reflex) with interchangeable lenses, a power winder, a dedicated auto flash and a host of accessory close-up lenses, filters, lens shades and carrying case.

We've overcome the limitation of the fuzzy 110 picture.

In fact, this unretouched enlargement was actually taken with the Pentax System 10 camera, using its telephoto lens.

The image is razor sharp and free of perceptible grain.

What also makes System 10 unique and so astonishing is its incredible lack of size and weight, coupled with the most sophisticated SLR technology.

It measures $2\frac{1}{5} \times 3\frac{9}{10} \times 1\frac{4}{5}''$ and weighs less than 6 oz.

System 10 is designed to be the ideal complement to your 35mm SLR. The whole system can fit in a small area of your current camera case, in place of a zoom lens (but your pocket will do in a pinch).

And frankly, there are times when you just don't want to carry a 35mm SLR around. But up to now, everything else has been a compromise.

Now, with the Pentax System 10, you'll never again have to say "I wish I had my camera with me."

Pentax System 10. Available in limited quantities only at select fine photographic stores.

TECHNICAL DATA

Type TTL metering 110 SLR camera, bayonet mount (80-degree setting angle).
Film Cartridge 110 film format; ASA set automatically for film in use.
Lenses Pentax-110 24mm f/2.8; Focusing range: 0.35m to ∞. Pentax-110 50mm f/2.8 telephoto; focusing range: .9m to ∞. Pentax-110 18mm f/2.8 wide angle; range: .25m to ∞.
Viewfinder Eye-level pentaprism finder with instant-return mirror; green or yellow LEDs indicates correct handheld exposure, or need for flash/tripod.
Exposure Metering Center-weighted, TTL metering at full aperture via SPD cell; EV Range: 3-17 (24mm F 2.8 lens; ASA 100).
Shutter/Automatic Diaphragm Programmed electronic shutter with programmed body diaphragm; Range: F 2.8 to f/13.5.
Other Features Two-stroke film advance lever, synch terminal for automatic flash synch with AF 130P auto flash; tripod socket, lens bayonet release button, battery holder tray; exposure count via back cover window, winder capability.
Power Source: Two 1.5V silver-oxide batteries.
Weight 172 grams (6.1 ozs) w/lens. Size (2.2" x 3.9" x 1.8")

PENTAX®
SYSTEM 10

Let it snow, let it snow, let it snow...

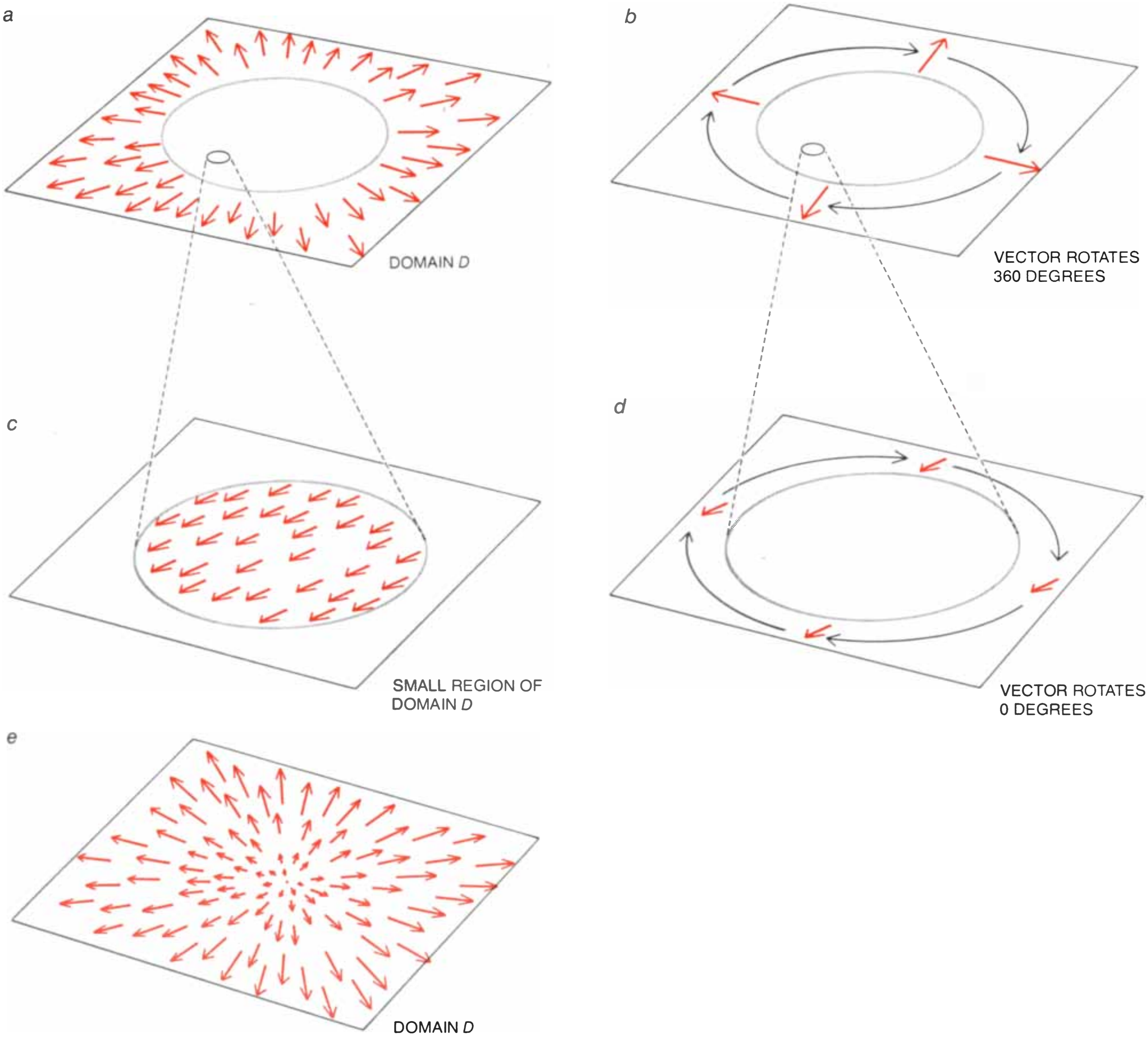
Button up your overcoat and join us making fun of winter...at the carnivals, follies, sleigh rides, ice slides and polar bear dips – all across Ontario. Take to the woods, on hundreds of miles of pristine cross-country trails, where the world is white, the silence golden. Or head for the hills and test your stuff on some of the longest and best maintained runs east of the rockies. Après ski, come in from the cold...our first-class resorts offer saunas, pools, firelit lounges, gluhwein and good times, and we treat you royally. This year you'll enjoy a substantial premium on every dollar – and the removal of Ontario's 7% room tax is an added plus. For information call us COLLECT (416) 965-4008, or write: Ontario Travel, Dept. L.S., Queen's Park, Toronto, M7A 2E5.

Ontario Canada

on the circumference of a circle. The situation is the same in three-dimensional space, where the directions toward infinity correspond to the points on the surface of a sphere. For a soliton to exist in either of these spaces the field must approach distinct vacuum states along each of the possible directions.

The topological peculiarity that gives rise to solitons is not possible for all two- and three-dimensional fields, or even for all those fields that have multiple vacuum states. One requirement is that the field itself be defined by more than one quantity at each point. The fields I shall discuss are specified by two values at each point in two-dimensional space and by three values in three-dimensional space. They are called vector fields because the two or three quantities can be regarded as giving the components of a vector along two or three orthogonal axes. The vector field is designated by writing the symbol for the field with an arrow over it: $\vec{\varphi}$. The intrinsic energy of a two-component vector field can be represented as a surface drawn over a plane. The plane is not that of the two-dimensional space itself but a plane giving all possible values of the vector field, or all possible combinations of magnitude and direction for the vector $\vec{\varphi}$.

One plausible equation for calculat-



CONFINEMENT OF A SOLITON in a two-dimensional space is demonstrated by a topological argument. The space (a plane) is filled with a vector field, a field defined at every point by a magnitude and a direction. Outside some arbitrary domain, *D*, the field is observed to radiate from *D* and its magnitude is everywhere equal to 1 (*a*). A field of that magnitude (but only one of that magnitude) is assumed to have zero intrinsic energy, and so everywhere outside *D* the field is in a vacuum state. At issue is whether the vacuum can continue inside *D*. That it cannot is proved by the following procedure. Imagine touring the perimeter of *D* while carrying an arrow that always remains parallel to the field (*b*). After a complete circuit the arrow resumes its original orientation, but it has rotated one turn. Now examine some microscopic region inside *D*; if the region selected is small enough, the field inside it must be uniform, with all the vectors parallel (*c*), since large variations in the field cannot be compressed into an arbitrarily small space. During a tour of this region (*d*) the arrow does not rotate at all. Many other circular tours can be imagined at scales intermediate between these two; for the larger tours the arrow would still rotate by a full turn and for the smaller ones it would not rotate. At some scale, however, there must be a transition, where the rotation of the vector field changes from 360 to zero degrees. A theorem in topology states that such a transition is possible only if at some point inside *D* the field itself vanishes, so that the direction of the arrow is undefined. It follows that somewhere inside *D* the field must be zero, and a smooth decline to zero at the center is the most plausible configuration (*e*). A field with a magnitude of zero, however, is not a vacuum state; it has energy greater than zero. The energy is that of a soliton, confined by a "twist" in the structure of the field.

ing the intrinsic energy as a function of the vector field is $E_I = (\vec{\varphi})^2$. The intrinsic-energy surface described by this equation is a paraboloid, the surface generated when a parabola is rotated around its axis. The intrinsic energy has just one minimum, at the apex of the paraboloid, which corresponds to the origin of the intrinsic-space plane. Therefore the intrinsic energy is zero only when both the magnitude and the direction of the field are zero. Because the field has only this one vacuum state, corresponding to no field at all, topological solitons are not possible.

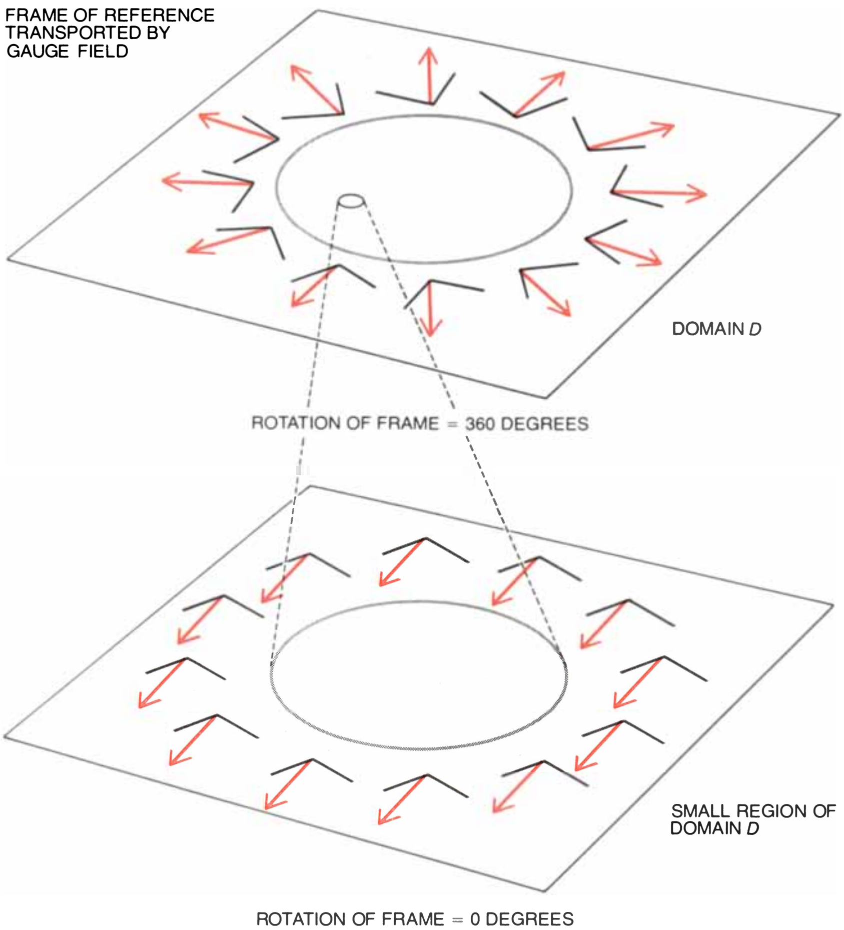
Suppose the intrinsic energy is given by a somewhat different equation, $E_I = (\vec{\varphi}^2 - 1)^2$. In this case the intrinsic energy is not zero at the origin of the plane; when $\vec{\varphi}$ is equal to zero, the intrinsic energy is equal to $(-1)^2$, or simply 1. Hence the state with no field is not a vacuum state. The true vacuum state is found when the magnitude of $\vec{\varphi}$ is equal to 1, since $(\vec{\varphi}^2 - 1)^2$ is then equal to zero. The equation $(\vec{\varphi}^2 - 1)^2 = 0$ is the equation of a circle; it is satisfied by all the points on the circumference of a circle whose center is the origin and whose radius is 1. The field is therefore in a vacuum state whenever the magnitude of the vector is 1, regardless of its direction. The complete energy surface has a curious structure. At large distances from the origin it looks much like a paraboloid, although it is steeper. Near the origin the surface reaches a minimum at all points on the circle where the magnitude of $\vec{\varphi}$ is equal to 1, then rises again inside this circle. The surface looks like a quartic paraboloid with a dent poked into its apex [*see illustration on page 102*].

What does this energy surface imply about the possible configurations of a two-dimensional vector field? One consequence of particular importance to this discussion is that the field can have zero intrinsic energy if the magnitude of the field vector is everywhere equal to 1, regardless of the vector's direction. Consider a field whose magnitude and orientation have been measured at all points on a plane outside some arbitrary domain *D*. The magnitude of the vector has been found to be equal to 1 everywhere, and its direction is such that the vector always points away from the origin, which is inside the domain *D*. Hence outside *D* the plane is filled with vectors that all have unit length and that seem to radiate from *D*. Everywhere outside *D* the field is in a vacuum state, since the unit vector corresponds to zero intrinsic energy. The question is: Can the vacuum state also extend inside *D*, so that the field has zero intrinsic energy everywhere? It cannot. The proof that it cannot is topological in nature and it shows that *D* must envelop a region of finite energy: a two-dimensional soliton.

Imagine taking a walk around the domain *D*, carrying an arrow that always points in a direction parallel to that of the field. On returning to the starting place the arrow will be pointing in the same direction it was when the walk began, but it will have described a full circle, a rotation through 360 degrees. Now imagine examining some microscopic region inside *D*. It might seem most plausible for the pattern of radiating vectors to be reproduced at smaller scale in this region, but a simple argument shows that such a field configuration is impossible. If the field were to make a 360-degree rotation within a microscopic region, the rate of spatial variation and hence the potential energy would be enormous. As the area of the sample approached zero the potential energy would be unbounded. It follows that if a small enough sample area can be chosen, the field within it must be uniform: the vectors at all points must have not only the same magnitude but also the same direction. Thus a microscopic sample of *D* would reveal a field in which all the vectors are parallel. An arrow carried around this sample would never deviate from its original orientation; it would not rotate at all.

The procedure of carrying an arrow around a closed loop in a field illustrates a theorem in topology. The theorem states that if the field has no discontinuities and does not vanish at any point, then the arrow must rotate an integer number of times during any circuit. The arrow can fail to rotate at all, in which case the integer is zero, or it can turn once, twice and so on, but it cannot make half a turn. No proof of this theorem will be presented here, but it is supported by common sense, which indicates that the arrow must return to its original position at the end of any complete loop.

The theorem is satisfied by the loop



GAUGE FIELDS alter the description of a soliton in two or three dimensions. For a soliton to exist, a field radiating from a domain, *D*, must be in a vacuum state outside *D*. Setting the magnitude of the field equal to 1 makes the intrinsic energy zero, but the potential energy must also be taken into account. Because a radiating pattern of vectors changes direction from point to point, the potential energy does not seem to be zero. The quandary is resolved by introducing a new field, the gauge field, which brings with it a prescription for transporting a frame of reference from point to point. Outside *D* the frame rotates by exactly the amount needed to keep the direction of the field vector constant at all points (when the direction is measured with respect to the transported frame). In a small region inside *D* neither the vector nor the frame rotates. With the introduction of the gauge field the potential energy of the original field is eliminated but the topological confinement of the soliton is retained. It is no longer the original field vector but the gauge field whose net rotation changes from 360 to zero degrees.

outside *D*, where the vector rotates through 360 degrees, and by the microscopic loop inside, where the rotation is zero. What happens, however, during some loop at an intermediate scale? At some point, as the size of the region examined is reduced, the rotation of the field must make a transition from 360 to zero degrees. The topological theorem forbids a loop with a fractional rotation, but it does include two escape clauses. One is the possibility that the field has a discontinuity, where its value changes abruptly from one point to the next. That solution is excluded, however, by the requirement that all physical fields exhibit only smooth and continuous changes. The only remaining possibility is that somewhere inside *D* the value of the field falls to zero. At that one point the net rotation can change from 360 to zero degrees because where the field vanishes the direction of the vector is undefined.

This elaborate argument has been presented for one purpose: to prove that inside the domain *D* there must be a point where the magnitude of the field is not equal to 1; instead it must be equal to zero. It then follows that at this point the intrinsic energy is not zero. In fact, the requirement that a field vary only by smooth increments creates an entire region inside *D* where the energy is greater than zero. This lump of energy cannot spread out over the plane; it is confined by the "twist" where the net rotation of the field changes by one turn.

A three-dimensional soliton is similar in structure, and its confinement can be explained by analogous arguments. The points of minimum intrinsic energy now make up the surface of a sphere, and the intrinsic energy rises both inside this sphere and outside it. A field configuration corresponding to these vacuum states can again be postulated for all points outside some domain, but now the domain encloses a spherical volume rather than a circular area. The same topological argument shows that the vacuum state cannot be extended to all regions inside the domain, and so there must be a region where the field has a value that gives an intrinsic energy greater than zero. This ball of energy is a three-dimensional soliton, and it closely resembles an elementary particle in ordinary space.

In the foregoing analysis of two- and three-dimensional solitons an important element has been neglected. The configuration of the field outside the domain *D* has been defined as the vacuum state because it has zero intrinsic energy, but no account has been taken of its potential energy. Because the field vector changes direction at every point, the potential energy is not zero. Indeed, it can be proved that the potential energy is infinite, so that the field configuration not only fails to describe a localized wave but also seems to be unattainable.

There is a way out of this impasse, but it is not a simple one. The infinite potential energy outside the domain *D* can be reduced to zero by postulating a new kind of field, called a gauge field. In the past several years gauge fields have become much-studied structures, not only in physics but also in mathematics, where they are known by a different name: connections. I shall not explore the workings of a gauge field in detail but shall concentrate on those properties that are important to the theory of solitons.

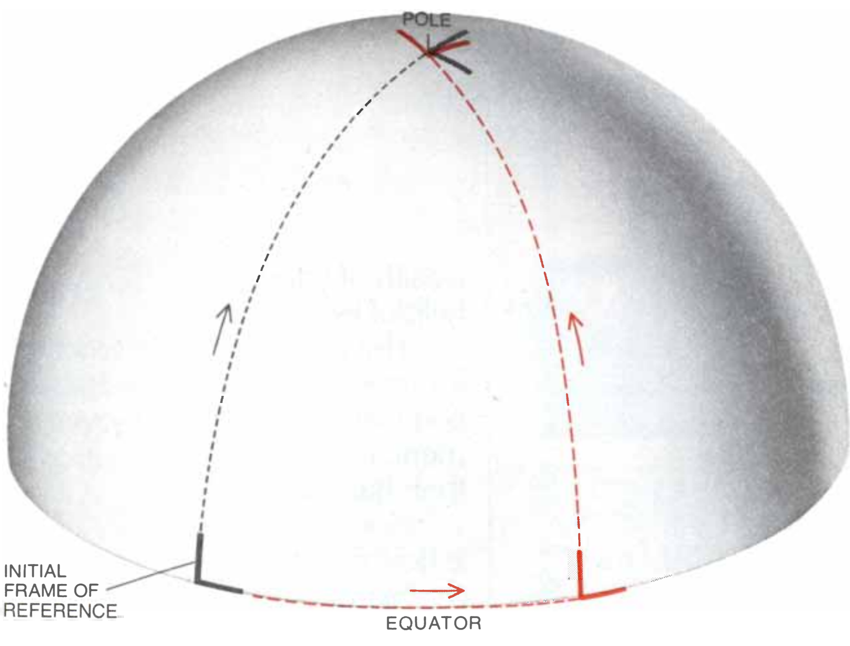
In order to eliminate the potential energy of the field without also destroying the soliton, the change in orientation of the vector from point to point must be abolished, while the overall rotation of the field by 360 degrees must be preserved. Although these two goals may seem to be irreconcilable, they are both accomplished by a gauge field.

In describing the two-dimensional field it was assumed implicitly that the direction of the vector would be measured at each point with reference to some fixed set of coordinate axes in the abstract plane where all possible values of $\vec{\varphi}$ are plotted. The frame of reference could be arbitrary, but it remained the same throughout space: it was a global frame. The introduction of a gauge field allows the frame of reference to rotate as one moves between neighboring points. The potential energy is then evaluated by measuring the variation of $\vec{\varphi}$ not with respect to a global frame of reference but rather with respect to local frames that can change from point to point.

Suppose that somewhere on the perimeter of *D* the field vector makes an angle of 30 degrees with a chosen frame of reference. If the frame is a global one, then after traveling 90 degrees around a circle the vector will have assumed a new orientation of 120 degrees. A suitably arranged gauge field, however, allows the frame of reference to rotate with the vector; no matter where a measurement is made, the orientation of the vector field is effectively constant and so the potential energy vanishes. On the other hand, the notion of a twist in the field configuration is preserved: it has been transferred from the field $\vec{\varphi}$ to the gauge field. Now it is the local frame of reference that undergoes a complete turn in a closed path around *D*.

It is important to emphasize that the gauge field does not simply specify the orientation of many independent frames of reference. Instead it prescribes how the orientation of a single frame changes when the frame is displaced. If it were necessary only to measure the changing orientations of a frame in a two-dimensional space, then it would suffice to specify one number at each point: the angle of rotation. Actually a two-dimensional gauge field is defined by two numbers at each point. One number tells how much the frame rotates during a displacement along the *x* axis and the other tells how much it rotates during a displacement along the *y* axis.

Because the gauge field is a prescription for transporting a frame between



TRANSPORT of a frame of reference according to the prescription of a gauge field is analogous to transport over the surface of the earth. If the frame is transported from its initial position on the Equator directly to the pole, it assumes one final orientation (*black*), but moving it first along the Equator and then to the pole yields a different final orientation (*color*). Hence the rotation during a displacement depends on the path followed. The rotation results from the curvature of the earth. In abstract space a gauge field has an effect equivalent to curvature.

neighboring points rather than a prescription for setting up a different frame at each point, the net rotation of the frame during a displacement depends on the path taken. In going from Rome to London by way of Berlin the frame might rotate by some definite angle. If the frame is carried from Rome to London by way of Paris, the angle of rotation will in general be quite different. Moreover, on a round trip—say from Rome to Berlin to London to Paris and back to Rome—the frame will generally not return to its original orientation.

The notion of transporting the frame on the earth's surface is not entirely metaphorical. Imagine setting up a frame of reference, consisting of two orthogonal axes, at some point on the Equator. Then transport the frame to the North Pole by sliding it parallel to itself, first by heading due north along a meridian, then by following the Equator for a distance before turning north along a different meridian. When the two frames meet again at the pole, their orientations will be different. The discrepancy results from the curvature of the earth.

There are special configurations of a gauge field where a frame transported around a closed loop does return to its initial orientation. In those configurations the rotation of the frame during a displacement is independent of the path taken. The gauge field is then said to be in a pure gauge form. It carries no energy and is in its vacuum state. When the final orientation of the frame does depend on the path, the gauge field carries energy, which becomes larger as the dependence on the path becomes more pronounced.

It is now possible to reexamine the energy of a two-dimensional field that includes a soliton. The introduction of a gauge field eliminates the potential energy of the field $\vec{\varphi}$ outside the domain *D*. Moreover, the gauge field itself is in a pure gauge form there and carries no energy because the frame of reference returns to its original orientation after any closed loop. The gauge field cannot continue in a pure gauge form throughout the interior of *D*, however. The frame rotates by 360 degrees when it is carried around a closed loop outside *D*, but there must be a small loop inside *D* where it does not rotate at all. By extension there must be loops of intermediate size where the frame rotates by some angle between 360 and zero degrees. In that case the frame does not return to its original orientation after the circuit, and so the gauge field must carry energy.

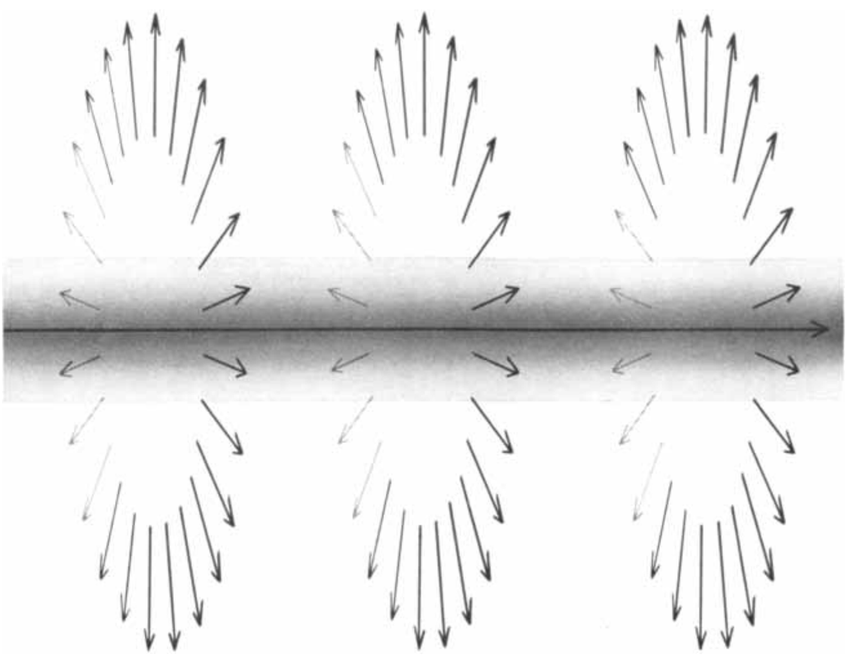
The twist in the field configuration outside *D* is now embodied in the gauge field. Both fields, the $\vec{\varphi}$ field and the gauge field, are in their vacuum state outside *D*, but the continuity of the field configuration demands that the fields carry energy inside the domain. The soliton that emerges is again made stable against dispersion by the topology of the field configuration.

The prototypical gauge field is the one that appears in Maxwell's theory of electromagnetism. It describes the relative orientation of frames of reference in an abstract space with two dimensions. The components of a vector field $\vec{\varphi}$ lie in this space. This field is not the electromagnetic field itself but describes the charged matter to which the electromagnetic field is coupled. In realistic situations $\vec{\varphi}$ and the gauge field are both defined at all points in three-dimensional space, although the abstract space where $\vec{\varphi}$ lies is still two-dimensional, the two dimensions corresponding to the two possible signs of charge.

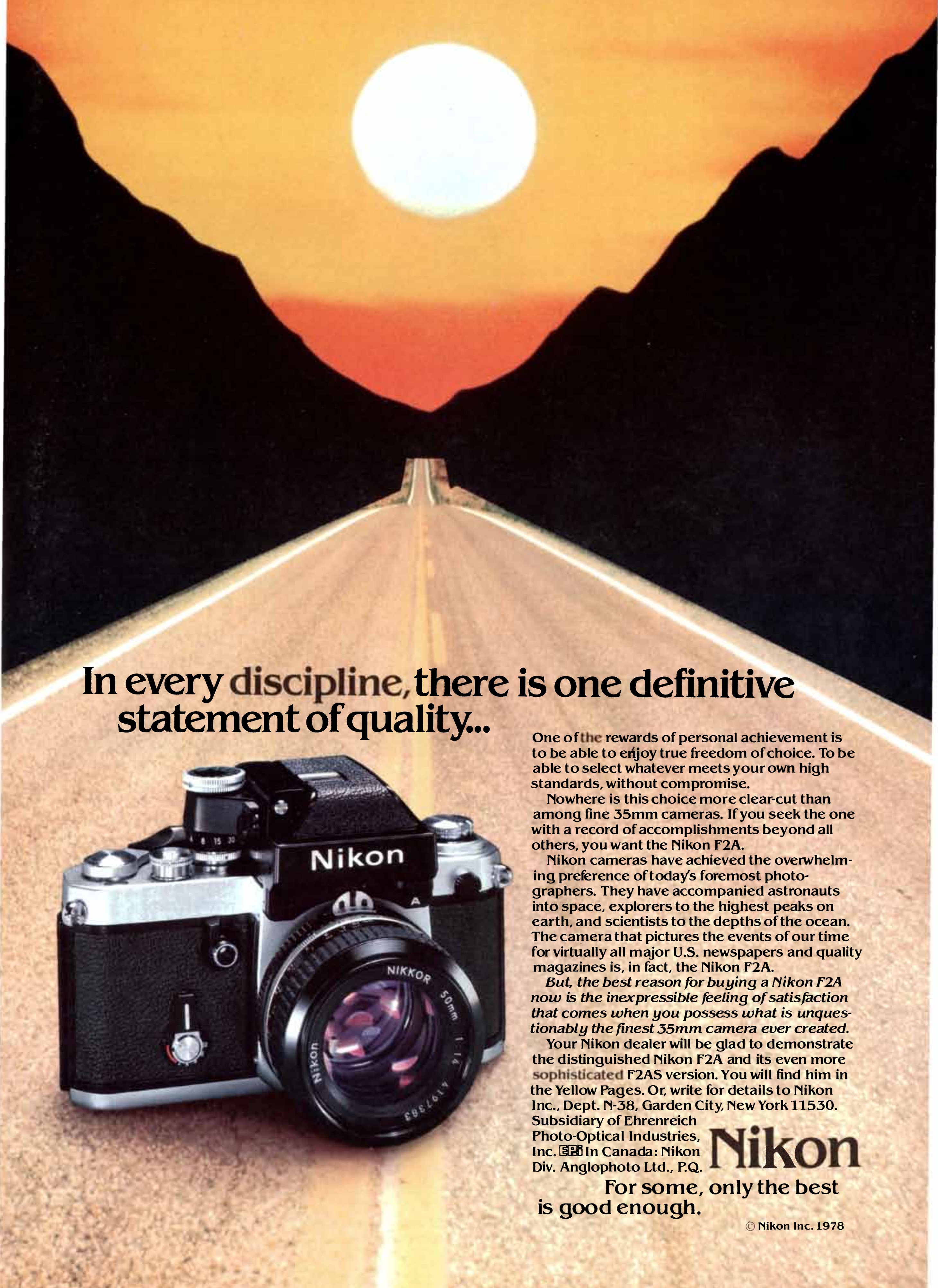
In most configurations of the electromagnetic field solitons are not possible because the minimum of the intrinsic energy corresponds to a zero value of the matter field $\vec{\varphi}$. In superconductors, however, the minimum is found at a field of nonzero magnitude. The field configuration then allows the existence of two-dimensional solitons, which do appear in superconducting materials. They manifest themselves as tubes of magnetic flux, where the fields are arranged in a vortexlike structure and the energy is confined to a narrow tube at the center of the vortex. When the vortexes are viewed in cross section, they resemble the two-dimensional soliton described above.

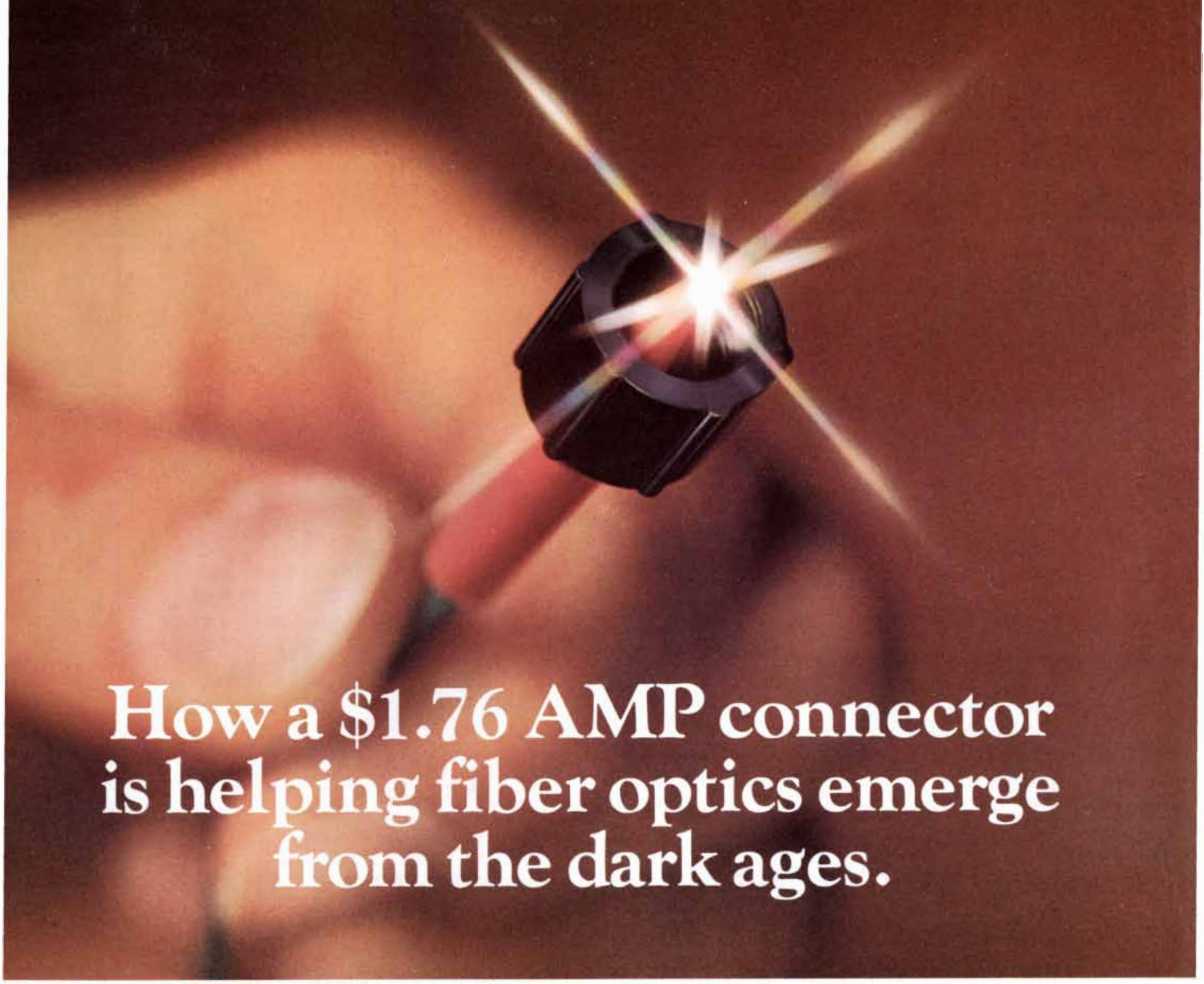
A gauge field defined in an abstract space with more than two dimensions is substantially more complicated than the electromagnetic field. The additional complexity arises because there are many axes around which the frame of reference can rotate, and in general the result of rotations performed around different axes depends on the order in which they are made. Gauge fields of this kind are called Yang-Mills fields, after C. N. Yang of the State University of New York at Stony Brook and Robert L. Mills of Ohio State University, who first discussed them in 1954. It was some time before the mathematical techniques needed to deal with the fields were mastered, but Yang-Mills fields have now been given a crucial role in the physicist's description of nature. A Yang-Mills field is the essential element in a theory that seems to unify two of the four fundamental forces of nature, the weak force and the electromagnetic one. Another Yang-Mills theory, although it is less well developed, might eventually explain a third fundamental force, the strong or nuclear force.

The discovery that solitons might be generated in a theory that describes a matter field coupled to a Yang-Mills field was made independently by Gerhard 't Hooft of the University of Utrecht and by Alexander M. Polyakov of the Landau Institute for Theoretical Physics in Moscow. By recognizing that the theory can have multiple vacuum states with a nontrivial topology, they discovered the three-dimensional soliton. It is a soliton of this kind that might exist as a real elementary particle. Such a soliton particle would be very massive



TUBE OF MAGNETIC FLUX in a superconductor is confined to a small cross section by a field that incorporates many two-dimensional solitons stacked with cylindrical symmetry. The superconducting electron pairs are described by a vector field with multiple vacuum states and with an intrinsic energy that is at a minimum when the magnitude of the field is greater than zero. The magnetic field is described by a gauge field. The twist in the topological configuration of the two fields bounds the magnetic flux and prevents it from spreading out in space.


In every discipline, there is one definitive statement of quality...
Nikon
One of the rewards of personal achievement is to be able to enjoy true freedom of choice. To be able to select whatever meets your own high standards, without compromise.
Nowhere is this choice more clear-cut than among fine 35mm cameras. If you seek the one with a record of accomplishments beyond all others, you want the Nikon F2A.
Nikon cameras have achieved the overwhelming preference of today's foremost photographers. They have accompanied astronauts into space, explorers to the highest peaks on earth, and scientists to the depths of the ocean. The camera that pictures the events of our time for virtually all major U.S. newspapers and quality magazines is, in fact, the Nikon F2A.
But, the best reason for buying a Nikon F2A now is the inexpressible feeling of satisfaction that comes when you possess what is unquestionably the finest 35mm camera ever created.
Your Nikon dealer will be glad to demonstrate the distinguished Nikon F2A and its even more sophisticated F2AS version. You will find him in the Yellow Pages. Or, write for details to Nikon Inc., Dept. N-38, Garden City, New York 11530. Subsidiary of Ehrenreich Photo-Optical Industries, Inc. EPOI In Canada: Nikon Div. Anglophoto Ltd., P.Q.
Nikon
For some, only the best is good enough.
© Nikon Inc. 1978



Many companies don't call AMP until they need electronic connectors or switches. But why wait till then? If you involve us early enough, we can help you save money. A lot of money.

Soon, tiny glass fibers transmitting light energy signals will be used where once hundreds of wires were needed.

They'll be adopted by many industries. And they'll touch our lives in many ways.

The future of fiber optics is bright. And one of the reasons is because we've already found a better way to connect the fibers.

Our connector precisely aligns each fiber automatically so that it can be polished and connected quickly. It's made a simple few-minute task out of a job that took skilled operators up to an hour.

Already, significant cost savings can be seen. Clearly.

Can we produce the same kind of savings for you?

If your company manufactures electrical/electronic products, the answer is probably yes. But it's important for you to call us early.

AMP
INCORPORATED

AMP has a better way.

That's when our experience in developing connectors, switches and application equipment for hundreds of industries worldwide can help you the most.

Early Involvement

At AMP, we call this approach to solving our customers' problems "Early Involvement." It's our better way. And it's what makes a $1.76* AMP connector worth so much more than $1.76.

Ask for a copy of our brochure "AMP Has a Better Way." Call (717) 564-0100, Ext. 8420.

Or write to AMP Incorporated, Harrisburg, PA 17105.

*Quantity price for two connector halves and splice bushing

Barcelona • Brussels • Buenos Aires • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • London • Luzern • Mexico City • Montreal
Paris • San Juan • Sao Paulo • Stockholm • Sydney • Turin • Toronto • Tokyo

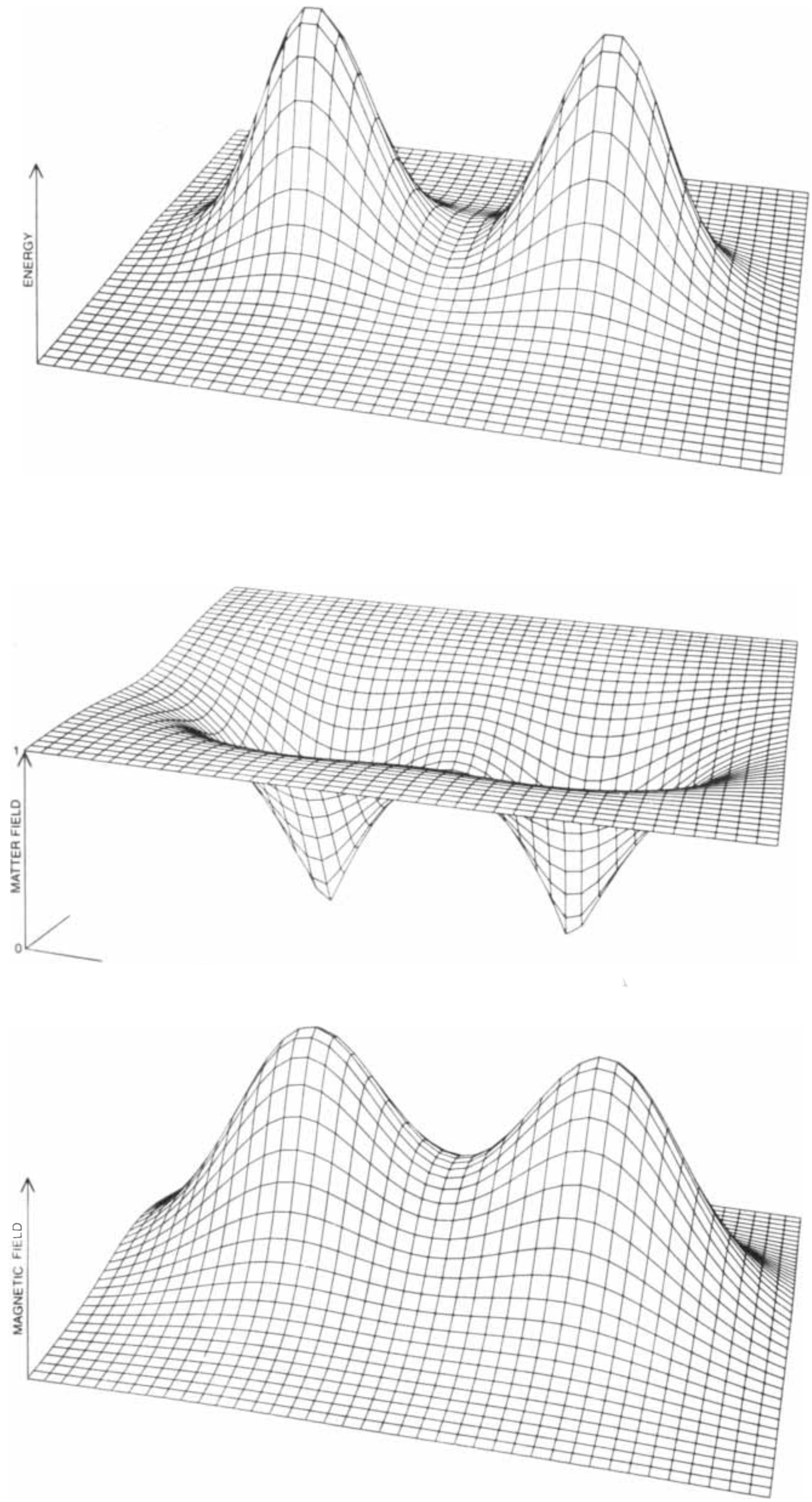
and would be a magnetic monopole, that is, it would carry an isolated magnetic charge. This property, which none of the known particles share, also derives from the topology of the field; in the same way that the twisted arrangement of vacuums confines energy it also encloses a unit of magnetic charge.

The theoretical exploration of the properties of these particles has only just begun, but a few interesting findings have already emerged. One result concerns relations between two fundamental categories of particles, the fermions and the bosons. These categories are distinguished by the intrinsic angular momentum, or spin, of the particles, and by their statistics, or behavior in groups. The spin of a fermion is a half integer, the spin of a boson an integer. The quantum-mechanical statistics of the particles specify that no two fermions can occupy the same state, whereas bosons can be brought together in unlimited numbers. Two fermions can combine to form a composite particle with the properties of a boson, just as two half integers add up to an integer. It would appear, however, that there is no way to combine bosons to yield a fermion.

When the field has only a global vacuum, the prohibition on making fermions out of bosons is indeed absolute, but it is not so when solitons are present. A mechanism for the conversion has recently been found by 't Hooft and Peter Hasenfratz, who is now at the University of Budapest, and independently by Roman W. Jackiw of the Massachusetts Institute of Technology and me. In the presence of a soliton a system with half-integer spin can emerge from a field whose only components are bosons. Alfred S. Goldhaber of Stony Brook has shown that the system would have not only the spin characteristic of a fermion but also the statistics.

Another novel result, obtained by Jackiw and me, shows that a fermion might be split in half under the influence of a soliton. We found a mode of interaction between solitons and fermions where the structure of a soliton is altered by the field of a fermion. The soliton exists in two states of identical energy, one state having the character of half a fermion and the other state that of half an antifermion.

The prospects for making and detecting soliton particles in the laboratory are quite uncertain. They depend in large measure on what theory is ultimately found to describe most accurately the interactions of elementary particles. If the theory is one that admits soliton solutions, then it is generally acknowledged that the solitons will appear in nature. There are many candidate theories, however; some of them incorporate field equations with a topology suitable for solitons but others do not. It is no surprise that soliton particles have not yet been observed in experiments



MERGER OF TWO FLUX TUBES in a superconductor illustrates the coalescence of two solitons. The matter field, whose magnitude is related to the density of superconducting electron pairs, is shown in the middle graph; this is the field in which the solitons appear. The matter field has two depressions, where the density of electron pairs falls to zero; everywhere else it rises toward a uniform value of 1. The energy of the field, shown in the upper graph, is zero at the periphery and rises to a maximum at each point where the density of electron pairs is zero. Thus the two regions where the matter field declines represent confined quantities of energy. The magnetic field passing through the solitons is shown in the lower graph; it too is confined to the regions of diminished electron-pair density. The graphs were constructed with the aid of a computer by the author and Laurence Jacobs of Brookhaven National Laboratory.

with particle accelerators. The mass of such a particle, measured in energy units, is thought to be some trillions of electron volts. One trillion electron volts is roughly 1,000 times the mass of a proton and more than four times the mass of a uranium atom. It will be at least several years before accelerators can create such heavy particles.

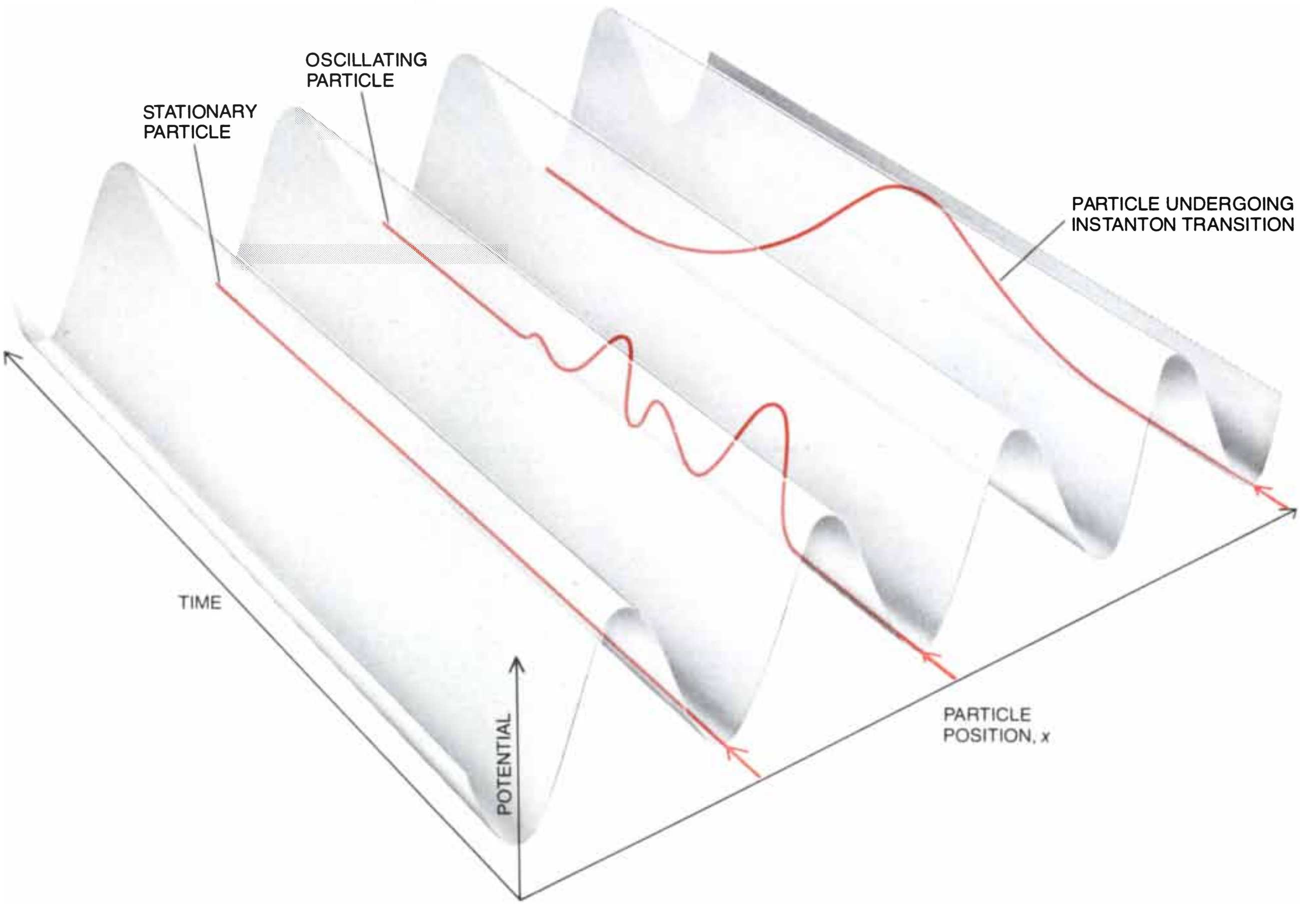
All the solitons described above are structures that are localized in space (whether the space has one, two or three dimensions). They are of interest because they are confined permanently to a definite region of space. In the past few years another kind of soliton has been discovered, one that is confined to a small region both in space and in time. It is a phenomenon that exists only at a particular place and at a particular moment. This new kind of soliton, which has been given the name instanton, is interpreted not as an object but as an event, not as a particle but as a quantum-mechanical transition between various states of other particles.

The nature of an instanton can be elucidated by examining the motion of a particle under the influence of a potential, which specifies the value of a force acting on the particle. For the sake of simplicity the potential can be made one-dimensional, and for instantons to appear it must be made periodic, that is, there must be several equivalent but distinct points of minimum energy. There are many examples of such potentials in nature; a familiar one is a roller coaster, where the potential is the gravitational field of the earth and a point of minimum potential energy is encountered at the bottom of each drop. It should be pointed out that the potential energy in this system is not that of a field but that of a particle in a field. In the example of the roller coaster potential energy is proportional to height above the lowest point on the track.

A graph of a simple periodic potential shows a straight line, which represents the space where the particle moves, and an undulating line (perhaps a sine curve), which gives the value of the potential for each point in the one-dimensional space. In order to trace the evolution of the system another axis, representing time, must be added to the graph. The undulating line is then converted into an undulating surface. Moving across the undulations is equivalent to a change in position; moving parallel to them is not a motion in space but instead reveals a steady state of the system at successive moments.

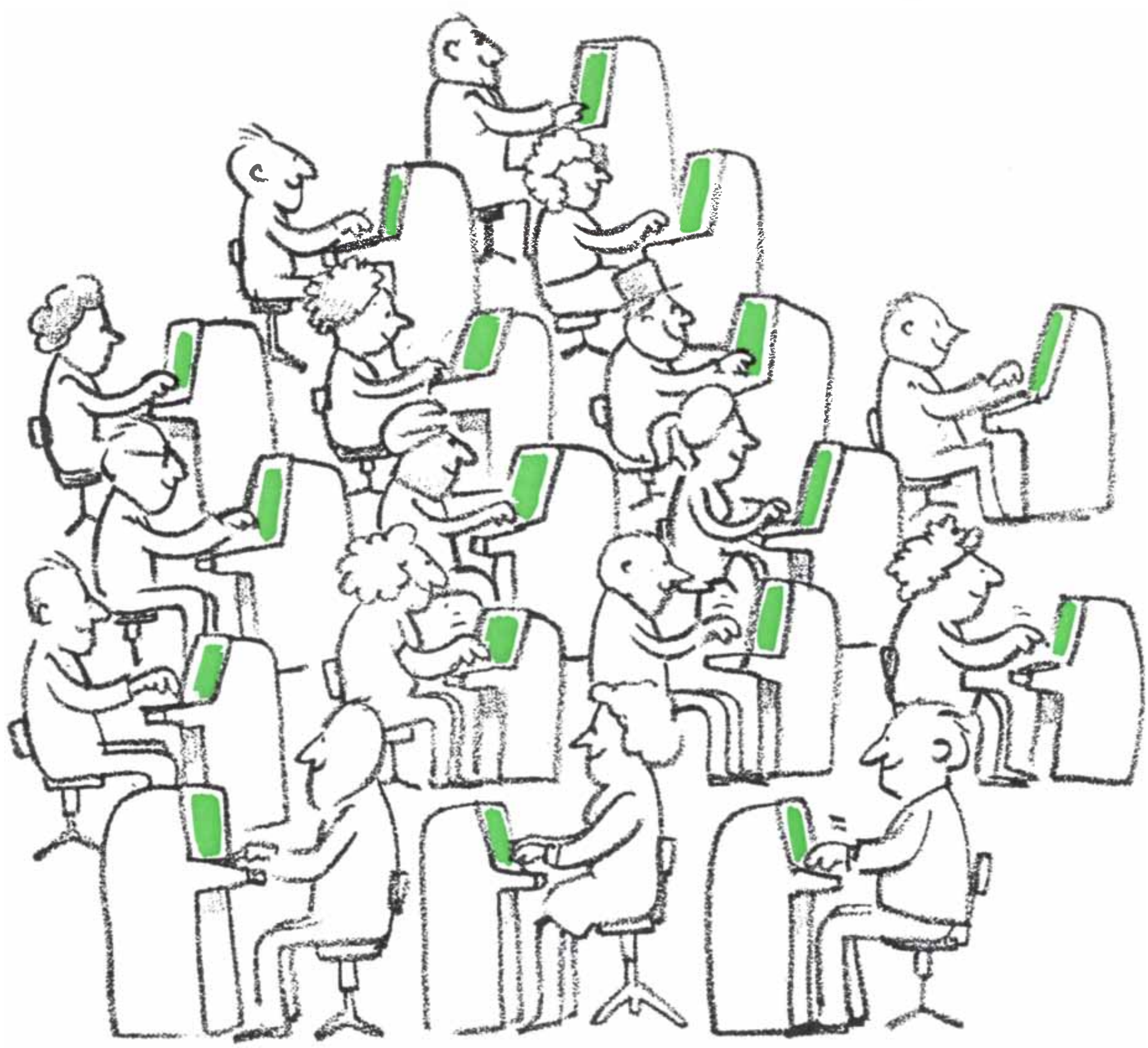
A particle in a state of minimum energy lies at the bottom of one well in the potential curve, and it can be expected to remain at rest there forever. The trajectory of the particle in time is therefore a straight line that follows the bottom of the valley. A particle might also oscillate about the equilibrium point at the bottom of a well, a motion that is represented along the time axis by a sinuous line wandering from one side of the valley to the other.

The event of greatest interest is the movement of a particle from a stable position at the bottom of a well over a potential barrier and into an adjacent well. On the time axis this evolution corresponds to a line that begins in one valley of the potential surface, climbs over



INSTANTON is a soliton confined not only to a region in space but also to a moment in time. It is interpreted as a quantum-mechanical transition between two states of motion of a particle. The particle moves in a one-dimensional potential, or force field, and the evolution of its motion is recorded on a potential surface that extends along the time axis. The history of a particle that remains stationary at one of the points of minimum potential is represented by a straight line that follows a valley in the surface. A particle oscillating about an equilibrium point traces an undulating curve in one of the valleys. The trajectory of an instanton climbs from the bottom of one valley, crosses a ridge and returns to stable equilibrium in an adjacent valley. The path describes the motion of a particle that disappears at one point of equilibrium and reappears at another. The transition is equivalent to the quantum-mechanical phenomenon of tunneling.

HOW TO GET THE MINI THAT'S USED BY THE MANY.



When you buy an NCSS 3200 computing system, you get software that's performance-proven by over 10,000 customers in over 5 million hours of on-line use.

So, unlike other systems, you're no guinea pig for untested software.

With the NCSS 3200, you get a system that starts off delivering what others may take years to achieve.

You also get a system that's backed by an international network of 3200-compatible machines. A phone call is all it takes to put you on-line and increase your capacity instantly.

You can save a bundle on software maintenance because we'll do it for you. Just as we now do for thousands of on-line customers.

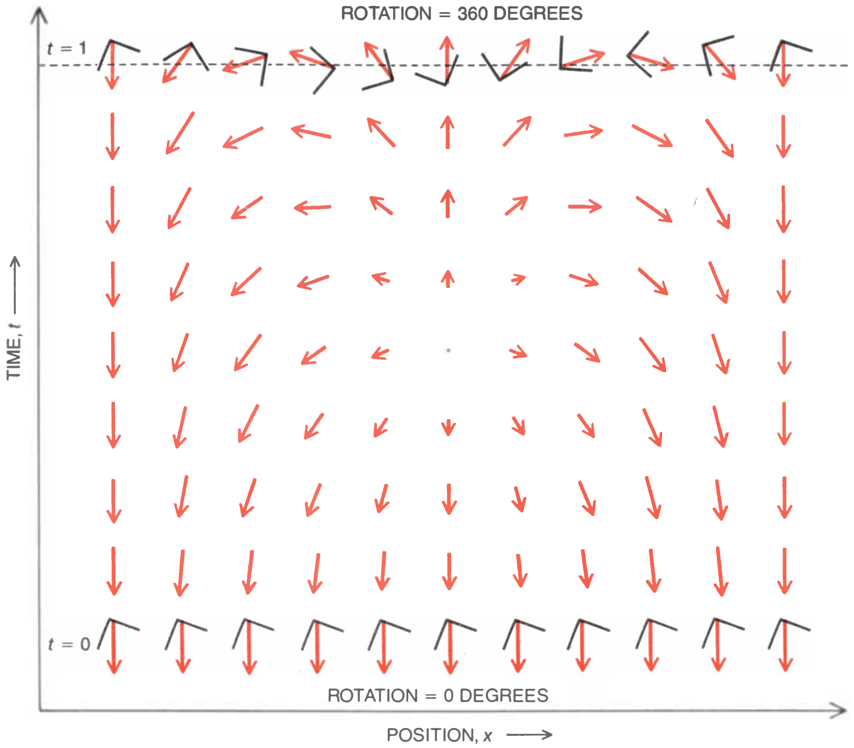
So if you're ready to invest in a mini, get the one that makes your investment pay off right from the start. NCSS 3200...the mighty mini with mainframe muscle.

For more information, write or telephone toll free:

NATIONAL CSS, INC., Dept. R-D
187 Danbury Road
Wilton, Connecticut 06897
(800) 243-6119
In Connecticut: (800) 882-5575

NCSS 3200

WE SOLVE YOUR INFORMATION PROBLEMS ONCE AND FOR ALL.



EVOLUTION OF A FIELD during an instanton transition exhibits a topological twist much like that of a purely spatial soliton. Initially ($t = 0$) the field is uniform, being made up of vectors that are all parallel and all of unit magnitude. This configuration is assumed to describe a vacuum state. At a later moment ($t = 1$) the vectors again have unit magnitude, but they are found to rotate through one turn. If the frame of reference is defined by an appropriate gauge field, the later configuration can also be regarded as a vacuum state. For the field to evolve from the first state to the second one, however, its net rotation must change from zero to 360 degrees, a transition that is topologically possible only if the field vector somewhere declines to zero. At that point the intrinsic energy of the field is greater than zero, and so the instanton describes an evolution from one vacuum state to another through a state with finite energy.

a ridge and descends into a neighboring valley. At one moment the particle is stationary at a point of minimum energy; the next moment it is found still at rest but at another minimum-energy position. The transition between these states is an instanton.

In the classical physics that preceded the development of quantum mechanics such a transition was impossible. If a train of roller-coaster cars is at rest at the bottom of a hill, one can predict with confidence that it will not spontaneously climb over the hill and come to a stop in the next valley. Energy would be conserved in this imaginary process—at least for an ideal, frictionless roller coaster—because all the energy that would have to be expended to raise the train to the top of the hill could be recovered on the downhill run. The transition is nonetheless forbidden because in classical physics energy must be conserved at all moments, not merely in a final accounting.

Quantum mechanics provides a kind of deficit financing that makes the instanton transition possible. A seeming violation of the conservation law is allowed provided that the violation does not last too long and that the books balance in the end. Through this mechanism, which is called tunneling, a particle can cross a potential barrier even though it has too little energy to surmount the barrier. The instanton is a structure in a classical field theory that describes this fundamentally quantum-mechanical process.

The graph of an instanton is geometrically identical with the graph of a soliton in the sine-Gordon field; only the labels on the axes have been changed. Indeed, instantons were first found as ordinary solitons in a field theory of four spatial dimensions. The discovery was made by A. A. Belavin, Polyakov, A. S. Schwartz and Yu. S. Tyupkin of the Landau Institute, who correctly interpreted the soliton not as a four-dimensional object but as an evolution of fields in three spatial dimensions and one time dimension. They also showed that instantons must appear in a large class of field theories, including the theories most commonly applied to the interactions of elementary particles.

Soon afterward the significance of instantons for the structure of the quantum-mechanical vacuum was analyzed by 't Hooft, by Curtis G. Callan, Jr., and David J. Gross of Princeton University and Roger F. Dashen of the Institute for Advanced Study and by Jackiw and me. (It was 't Hooft, incidentally, who suggested the name instanton.) The existence of instantons implies that the vacuum state in quantum mechanics is not unique but has a periodic structure, rather like the potential valleys of the electromagnetic field inside the lattice of atoms in a crystal. Of course, the periodic field in a crystal results from an orderly arrangement of atomic nuclei, whereas the pattern generated by the instanton solutions is an intrinsic structure of space-time. The discovery of the structure had been quite unexpected.

Although instantons are a recent innovation in field theory, they have already led to the resolution of a disturbing puzzle in the physics of subnuclear particles. The puzzle concerns the masses of the particles called mesons, which are thought to be composite objects made up of the more fundamental entities called quarks. Each meson consists of a quark and an antiquark bound together by a gauge field. Two of the quark types (labeled u and d) and their corresponding antiquarks ($\bar{u}$ and $\bar{d}$) are thought to be comparatively light. Since there are four ways to make a quark-antiquark pair from these objects ($u\bar{u}$, $u\bar{d}$, $d\bar{u}$ and $d\bar{d}$), it would seem there should be four mesons of comparatively low mass. Three such mesons have been known for many years: they are the negative, positive and neutral pi mesons, or pions, which have masses equivalent to an energy of about 140 million electron volts. The fourth light meson, which seems to be an inescapable prediction of the theory, has never been found.

There is another particle, however, that could fill the role. It is the eta meson, and it has all the appropriate properties except one: its mass is about 550 million electron volts. The introduction of instantons has now explained the anomaly of the eta meson's mass. The instantons appear as excitations, localized in space and time, in the gauge field that binds the quarks together. They alter the distribution of mass among the mesons because they have different effects on the various quark combinations. Roughly speaking, an instanton is transparent to a pion, but it acts as an obstacle to the propagation of an eta meson and thereby increases its effective inertial mass.

Solitons and instantons are the offspring of field theories that can be forbiddingly complex, yet they have a rich and elegant mathematical structure. Indeed, physicists investigating solitons have discovered that mathematicians have been studying equivalent objects for many years entirely for their geometrical interest. Mathematical analysis and physical intuition have forged powerful tools to expose the nature and properties of the soliton.

The Source of the Earth's Magnetic Field

The source is believed to be a dynamo set up by the ponderous flow of matter in the earth's core, but what drives the flow? It may be the gravitational energy released by the sinking of heavy material

by Charles R. Carrigan and David Gubbins

It has been known since 1600 that the earth as a whole is a magnet, but the source of the magnetic field is still not fully identified. The source is not permanently magnetized minerals in the earth; most of the planet is much too hot for any magnetic material to retain its magnetism. Furthermore, permanently magnetized minerals cannot move about rapidly enough to account for the known long-term changes in the strength, direction and pattern of the earth's magnetic field. The analysis of earthquake waves indicates that at least part of the earth's core is fluid, and it is almost universally believed the ponderous flow of matter in the core generates electric currents, which in turn induce the magnetic field.

There is little agreement, however, on how the metallic fluid in the core flows, on what energy source drives the fluid's motion and on how that motion gives rise to the magnetic field. Because the core cannot be scrutinized directly and because its high temperatures and pressures cannot be reproduced easily in the laboratory these questions are hard to answer. Nevertheless, answers are starting to emerge from a combination of innovative theoretical and laboratory approaches. Such answers suggest that the core-fluid motions might be driven by gravitational energy liberated by the migration of heavy materials to the center of the core and of light materials to the outside of the core.

Only a few things are known with any certainty about the earth's core and its magnetic field. Seismological, or earthquake-wave, data indicate that the core consists of a giant metallic sphere about the size of the planet Mars. Under normal conditions the core fluid conducts heat and electricity much better than copper does, and it probably has about the same viscosity as water. With a mean radius of 3,485 kilometers, the core constitutes about a sixth of the earth's volume and about a third of its mass. The density of the core varies from a minimum at its edge of nine times the density of water to a maximum at its center of 12 times the density of water. The density estimates combined with hypotheses about the origin of the solar system suggest that the core is composed mostly of iron and nickel with a trace of lighter elements such as copper, sulfur and oxygen.

At the center of the core lies an inner core whose seismic properties are somewhat different from those of the surrounding region. With a radius of 1,220 kilometers, the inner core is about two-thirds the size of the moon and unlike the outer core is probably solid. Pressures at the center measure in millions of atmospheres, and at those pressures the melting point of iron lies between 3,000 and 5,000 degrees Kelvin.

The easiest way to get an idea of the strength and direction of the earth's magnetic field is to study the behavior of a compass needle. The strength of the field, as measured by the force needed to deflect a needle from its preferred direction, is extremely small. The maximum field of .3 gauss, near either the North Pole or the South Pole, is several hundred times weaker than the field between the ends of a toy horseshoe magnet. The needle tends to point along north-south arcs around the earth, but in few places does it point precisely to geographic north, which is one end of the earth's axis of rotation. Instead its direction varies sporadically from region to region, so that the field is seen to consist of irregular eddies.

The general north-south alignment of the lines of magnetic force with the earth's axis of rotation suggests that the field is mainly a dipole (the field that would be created by a loop of electric current in the plane of the Equator). The field, which is tilted at an angle of 11 degrees to the axis of rotation, is not a perfect dipole. The field cannot be descended from a magnetic field of the primordial earth because the electrical conductivity of the core material is too low. Without a constant supply of energy the electric currents maintaining the field would have died out in less than 10,000 years. And the field has clearly existed for a much longer time.

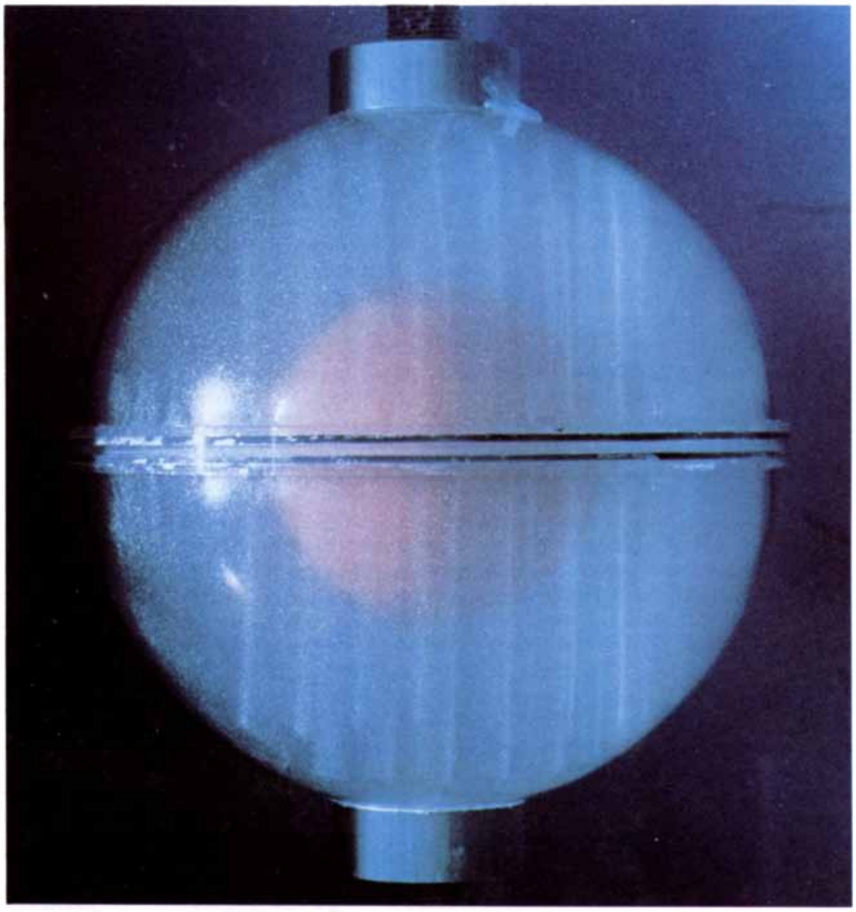
Since the 17th century elaborate charts of the strength and direction of the magnetic field at points all over the earth's surface have been made for the purposes of navigation at sea. Such charts provide a detailed record of the secular (long term) variations in the field over the past four centuries. The charts reveal two major trends in the secular variations: first, a slow but steady decrease in the intensity of the main dipolar field at a rate such that if it continued, the field would vanish altogether in 3,000 years, and second, a slow westerly drift in the field's irregular eddies amounting to one degree of longitude every five years. The drift suggests that fluid in the core is moving at a rate of about one millimeter per second, or about 100 yards per day.

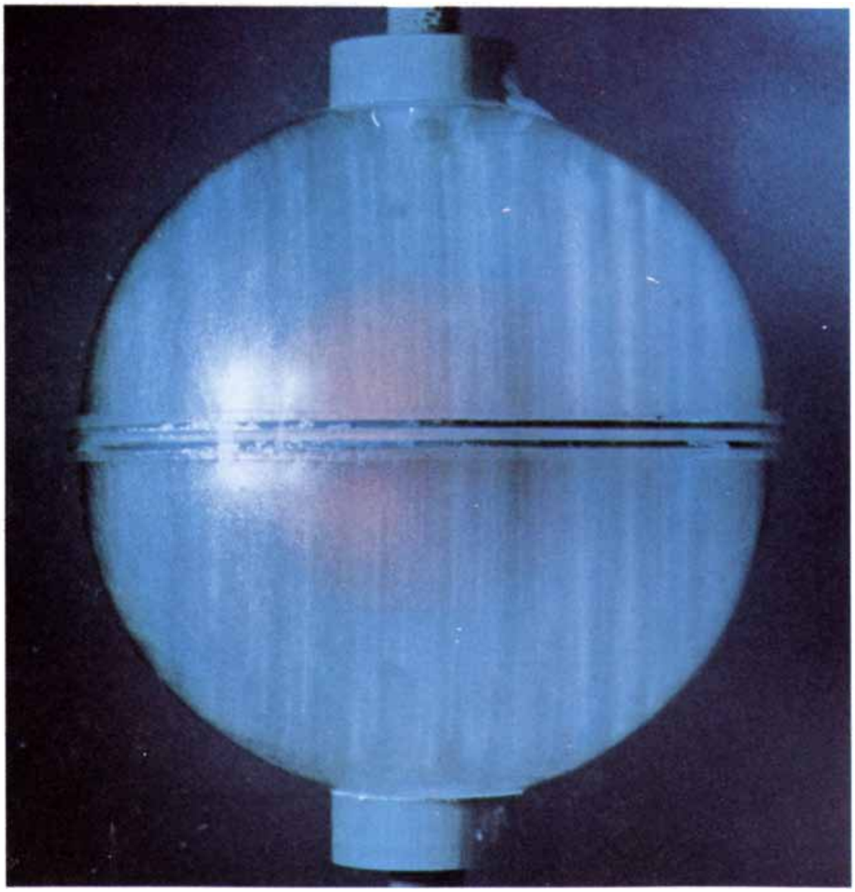
Rocks in the earth's crust also provide information about the magnetic field. The rocks serve as records of how the strength and direction of the field have evolved over geologic time. In paleomagnetic experiments the rocks' age and permanent magnetization are measured. From these measurements the field's characteristics can be dated, because the field presumably caused the permanent alignment of magnetic domains within the rocks at the time of their formation. Such experiments indicate that the earth has had a significant field for at least 2.7 billion years, a period that spans a substantial fraction of the earth's 4.6-billion-year existence. During all this time the field has fluctuated in magnitude and every million years or so has reversed its direction.

Any hypothesis about the source of the field must account for its mainly dipolar form, its slow decrease in strength, its westerly drift and its pole reversals. It is difficult to form a reasonable hypothesis because the 3,000 kilometers of matter that lies between the earth's surface and the top of the earth's core obscures what the field is really like at the site of its origin in the earth's interior. Near the core the field is probably about 10 times stronger than it is at the surface. In addition the field there is probably much more complicated than the surface dipole and is probably changing at a much faster rate.

Such rapid changes cannot be detected at the earth's surface because they would be filtered out by electrical conduction in the intervening mantle. Nothing is known about the intensity of the field inside the core itself. In particular it is impossible to detect toroidal fields: fields whose lines of force lie parallel to spherical surfaces. Theoretical considerations indicate that toroidal fields, of which the simplest type has the form of a ring, are very likely present and may be much stronger than the surface dipole. The impossibility of directly detecting toroidal fields is a major obstacle to constructing models of the fields in the core.

Today only one model of the core has survived: the "self-sustaining dynamo" model, developed in the 1950's by Walter M. Elsasser of the University of California at San Diego and E. C. Bullard of the University of Cambridge. A dynamo is any machine that converts mechanical energy into electrical energy. A simple example of such a machine is the disk dynamo invented by Michael Faraday. He put a bar magnet under the edge of a copper disk mounted on a spindle. When the disk was spun on the spindle, a small current was generated in the disk by its motion through the magnet's field. Because the current exerts a





ROTATING PLASTIC SPHERE that consists of a smaller solid sphere (*pink*) surrounded by water is a model of possible fluid motions in the core of the earth that give rise to the earth's magnetic field. To generate buoyancy forces a temperature gradient of one degree Kelvin was imposed between the inner sphere and the outer one. When the buoyancy forces are just strong enough to drive the fluid against the retarding effects of forces resulting from viscosity, the fluid motions are highly ordered small-scale circulations that take the form of long, slowly turning rollers that are aligned parallel to the rotation axis of the sphere (*top*). When the temperature gradient is increased, the buoyancy forces are strong enough to drive circulations throughout the entire fluid (*bottom*) rather than throughout only a thin layer adjacent to the inner sphere (*top*). Microscopic platelets were added so that the circulations could be seen in these photographs. Original model was built by Friedrich H. Busse and one of the authors (Carrigan) at the University of California at Los Angeles.

force in opposition to the disk's motion, the disk slows down. In short, rotational mechanical energy is converted into electrical energy.

In a self-sustaining dynamo such as the earth's core the generated electric currents serve to reinforce the magnetic field so that no external magnetic-field supply is needed beyond that which originally triggered the dynamo. A simple example of a self-sustaining dynamo is a Faraday disk dynamo where the bar magnet is replaced by a coil of wire. When an electric current is fed through the coil, a magnetic field identical with that of a bar magnet will be created. Hence the coil induces a current in the disk in exactly the same way that a bar magnet does. If the current induced in the disk is fed back into the coil, the magnetic field will be sustained, provided the disk is kept spinning.

The earth's dynamo is self-sustaining: after having been started with a magnetic field that could be very small, the dynamo will generate its own field without any external supply of magnetism. The earth's dynamo could have been initially activated by the small magnetic field that permeates the entire galaxy. The dynamo mechanism would then take over and generate a much stronger field of its own. Although the metallic liquid in the earth's core obviously does not resemble a solid disk, it could in principle flow in such a way that it would act as a mechanical dynamo.

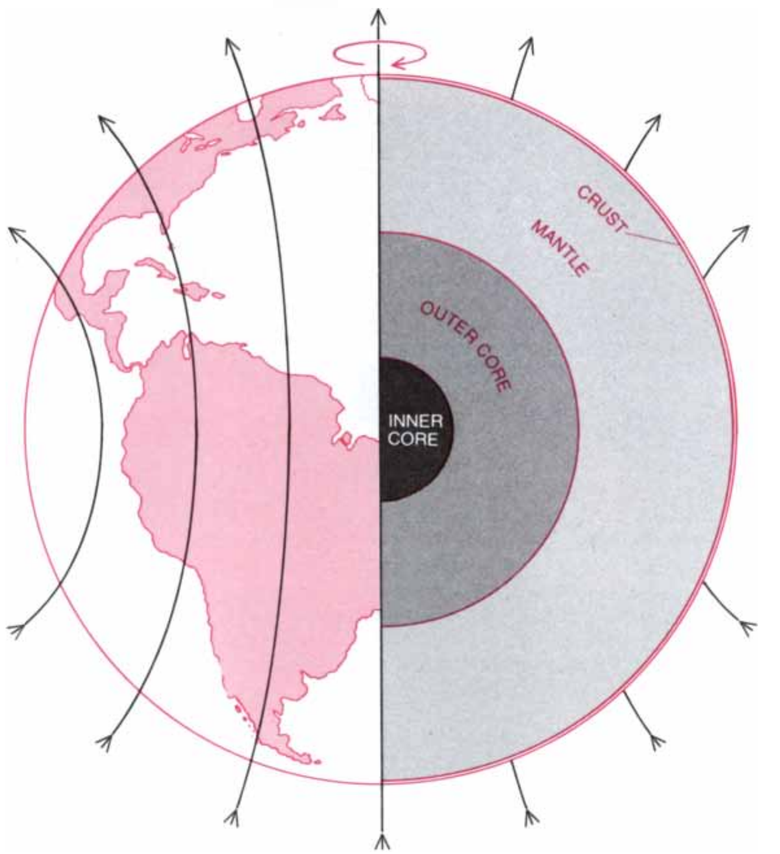
The crucial question is how the core liquid flows to act as a dynamo. Although a self-sustaining dynamo does not require a constant supply of magnetic field, it does require a constant supply of mechanical energy to keep the conducting material moving. In the case of the earth's core this means not only that the metallic fluid must flow in the right manner but also that some energy source must sustain the flow.

It is time now to consider in detail a possible mechanism by which the magnetic field is generated. Some form of convection initially stirs the core fluid, which then moves in dynamo fashion because of the forces due to gravity, magnetism and rotation. Rotation clearly must play a fundamental role in the origin of magnetic fields, since not only the earth but also other planets, the sun and other rotating stars have magnetic fields that are aligned or related to their axis of rotation. The connection between magnetism and rotation is so striking that in the early part of this century many physicists, including Albert Einstein, contemplated a new physical law that would require every rotating mass to give rise to a magnetic field. Talk of such a law was finally abandoned in the 1950's because of two crucial experiments. P. M. S. Blackett of the Imperial College of Science and Technology failed to find a magnetic field around a rotating gold cylinder. And S. K. Runcorn of the University of Cambridge measured the strength of the earth's magnetic field at different depths in a mine shaft and failed to detect the characteristic values such a law would predict.

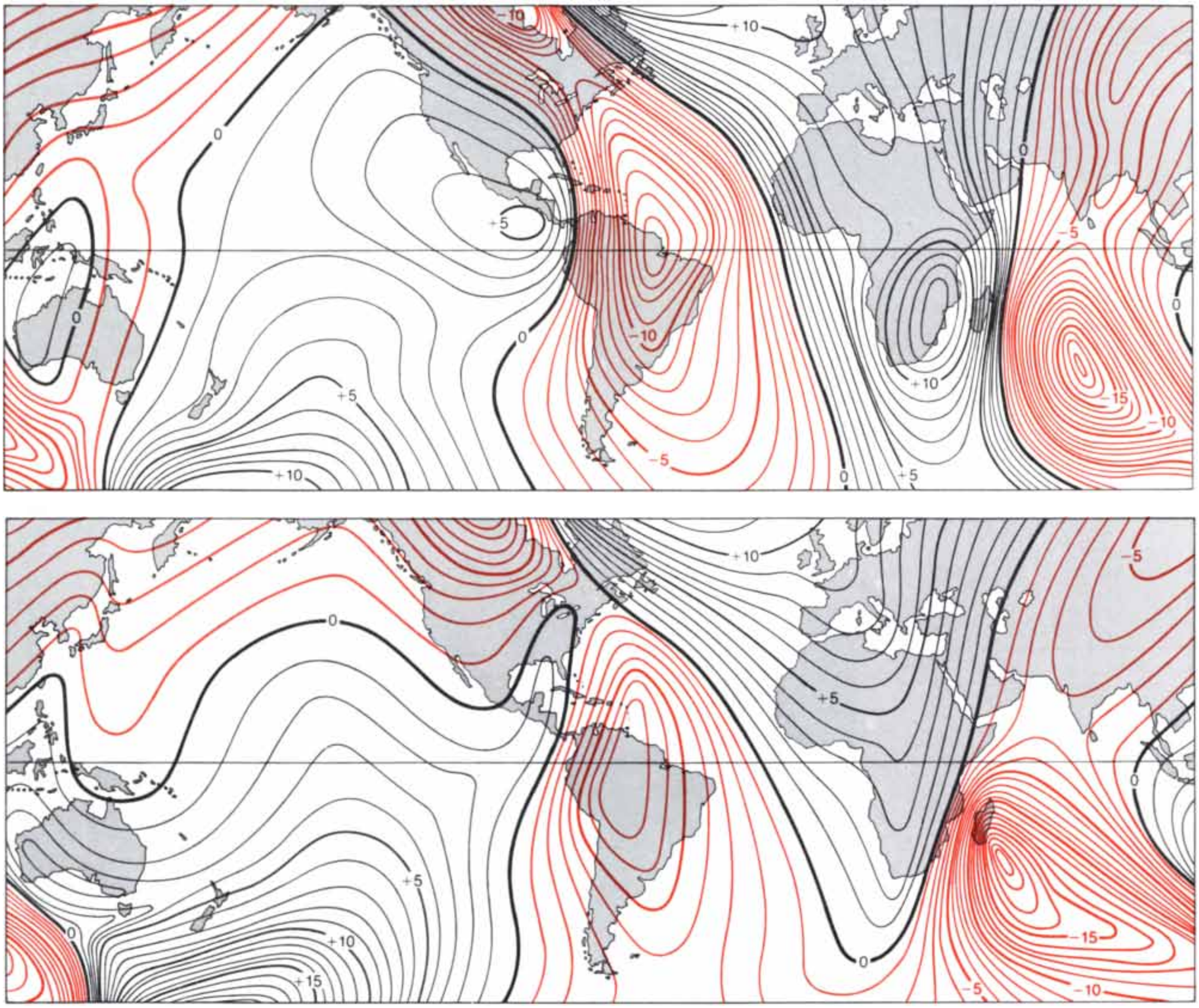
It is now fairly certain that rotation is linked to the origin of the earth's field through the Coriolis force that rotation exerts on the core fluid. The Coriolis force acts on any mass moving with respect to a rotating frame of reference. The mass is accelerated in a direction perpendicular to its motion. In the atmosphere and the oceans the Coriolis force is responsible for large-scale cyclonic motions of air and sea currents. The Coriolis force, however, can change only the direction of currents. It cannot change their speed because it acts only perpendicularly to the direction of flow, and so it cannot drive currents against the retarding effects of other forces.

Gravitational buoyancy is most likely to be the force that actually drives the motion. It causes less dense material in a dense fluid to float or convect upward. Gravitational buoyancy is the force that moves atmospheric currents, ocean currents and even continents. This force can also generate radial fluid motions, whose presence in the core is required by theoretical considerations of the earth's magnetic field. The buoyancy forces responsible for convection depend on density differences that can result from variations in the fluid's temperature and composition. Convection tends to redistribute the material in the fluid so that buoyancy vanishes. For the fluid to continue circulating, buoyancy forces must persist, and such forces can be maintained either by a continuous source of heat or by a continuous source of less dense material. In other words, since buoyancy forces do mechanical work, an unceasing supply of energy is needed to maintain them.

Geophysicists are now at the stage where they are trying to construct idealized models of the earth's core that exhibit some of the core's important features. The nature of convection in rotating fluids that are permeated by magnetic fields is under study. Friedrich H. Busse of the University of California at Los Angeles developed an idealized fluid-dynamical model in order to analyze the effect of a radial variation in temperature on a rapidly rotating fluid sphere. Like the gravitational field in the



EARTH'S CORE, whose radius is 3,485 kilometers, constitutes about a sixth of the earth's volume and about a third of its mass. At the center of the core lies a solid inner core whose radius is 1,220 kilometers. The flow of matter in the liquid outer core could generate electric currents, which in turn could generate field of the earth. Colored lines represent lines of force.

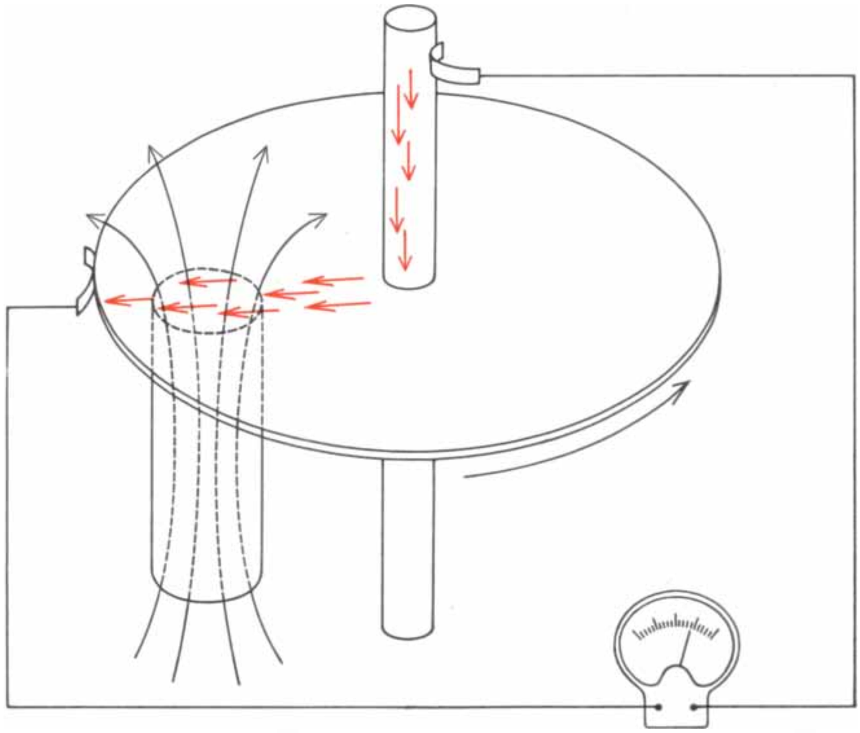


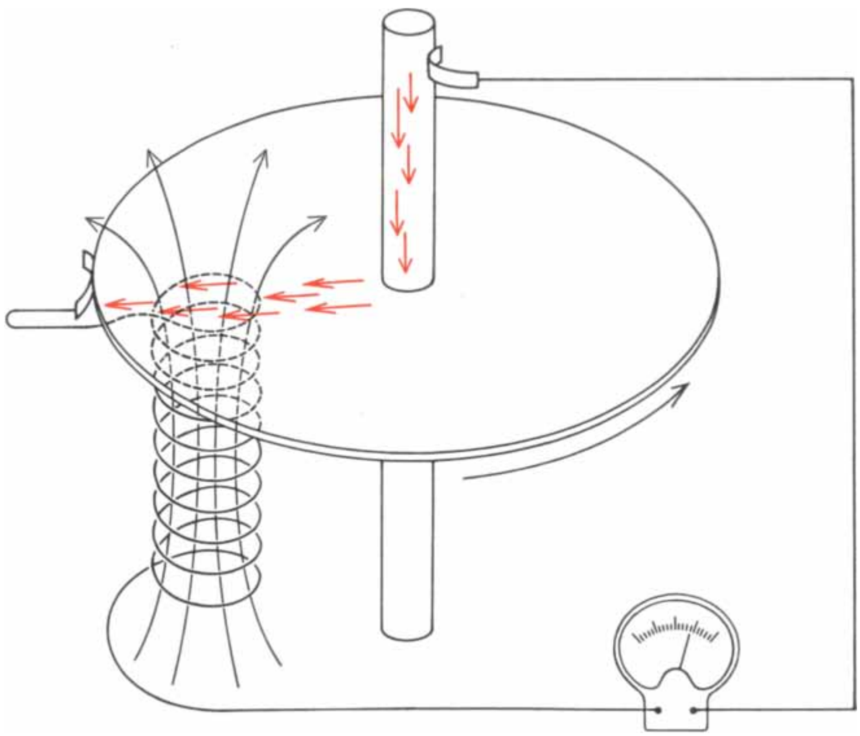
SECULAR VARIATIONS, or long-term changing eddies, in the magnetic field of the earth are indicated by the differences in the contours on these two maps of the world. The contours plot the rate of change in minutes of arc per year of the deviation of a compass needle from geographic north. The upper map shows the contours in 1912, the lower map the contours in 1942. A contour labeled with a positive number (*black*) indicates the rate at which a compass needle on the contour moved away from geographic north, a contour labeled with a negative number (*color*) the rate at which a needle on the contour moved toward geographic north. For example, on the North Island of New Zealand in 1942 a compass needle would have moved away from geographic north at five minutes of arc (+5) each year.

earth's core, the gravitational field in the sphere pointed inward toward its center. This early study ignored the effects of magnetic fields and dealt only with the effects of rotation on fluid flows that were driven by gravitational buoyancy resulting from temperature differences. Such results, however, should hold for other sources of buoyancy as well. Busse's theoretical calculations indicated that for sufficiently large temperature variations the buoyancy-driven flows are highly ordered by rotation and could give rise to the motions required for a dynamo.

On the basis of this analysis Busse and one of us (Carrigan), who was then (in 1973) a graduate student, set up a laboratory model of buoyancy-driven flows. The model consisted of a rapidly rotating plastic sphere that was filled with water, whose relevant properties supposedly match those of the core fluid. When the sphere was rapidly spun, the water "felt" the Coriolis force, just as the core fluid does. The sphere's small mass meant that its own gravitational force was negligible compared with the centrifugal force exerted on the fluid as a result of the rapid rate of rotation. In the earth's core, on the other hand, the gravitational force far exceeds the centrifugal force resulting from the core's daily rotation. It turns out that when the Coriolis force is as large as it is in the model, the model's strong centrifugal force serves to adequately simulate the effects of the core's strong gravitational force. Buoyancy could then be introduced in the model by the imposition of a radial temperature difference between the inner and the outer sphere.

How well the model represents what is going on in the core depends on how well the important forces can be scaled down while the correct proportions are maintained among them. If the laboratory model is spun fast enough, the motion of the water in the sphere may resemble the motion of the iron in the core. As with any scale model, the important parameters are the ratios of the various forces that are present. In fluid dynamics the ratio of the viscous forces to the Coriolis force is called the Ekman number. In the earth's core the Ekman





FARADAY DISK DYNAMO generates electric currents (*colored arrows*) when a copper disk is turned through the magnetic lines of force of a bar magnet (*top*) or a coil of wire (*bottom*). In a self-sustaining dynamo (*bottom*) the generated electric currents serve to reinforce the magnetic field of the coil so that no external supply of magnetism is needed beyond that which originally served to trigger the dynamo. The metallic liquid in the core of the earth is believed to flow in such a way as to act as a mechanical dynamo in generating magnetic field of the earth.

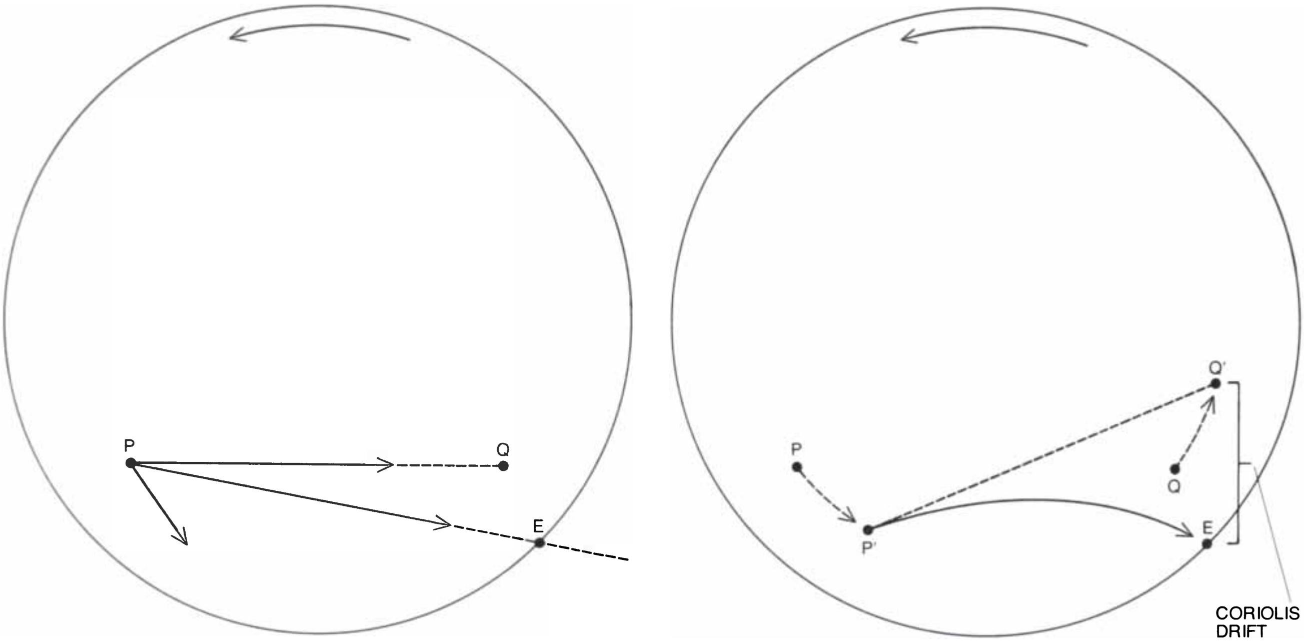
number may be as small as 10^{-15} because of the fluid's low viscosity and large-scale motions.

Whereas the viscosity of the water in the model may be comparable to the viscosity of the liquid in the core, the dimensions of the spheres in the model are of course millions of times smaller. Because of the small size of the model spheres viscous forces would dominate the Coriolis force if the model rotated only once per day, as the earth does. For the Coriolis force to be sufficiently strong compared with the viscous forces, or in other words for the Ekman number of the model to approach the Ekman number of the earth's core, the model must spin about 500 times per minute, or about a million times faster than the earth spins.

To preserve the correct ratio of buoyancy forces and viscous forces a temperature gradient of one degree K. was imposed between the sphere's center and its outside. If the buoyancy in the model had been too small, the water would not have circulated, because the driving forces would not have been large enough to overcome the retarding forces. If the buoyancy is just large enough for the water to start moving, the flow is confined to a narrow cylindrical region whose axis coincides with the sphere's axis of rotation. The flow within the cylindrical region consists of highly ordered small-scale circulations that take the form of long, slowly turning rollers aligned parallel to the rotation axis. The arrangement of the rollers resembles the arrangement of rollers in a machine bearing.

The roller appearance of the fluid motions is explained by a theorem of fluid dynamics. The theorem applies to rotating fluid bodies where the Coriolis force is much larger than other forces, such as those resulting from viscosity, buoyancy and the fluid's inertia. In that case only forces arising out of fluid-pressure differences are large enough to counteract the Coriolis force. The pressure forces are "conservative" in one sense: they cannot change the fluid motions along the direction parallel to the axis of rotation. In other words, if the rotating body were cut into thin slices at right angles to the rotation axis, then the theorem requires that the lines of flow in one slice be identical with the lines of flow in any other. In particular a cross-sectional slice of the roller-circulation pattern would look like the patterns in the slices directly above and directly below it.

In a spherical body such as the earth's core or the model, however, the spherical boundary prevents the flow from being identical in every cross section perpendicular to the axis of rotation. The boundary at each end of a roller is oppositely sloped, and so to an observer of the model outward circula-



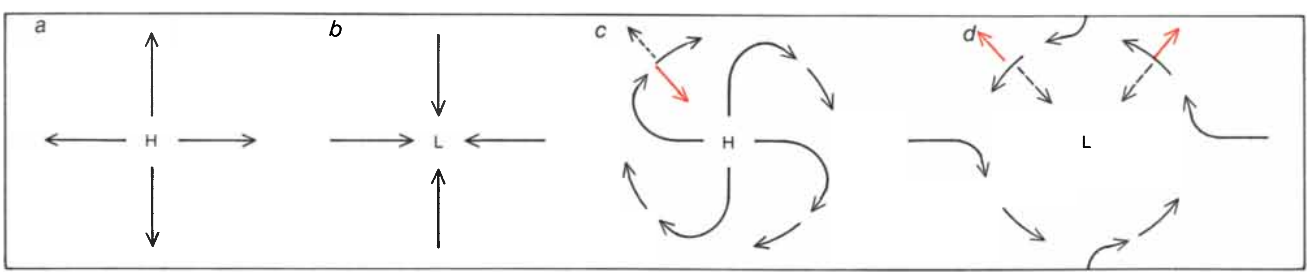
CORIOLIS FORCE, which acts on any mass moving with respect to a rotating frame of reference, may give rise to large-scale cyclonic motions of the fluid in the core of the earth. A merry-go-round that rotates counterclockwise as seen from above illustrates the effect of the Coriolis force. In the drawing at the left a man at *P* tries to throw a ball to a man at *Q*. The motion of the man at *P* (*short arrow*), however, causes the ball to head in the direction *PE*. In the drawing at the right the man at *P* has moved to *P'*, the man at *Q* has moved to *Q'* and the ball has moved to *E*. To observers on the merry-go-round ball appears to have moved on a curve as a result of a force acting on it.

tion in the roller causes the fluid near the boundary to be pushed upward in the lower hemisphere and downward in the upper hemisphere. On the other hand, in the equatorial cross section the fluid motion is perfectly horizontal.

Although viscosity in the model is small, its effect on the balance between the pressure differences and the Coriolis force is enough to allow the existence of roller circulations. In the core, however, the viscosity is too small to have any such effect, although the magnetic fields might upset the force balance in a similar way. Viscous forces and magnetic fields also influence the diameter of the roller circulations. In the model the rollers have a diameter of about 10 percent of the sphere's radius for a fluid with the viscosity of water. If the viscosity is increased, the rollers become fatter. In the core, where the effect of viscosity is negligible, the diameter of the rollers would be an even smaller fraction of the core radius. The experiment does not indicate, however, whether or not the magnetic field would act as viscosity does in making the rollers fatter. In fact, the absence of magnetic fields is a serious limitation of this model.

Suppose now the rotating sphere was filled with a liquid metal such as mercury or sodium instead of water. Would the sphere then generate its own magnetic field? The answer is no. Any electric current that was fed into the fluid would die away in a fraction of a second, which is too short a time for dynamo action to take effect. On the other hand, electric currents in the earth's core can last for about 10,000 years without having to be regenerated, and that is a long enough time for the ponderous core-fluid motions to act as a dynamo. The lifetime of electric currents in a body is proportional to the square of the body's radius multiplied by its electrical conductivity. Because the conductivities of the model and the earth's core are roughly the same, it is the vast difference between the radius of the model and the radius of the core that explains why currents in the core can last 10^{17} times longer than currents in the model. For the model to have a chance of working it would have to be scaled up to the size of the core. Alternatively, the model would also work if it consisted of a fluid whose conductivity approached infinity. In that case the model could operate successfully at its present size.

With the failure of laboratory experiments to model the effects of magnetic fields geophysicists have turned their attention to theoretical considerations. The recent theoretical work of G. O.



NONROTATING SYSTEM is not affected by the Coriolis force. Fluid simply flows outward from regions of high pressure (*a*) and inward toward regions of low pressure (*b*). In a rotating system such as the earth's core the Coriolis force deflects the flow to the right in the upper hemisphere. When the Coriolis force (*colored arrows*) cancels the pressure forces (*black broken arrows*), the fluid circulates around regions of high pressure (*c*) and regions of low pressure (*d*). This phenomenon accounts for the roller circulations in model of earth's core.

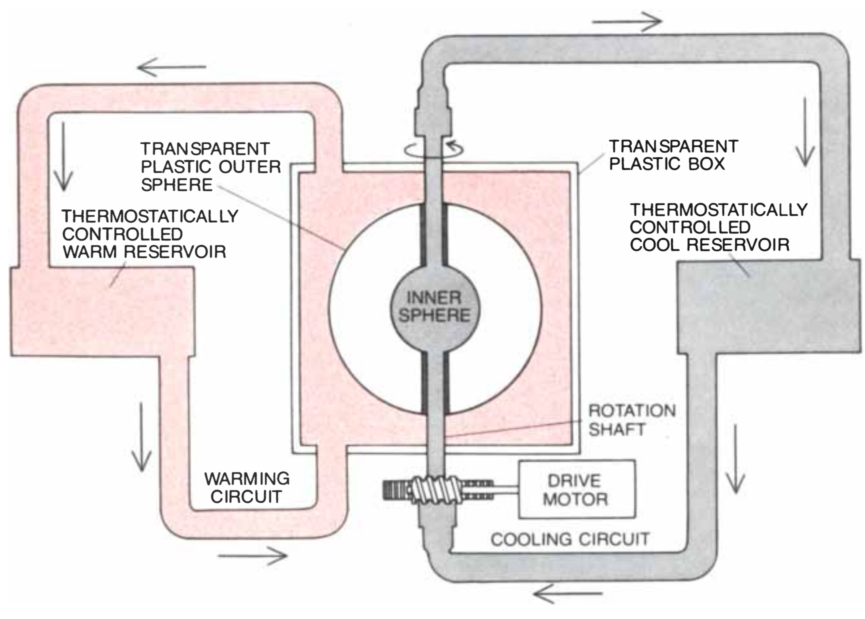
Roberts of the University of Cambridge and of other investigators shows that fluid motions can generate a magnetic field if they possess a net "helicity": the degree to which flow streamlines are twisted into either right- or left-handed coils. Net helicity arises whenever there are more coils in one direction than there are in the other. In the laboratory model helicity arises in the rollers because of the flow induced by the spherical boundaries. On the basis of the similarity between the fluid motions in Roberts' theoretical work and the fluid motions in the laboratory model, Busse demonstrated that convection patterns could generate a field that was mainly a dipolar one. Moreover, he proved that by virtue of the dynamo action of the fluid motions the magnetic field could increase substantially from a small initial value.

Imagine now that a laboratory experiment was set up in which a magnetic field grew in intensity from a microscopic size. Initially the conducting fluid in the experiment would not be affected by the magnetic field because initially the field would be too small. As the magnetic field grew, however, it would influence the fluid motions. It is a fundamental principle of electricity and magnetism that a magnetic field exerts a force on a conductor in opposition to the conductor's motion. If the earth's dynamo were a simple disk dynamo driven by a handle attached to a spindle in the center of the disk, the handle would become increasingly difficult to turn as the magnetic field put up increased resistance. The magnetic field would suppress the conductor's motion until an equilibrium was reached in which the field achieved a steady value. In a fluid conductor such as the earth's core, however, there is another degree of freedom: the direction of flow. The magnetic field could also alter the fluid's direction and so change the flow pattern. Such changes in direction could reduce the dynamo effect even though the speed of the fluid might not decrease.

Because a laboratory model in which the field grows from a microscopic size has not yet been constructed it is not known how a steadily increasing magnetic force would change the flow pattern. This is the subject of the continuing work of Busse and of P. H. Roberts and his co-workers at the University of Newcastle upon Tyne. Busse has calculated the changes in the fluid's speed and direction by supposing the magnetic forces are small, whereas P. H. Roberts' group is concentrating on what happens when those forces are large.

If the magnetic forces turn out to be large, the flow pattern may be drastically altered. In the hypothetical model the flow pattern of vertical rollers would probably become coarser. There would be fewer but larger rollers, and suddenly, when a critical field strength was reached, the pattern would change to one of large-scale motion, mainly in the horizontal direction. Little is known about this possibility, but the magnetic field would probably be large and toroidal. The field would lie in a ring within the spherical model. What actually happens in the earth's core depends on the equilibrium field strength and on many other factors that have not yet been studied.

The existence of a large toroidal field

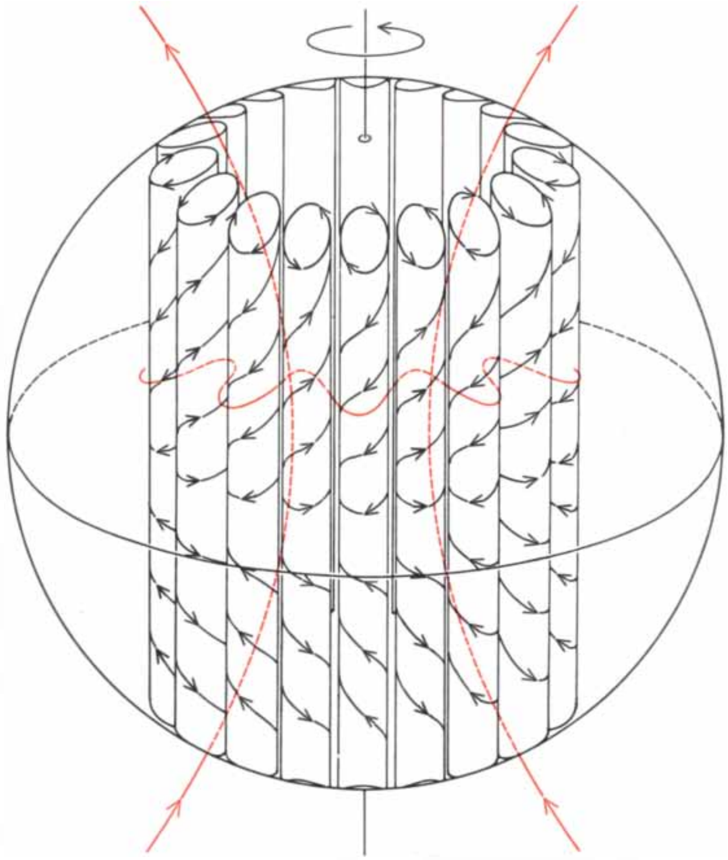


TEMPERATURE GRADIENT in the model of the earth's core shown in the photographs on page 119 is maintained by the heating system shown in this highly schematic diagram. In the model the temperature increases as the distance from the center increases. Such a temperature gradient, which is the opposite of the gradient in the core, is designed to compensate for differences between gravity in the core and centrifugal force that substitutes for gravity in the model.

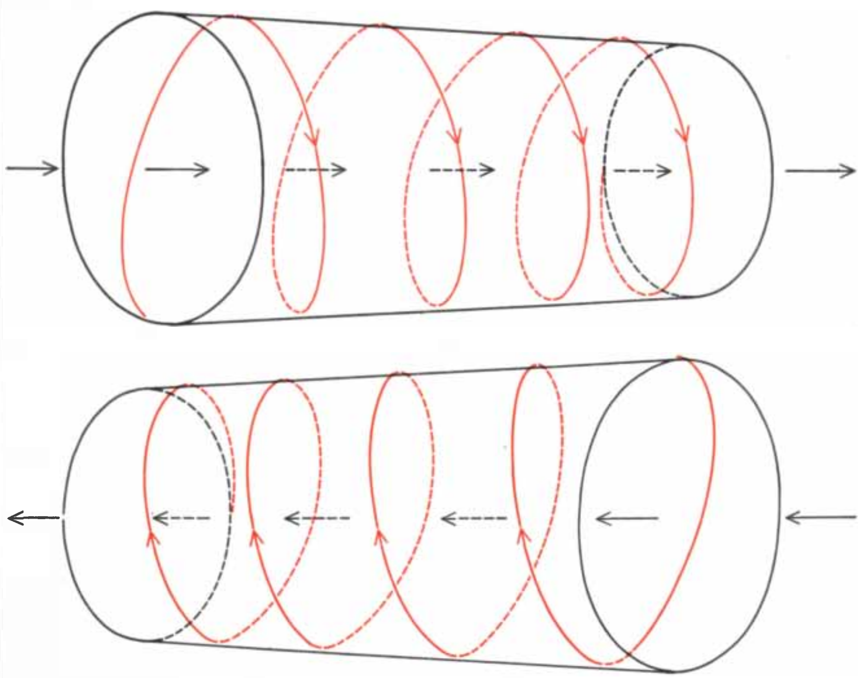
is the basis of an account of secular variation worked out by Raymond Hide of the United Kingdom Meteorological Office and S. I. Braginsky of Moscow M. V. Lomonosov State University. They maintain that under the influence of both rotation and a magnetic field the core fluid could support waves with periods on the order of thousands of years. Their account suggests that the westward drift of the magnetic field might be a wave phenomenon. Just as the advance of an ocean wave does not involve the forward motion of the water itself, so the westward drift of the magnetic field need not imply that the entire core is rotating with respect to the mantle if the movement is a wave motion. Hide and Braginsky's account applies only if the toroidal field is strong, and so this is one of the more compelling reasons for believing in the existence of a strong field.

Any successful account of geomagnetism must explain what caused the pole reversals at various times in history. Paleomagnetic work suggests that during a reversal the dipolar field first decays in intensity for roughly 10,000 years, then suddenly changes polarity and finally grows slowly in intensity again. The other possibility, that the dipole swings over while maintaining the same strength, is not supported by the paleomagnetic data. The mathematical equations that describe core dynamics remain the same when the sign of the magnetic field is changed, so that the dynamical state of the core is the same in both the normal and the reversed state. No one has developed an explanation of why the sign reversals take place. The apparently random reversals of the earth's dipolar field have remained inscrutable.

Disk dynamos might shed some light on the reversal phenomenon, because under certain circumstances they exhibit apparently random reversal behavior. The reversals arise because of coupling among various parts of the disk dynamo's electric circuit. Such coupling is similar to that among pendulums that are attached to a common support. For example, if three pendulums, spaced a short distance apart, are hung from a horizontal string and one of them is set swinging, then eventually the other two will start swinging. At certain times one of the pendulums may even stop for a short period as energy sloshes back and forth among the three of them. Although the behavior of the pendulums is deterministic (that is, future motions can be calculated from initial conditions and the laws of dynamics), onlookers see an apparently random exchange of kinetic energy among them. By the same token disk dynamos give rise to reversals that appear to be random but are actually deterministic. There are mathematical similarities between fluid dynamos and disk dynamos, and



CONVECTION CIRCULATIONS in the rotating spherical model have the appearance of slowly spinning rollers. Because the rollers end on the surface of a sphere, the fluid motions there are not parallel to the equatorial plane of the model. The boundary at each end of a roller is oppositely sloped, and so to an observer of the model outward circulation in a roller causes the fluid near the boundary to be pushed upward in the lower hemisphere and downward in the upper hemisphere. These motions are capable of generating the earth's dipolar magnetic field.



HELICITY is the degree to which fluid streamlines are twisted into helixes, or coils. In the model of the earth's core helicity arises in the rollers. The fluid in the upper roller traces out a right-handed helical path (*color*), which means that the fluid is circulating in the same sense as a right-handed corkscrew. The helicity in the lower roller is also right-handed even though the flow (*black*) is in the opposite direction. The role of helicity was discovered in the late 1960's.

so the apparently random reversals in the earth's magnetic field could be deterministic too. The formidable complexity of the mathematical equations that govern fluid dynamos has so far prevented physicists from searching seriously for reversing models.

A different but commonly entertained account of pole reversals suggests that the fluid dynamo "switches off" for a while, allowing the dipolar field to decay, and then "switches on" again, generating a field that happens to grow in the opposite direction. Although this account does not contradict the available information, it seems contrived. There is no reason to believe the core remains "on" for the million years between reversals but turns "off" for the few thousand years during which the poles reverse. It seems more likely the reversals are a natural feature of a dynamo that is constantly working to generate magnetic fields, as the disk dynamos are.

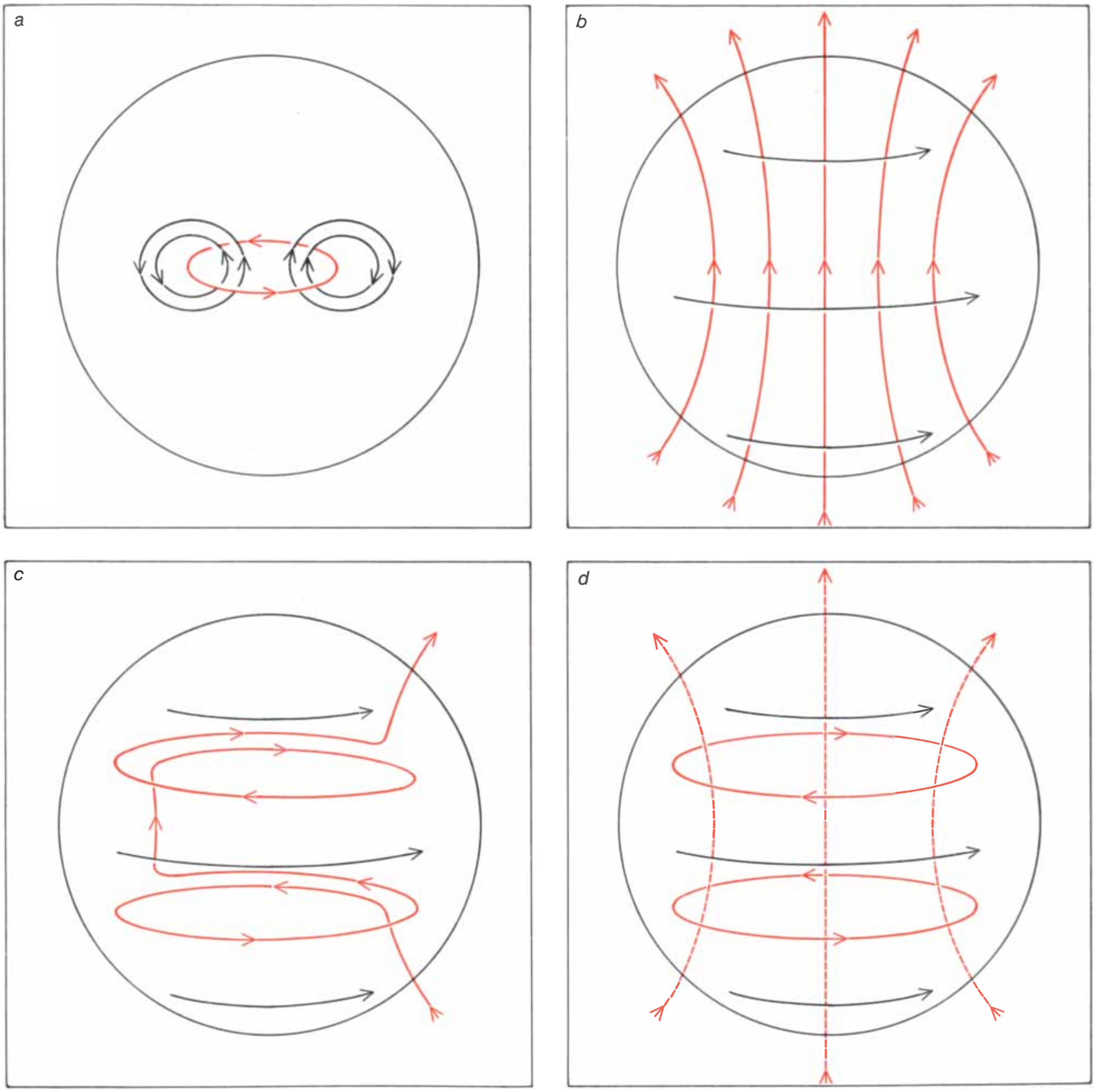
Over extremely long periods of time the character of the pole reversals has changed. For example, during the Cretaceous period, which lasted from 135 to 65 million years ago, there were no reversals for more than 20 million years. Such long-term behavior is presumably related to fundamental changes in the driving mechanism of the dynamo or in the shape of the boundary between the core and the mantle, but once again these possibilities have scarcely been investigated.

It is now time to take a closer look at possible energy supplies for core-fluid motions. An energy source is essential for any buoyancy forces because they do work against viscous and magnetic forces. In the laboratory experiment with the rotating sphere the energy comes in the form of the heat supplied to maintain the temperature gradients. In the earth's core the situation may not be quite so simple. There the energy must have been supplied at a more or less constant rate for several billion years. The energy may originate in a variety of forms (gravitational, chemical and heat), although it will ultimately be converted into the heat that flows out into the mantle. The energy supply cannot be so great as to melt the mantle or to drive more heat to the earth's surface than is actually observed. Such constraints turn out to be less trivial than they may at first appear to be.

Heat energy for thermal buoyancy could come from radioactivity if the core fluid contained a large enough quantity of radioactive elements. The main heat-producing elements are uranium, thorium and potassium. Theoretical accounts of the formation of the core suggest that uranium and thorium tend to migrate into the mantle and the crust, leaving behind only trace amounts in the core. Some workers have suggested, however, that the core contains an abundance of potassium, in which case it would also have a substantial amount of the radioactive isotope potassium 40. The suggestion that potassium is abundant in the core is not widely accepted because it rests on uncertain assumptions about the core's chemistry and composition.

A more promising idea is that the earth is cooling and hence liberating heat augmented by the latent heat of the core liquid as it freezes to form the solid inner core. The thermal capacity of the core is so great that a steady temperature drop of 100 degrees K. over the past three billion years would provide enough heat. For heat to drive the dynamo there must be convection currents throughout the mantle to carry away the heat. Cooling also causes the earth to shrink and hence to lose gravitational energy. If the contrast in density between the material of the inner core and that of the outer core is as large as seismologists think it is, gravitational energy will be released in considerable quantity. Such energy must ultimately leave the core as heat and be carried to the surface by convection in the mantle.

Experimental measurements place an



TOROIDAL MAGNETIC-FIELD LINES are probably present in the core of the earth. A ring (*colored line*), which is the simplest type of toroidal field, is shown in the top drawing at the left. The electric currents (*black lines*) that give rise to this field must point radially but cannot enter the earth's mantle because it is an insulator. The currents cannot travel through the mantle, so that any toroidal magnetic fields cannot be detected at the surface of the earth. The three other drawings illustrate how fluid motions acting on an initial dipolar field could generate large toroidal fields. At the top right a dipolar field (*colored lines*) is superposed on a rotating fluid (*black lines*) whose speed of rotation varies from point to point in the core. At the bottom left field lines are pulled out of shape by the nonuniform rotation of the fluid. The fluid pulls each dipolar field line into a toroidal shape. At the bottom right other effects on the motions of the fluid cause the field lines to form rings. This predominantly toroidal field points in opposite directions in the Northern and the Southern Hemisphere.

upper limit of about 4×10^{13} watts on the average heat flow through the surface of the earth. Most of this heat is known to have come from radioactivity in the crust, leaving only about 10^{13} watts to have originated in the core. If the core heat has come from the radioactive isotope potassium 40, the concentration of potassium in the core would have to be 800 parts per million. It is more likely that the heat has been supplied by the cooling of the core by 100 degrees K. over three billion years. In that case the heat would have come from the three comparable sources we have mentioned: the drop in temperature, the release of latent heat as liquid froze to form the solid inner core and the release as heat of some of the gravitational energy liberated by the shrinkage of the cooling core. The crucial question is whether or not 10^{13} watts is enough power to maintain the earth's magnetic field. This is not much power for an object as large as the earth's core. In fact, it is only about twice the total rate at which energy is currently being consumed in North America by human activities.

Engineers rely on thermodynamic considerations to assess the ideal energy conversion in electric-power stations without knowledge of the power station's turbines or generators, and a similar assessment can be made for the core without knowledge of certain aspects of its interior. In the case of the core it is estimated that between 10^9 and 10^{12} watts of power is lost to the currents that generate the magnetic field. Depending on how the core's dynamo is driven, an even larger quantity of heat will be needed to drive it. As with a power station that converts heat into electricity, the efficiency of the core's dynamo is the ratio of the electrical heating to the energy supplied. There is, however, one essential difference between the two systems: the electric currents in the core dissipate their heat in the core itself. Some of that heat could still drive convection currents, in the same way that a power station could use some of its output to heat its own boilers. On the other hand, some of the heat might hinder convection, in the same way that the efficiency of a power station would be impaired if its output were used to heat the cooling towers.

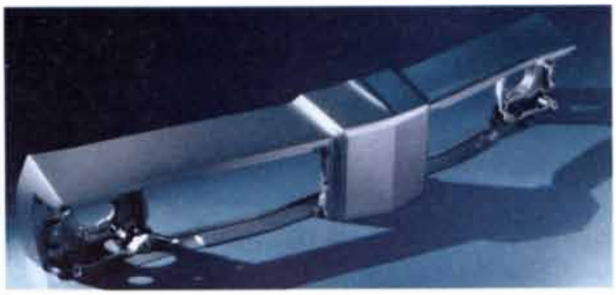
The ideal thermodynamic efficiency for a heat-driven dynamo is 20 percent, according to the calculations of George E. Backus at the University of California at San Diego. He found that the ideal efficiency was attained by a disk dynamo driven by an ideal heat engine operating between two fixed temperatures, provided the heat dissipated in the coil of the dynamo is fed back into the heat reservoir of the engine. Unlike such an engine, however, the earth's core is not ideal in any way and so the efficiency is less than 20 percent. In fact, the actual efficiency is probably about 5 percent. The reason is that the heat in the core can be dissipated not only by convection but also by conduction. When heat is transmitted by conduction, the fluid does not move and so conducted heat does not contribute to dynamo motion. An efficiency of 5 percent means that for every unit of heat generated to maintain the electric currents, 20 times as much heat must travel from the core into the mantle. As a result heat-driven dynamos could at best generate a small magnetic field, but certainly not a large toroidal field of the kind required by Hide and Braginsky's account of secular variation.

The poor performance of thermal dynamos led Braginsky and then one of us (Gubbins) to examine gravitational energy as a power source alternative to heat. Some of the gravitational energy that is lost by the contraction of the earth appears as compressional heating, but the amount of such heat is very small. Estimates of the difference in density between the solid material of the inner core and the liquid of the outer core have been made from seismological measurements by T. G. Masters of the University of Cambridge. Although these estimates are crude, they definitely do not allow a difference in density of less than 20 percent. This means that the solid inner core consists of a greater proportion of iron and nickel than the liquid of the outer core.

As the liquid freezes, iron will migrate into the solid inner core as light materials stay behind in the liquid outer core and provide buoyancy. Convection currents will stir the fluid and so uniformly redistribute the light material. The gravitational energy that is lost through this redistribution can be transformed into heat energy by electric heating, by viscous heating and by the molecular diffusion of the light material through the main body of the liquid. The properties of liquid iron are such as to make electrical heating the largest of these effects, because both the viscosity of iron and the diffusion rates of the light materials are low. The magnetic field can be regarded as a medium for converting gravitational energy into heat.

What is striking about the gravitationally powered dynamo is that under certain conditions it is almost 100 percent efficient. Most of its energy goes into generating the magnetic field. Such a model is much more satisfying than the heat-driven dynamo, where so much energy is wasted. A gravitational energy source could generate magnetic fields of hundreds of gauss without pushing too much heat into the mantle. It seems likely that gravitationally powered dynamos will play a significant role in future models of the core.

SMC, or sheet molding compound, is a stiff, lightweight plastic composite consisting of resin, powdered filler, and randomly oriented chopped fibers. With the continuing trend toward lighter automotive materials, SMC is becoming a promising replacement for steel.

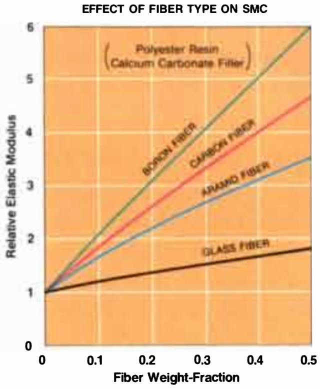


But can this plastic compete with metal? In light duty it can. In fact, thanks to redesigning, some SMC parts outrival their steel counterparts.

However, for structural application where stiffness is paramount, redesigning the part may not be enough. That's why engineers here at the General Motors Research Laboratories addressed the problem of "redesigning" the material, too.

The challenge was an imposing one: For a given use, find what ratio of which resins, fillers, and fibers provides the best mix for minimizing weight without adversely affecting stiffness and cost. And do it analytically. (Blending and testing would take forever.)

So our engineers modeled the resin-filler-fiber system. Then using micromechanics theories for calculating the stiffness of resin + filler and of resin + fiber, they devised a mathematical procedure that predicts the stiffness of any blend of all three components.



This unique analytical tool has two potent capabilities. It enables designers to determine an SMC composition that meets structural and cost goals — in other words, to tailor the material to the need. And it aids our engineers in exploring the cost-effectiveness of lighter, stiffer components such as glass microsphere fillers and carbon fibers.

Mathematical analysis: the best way of designing a material having endless combinations of ingredients.

We currently have openings for Ph.D.s in engineering or the physical, mathematical, or biomedical sciences. If interested, please send your resume to: GMR Personnel, Dept. 210. An Equal Opportunity Employer.

We've set the stage for SMC combos to perform better.

General Motors Research Laboratories
Warren, Michigan 48090

Physician, did you miss any of these significant developments in medical science?

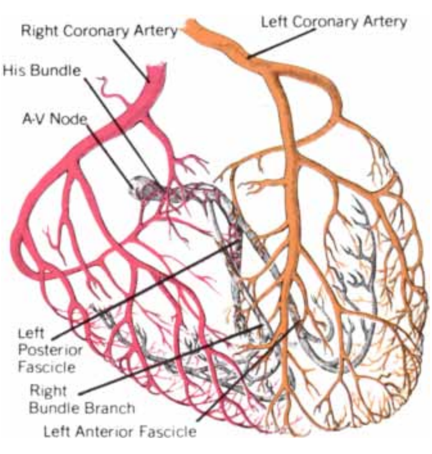
• The lidocaine level in the blood may fall below the therapeutic range when the drug is given as a single bolus followed by a continuous infusion. To avoid a "therapeutic hiatus," a second bolus should be given 20 to 30 minutes after the first bolus...

• Evidence is mounting that there is widespread over-diagnosis of pulmonary embolism at a time when heparin, in therapeutic doses, has become an important cause of drug-related death in the relatively well, hospitalized patient...

• The widely used street drug, phencyclidine, appears to cause persisting and possibly permanent brain dysfunction after repetitive use...

• There is alarming incidence of acute leukemia which may reach five percent in patients who appear to have been successfully treated with chemotherapy or radiation therapy for Hodgkin's disease...

Every day medical science evolves in ways that affect patient management and therefore patient well-being. But with more than 2,000 journals published, innumerable meetings held and hundreds of courses given, it is all too easy for developments that compel change in diagnosis and treatment to get by even the most conscientious reader and meeting-goer. And since such discoveries are today's news, they will not appear in the standard texts for years.



Branches of the right and left coronary arteries supply blood to the AV node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

We have a solution to this problem: It is called SCIENTIFIC AMERICAN *Medicine*. It makes such developments accessible to you when and where you need them, month after month.

Here is how SCIENTIFIC AMERICAN *Medicine* works. When you subscribe, you receive a two-volume, 2,000 page text, organized according to the familiar subspecialties so that you know where to find the answers to your questions.

The authors of SCIENTIFIC AMERICAN *Medicine* are leading authorities from the faculties of the Harvard Medical School and the Stanford University School of Medicine. These seasoned scholar-practitioners, close to the creation of new medical knowledge, write for you the practitioner, not the student. Therefore, at the time you receive your copy of SCIENTIFIC AMERICAN *Medicine*, you will hold in your hands the world's most advanced text of medicine. As long as you continue to subscribe it will stay that way.

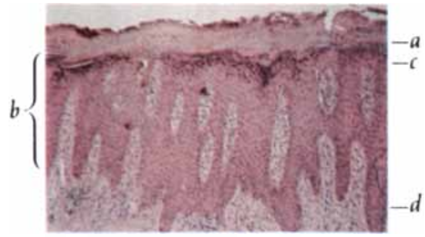
Each month you will receive several updated chapters. The new chapters which replace the old chapters contain reports of the significant advances in medical science that affect patient management. The index will also be updated each month. Furthermore, updated material will be summarized in a monthly bulletin accompanying the updated chapters so that you can quickly scan the information.

As the updated material is incorporated into the text, older more familiar material is deleted. Thus the amount of useful knowledge increases but the number of pages does not.

A Continuing Medical Education Program

For subscribers who wish it, SCIENTIFIC AMERICAN *Medicine* offers at no extra charge a continuing medical education program sponsored by the Stanford University School of Medicine's Office of Postgraduate Medical Education and administered by the Educational Testing Service of Princeton, New Jersey. The program consists of eight patient management problems per year, each written by a member of a Panel of Examiners drawn from leading clinical institutions. By completing all eight PMP s in a 12-month period the subscriber can earn 32 Category 1 credits per year, from the Stanford University School of Medicine, toward the AMA's Physician's Recognition Award.

This complete, concise medical knowledge service is yours for the drop of the accompanying card. Please fill it out and mail it to us. We think you will find that SCIENTIFIC AMERICAN *Medicine* is your answer to the challenge of bringing all that modern medicine has to offer to your patients.



Histopathologic changes in psoriasis.

Extensive psoriasis (top) cleared completely (bottom) after four weeks of treatment with anthralin.

THE DISTINGUISHED AUTHORS, AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN *MEDICINE*.

1. Cardiovascular Medicine

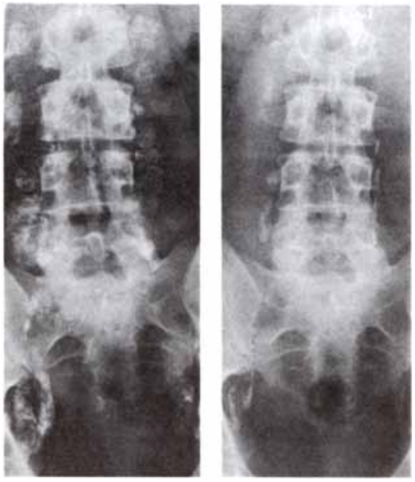
Edgar Haber, M.D., Professor of Medicine, Harvard Medical School, Physician and Chief of Cardiology, Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Professor of Medicine, Stanford University School of Medicine

Roman W. DeSanctis, M.D., Professor of Medicine, Harvard Medical School, and Physician and Director, Coronary Care Unit, Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Associate Professor of Medicine, Harvard Medical School, and Assistant Physician and Associate Director, Coronary Care Unit, Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Instructor in



Extensive retroperitoneal Hodgkin's disease before and after MOPP chemotherapy.

Medicine, Harvard Medical School, and Chief, Hypertension Unit, Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Professor and Chairman, Department of Dermatology, Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Professor of Medicine and Dean for Students and Alumni, Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Professor of Medicine and Head, Division of Gastroenterology, Stanford University School of Medicine

5. Hematology

Stanley L. Schrier, M.D., Professor of Medicine and Head, Division of Hematology, Stanford University School of Medicine

6. Immunology

John David, M.D., Professor of Medicine, Harvard Medical School, and Physician, Robert B. Brigham Hospital

7. Infectious Disease

Thomas C. Merigan, M.D., Professor of Medicine and Head, Division of Infectious Diseases, Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Professor of Medicine, Harvard Medical School, and Chief, Infectious Disease Unit, Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Assistant Professor of Medicine, Harvard Medical School, and Hospital Epidemiologist and Associate Physician, Massachusetts General Hospital.

Adolf W. Karchmer, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Associate Physician, Massachusetts General Hospital

Robert H. Rubin, M.D., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

8. Intensive And Emergency Care

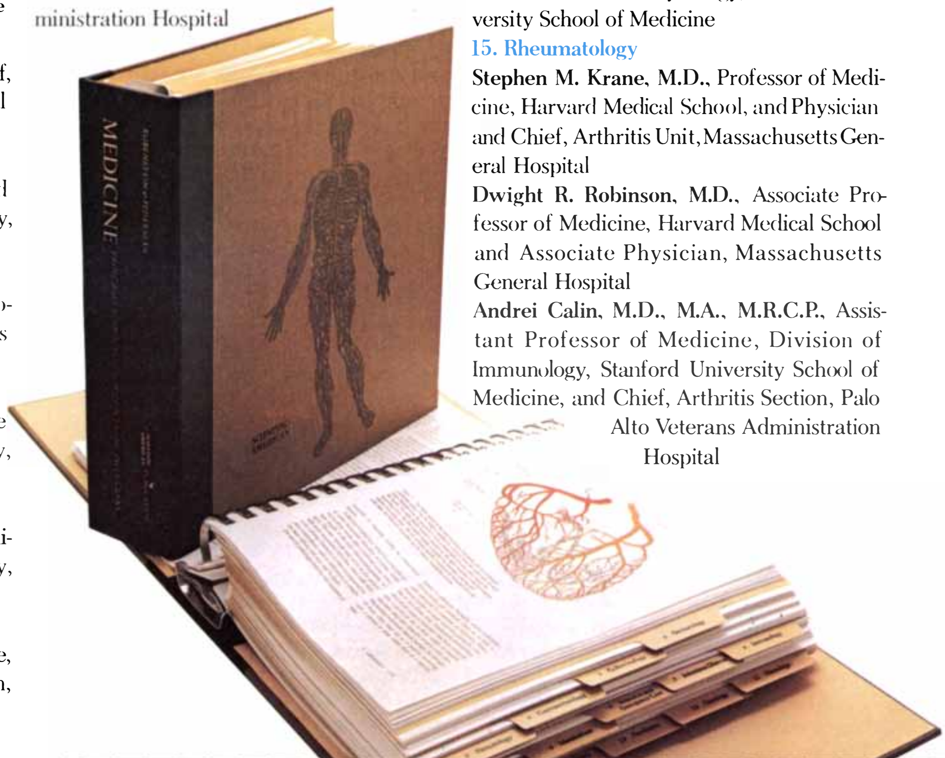
Edward Rubenstein, M.D., F.A.C.P., Associate Dean of Postgraduate Medical Education, and Professor of Clinical Medicine, Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Professor of Medicine, Harvard Medical School, Director of Research, Howard Hughes Medical Institute, and Physician, Peter Bent Brigham Hospital

10. Nephrology

Roy H. Maffly, M.D., Professor of Medicine, Stanford University School of Medicine, and Chief, Renal Service, Palo Alto Veterans Administration Hospital



11. Neurology

Robert W.P. Cutler, M.D., Professor of Neurology, Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Professor of Medicine and Radiology, and Chief, Division of Oncology, Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Associate Professor of Psychiatry, Harvard Medical School, and Chief, Psychiatric Consultation—Liaison Service, Massachusetts General Hospital

14. Respiratory Medicine

Eugene D. Robin, M.D., F.A.C.P., Professor of Medicine and Physiology, Stanford University School of Medicine

15. Rheumatology

Stephen M. Krane, M.D., Professor of Medicine, Harvard Medical School, and Physician and Chief, Arthritis Unit, Massachusetts General Hospital

Dwight R. Robinson, M.D., Associate Professor of Medicine, Harvard Medical School and Associate Physician, Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Assistant Professor of Medicine, Division of Immunology, Stanford University School of Medicine, and Chief, Arthritis Section, Palo Alto Veterans Administration Hospital

SCIENTIFIC AMERICAN MEDICINE

Post Office Box 5911, New York, N.Y. 10017

Gentlemen: Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement, and update it regularly by sending me new monthly subsections. As part of this service, I understand that the annual subscription of $150 for SCIENTIFIC AMERICAN *Medicine* is tax-deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a complete refund. SA2-79

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*.
☐ I shall also enroll in the CME Program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $150
☐ Please bill me

Name____________________

MD Specialty____________________

Address____________________

City__________ State__________ Zip__________

Genes That Violate Mendel's Rules

In sexual reproduction the parental genes are continually reshuffled and thereby exposed equally to the stringent test of natural selection. Some genes cheat, subverting the process to favor their own survival

by James F. Crow

Why is sexual reproduction so ubiquitous in the living world? An elaborate biparental method of procreation is certainly not particularly efficient. It requires two to do the work of one, and there are plenty of asexual reproductive processes, such as spore formation in some fungi, that are far more sparing of time and energy. If sheer numbers were the criterion of reproductive efficacy, sex would have been abandoned a billion years ago or would never have evolved at all.

The great role of sexual reproduction is an evolutionary one: it shuffles the genes of two parents to provide the genetic endowment of their progeny, and thus exposes to the stringent test of natural selection a maximum diversity of the characters and capacities represented in a species. If the system is to work well, if the testing is to be fair, the shuffle must be an honest one. It is not always honest, however. There are genes that cheat, perpetuating themselves in the population by tricking the reproductive process to work in their own favor.

The cheating is done at the stage of sexual reproduction called meiosis, the "reduction division" whereby one male germ cell divides to form four sperm cells, each of which has half the normal complement of chromosomes, and a comparable process takes place in egg formation in the female. Meiosis is followed by fertilization, which restores the original number of chromosomes, and the two processes together constitute the physical basis of Mendelian inheritance.

Chromosomes come in pairs and therefore so do the genes, which are segments of the long thread of DNA that is the core of a chromosome. In each individual's body and germ cells one member of each homologous chromosome pair has come from each parent. In meiosis the members of each pair segregate: one or the other of them, chosen at random, is transmitted to a sperm or egg cell and so to each offspring. The various pairs of chromosomes segregate independently, so that the genes on different chromosome pairs are thoroughly scrambled at each meiosis. Moreover, there is a regular process, called crossing-over, in which the two members of a pair line up side by side, break at corresponding points and exchange partners. As a consequence even genes on the same chromosome are not constrained to stay together but can participate in the meiotic shuffle.

The evolutionary advantage of this shuffling is that it enables genes from different individuals to come together and genes that have been together to separate. By thus scrambling the genes in every sexual generation Mendelian inheritance tests all the genes in many combinations. Genes that increase the capacity for survival, fertility and the survival of progeny are retained by the sieve of natural selection; less effective genes are lost. Evolution involves the continuous testing and retesting of gene combinations and the retention of those combinations that increase reproductive success, with the result that the species becomes better adapted.

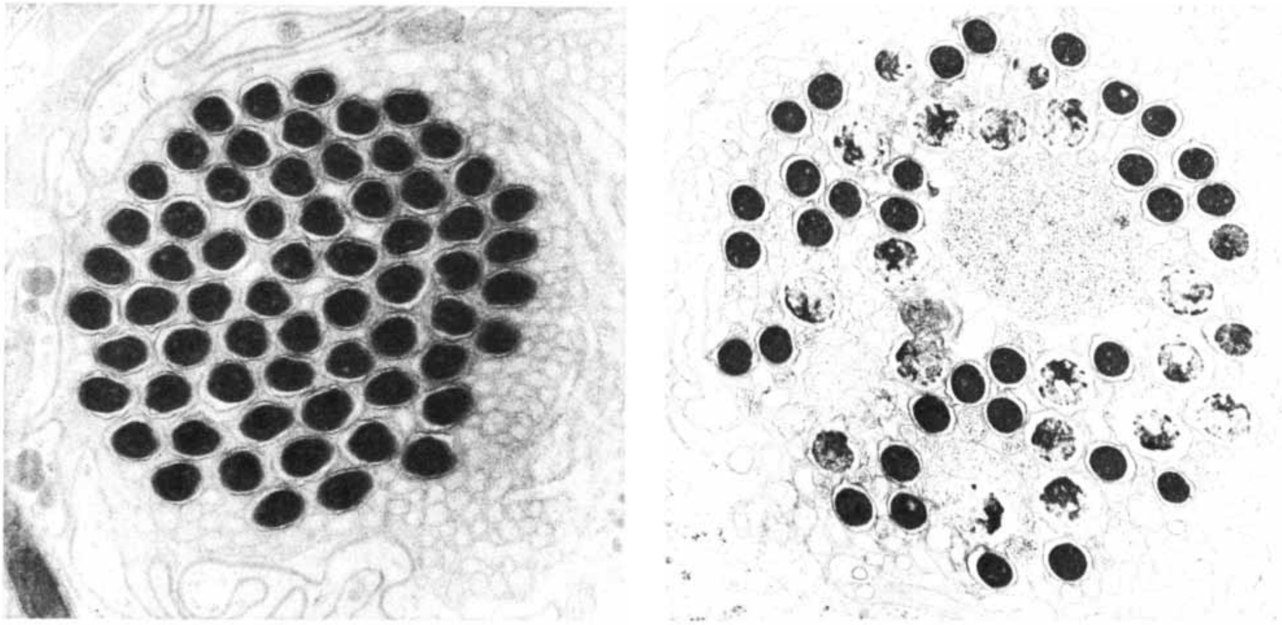
Meiosis ensures a scrupulously fair test for every gene combination by giving each gene the same chance as every other gene of being transmitted to the next generation. Genes that cheat in meiosis undermine the system generally by reducing its fairness. More directly, such genes usually affect processes other than meiosis and, like most new mutant genes, are almost always harmful; for example, one such gene found in mouse populations causes tail abnormalities in addition to meiotic disturbances. Other cheating genes have been discovered in corn, lily, tobacco, trillium, rye, mosquitoes, grasshoppers and the classical geneticist's favorite species: the fruit fly *Drosophila melanogaster*. This article will give an account of one particularly well-understood set of cheating genes, those that produce the trait *Segregation Distorter*, or *SD*, in drosophilas.

The story begins in 1956, when *SD* was discovered by Yuichiro Hiraizumi, who was then a graduate student of mine at the University of Wisconsin and is now at the University of Texas. Hiraizumi was investigating certain genes, on chromosome No. II of natural *Drosophila* populations, that affect variability. He mated males in which one member of each pair of these chromosomes was the "wild type" from nature and the other member came from a laboratory stock, and he traced the chromosomes' inheritance by noting the distribution of particular marker genes affecting eye pigment. The normal drosophila eye has two pigments, one cinnabar (a bright scarlet, like the mercury ore for which it is named) and the other brown. Together they give the eye its normal dark red color. In one mutant form no brown pigment is produced; the eye is therefore cinnabar and the mutation is called *cn*. In another mutation, *bw*, the cinnabar pigment is deleted and the eye is brown. When both mutations are present, as was the case in Hiraizumi's laboratory stock, the eye is without pigment and looks white. Each of the mutants is recessive, meaning that it has its effect only when the normal gene, which would mask it, is absent.

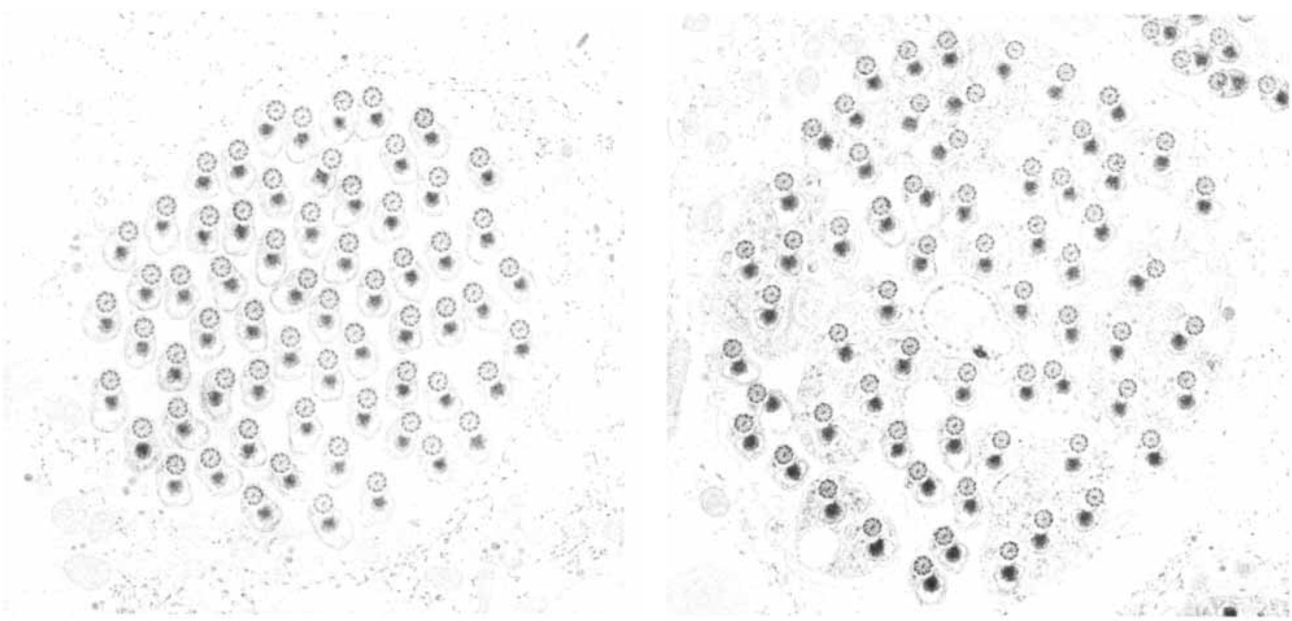
Hiraizumi mated hybrid red-eyed males (with one natural chromosome and one laboratory chromosome containing both mutations) to females that had two laboratory chromosomes and were therefore white-eyed. Because there is normally no crossing-over in male drosophilas each transmitted chromosome carried either both mutant eye-color genes or neither, so that according to the rules of Mendelian inheritance half of the progeny should have had white eyes and the other half should have had dark red eyes. Hiraizumi did observe the expected 50 : 50 distribution among the progeny of some 200 matings (except for some minor deviations caused by the variability-reducing mutant genes that were the original object of the study). Six of the matings, however, produced very strange results. Instead of half of the 100-odd progeny of each mating being red-eyed, almost all of them were. The red-eyed proportion ranged from 95 to 100 percent, and it was 99 percent or more in most cases.

What had caused six chromosomes, each descended from a different male in the wild population, to behave in this most unusual way? The possibility that there had been a selective death of white-eyed progeny early in embryonic development was quickly ruled out; far too many of the eggs that were laid developed into normal adult flies to allow such an explanation. Hiraizumi also established that whatever was skewing the eye-color ratio had its effect only in the course of sperm formation, not in egg formation; when he reversed the sexes in his mating, so that the females rather than the males carried the unusual chromosome, all the matings gave rise to the standard 50 : 50 ratio.

The wild flies that were the source of the unusual chromosomes had been collected in the fall from a clump of



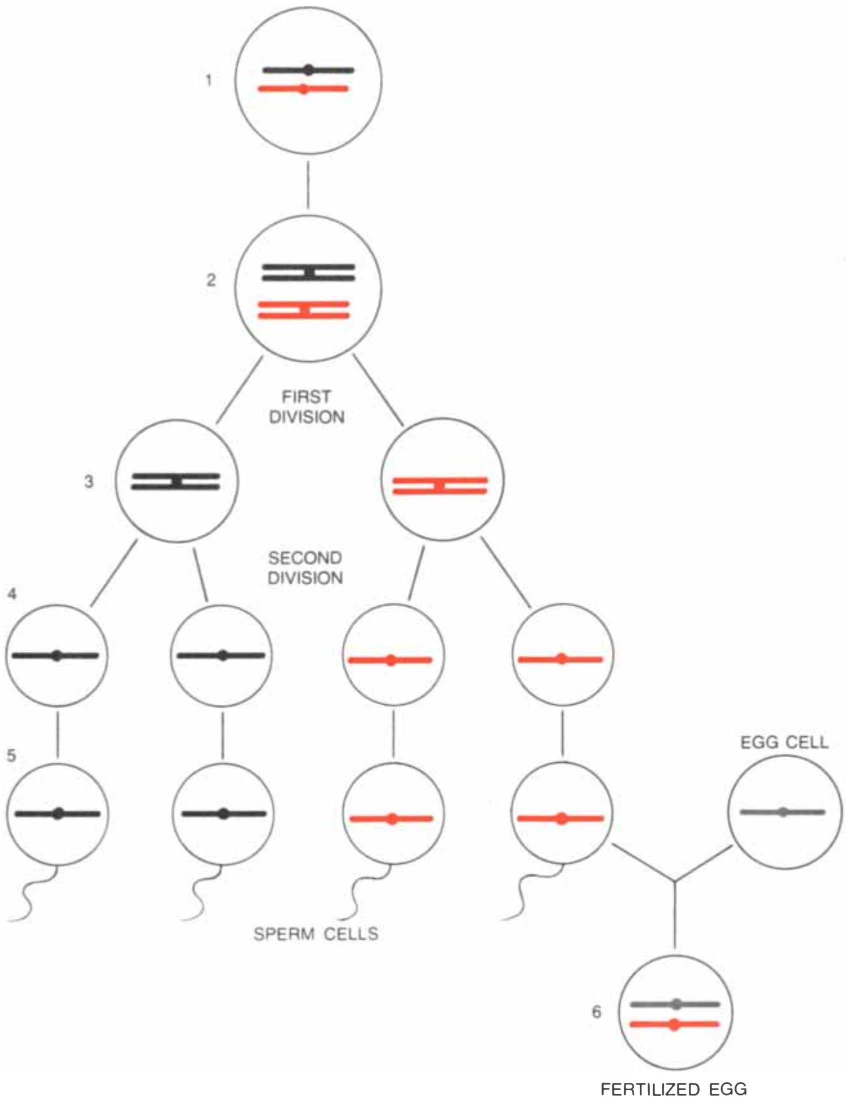
EFFECT OF SD CHROMOSOME, which subverts the process of meiosis in the fruit fly *Drosophila melanogaster*, is to prevent the proper development of half of the sperm heads. Normally the nuclei in a bundle of 64 immature sperm cells condense to form the heads of 64 sperms, as in the electron micrograph (*left*) made by Robert W. Hardy of the University of California at San Diego. In flies carrying an *SD* chromosome only 32 nuclei show the normal condensation, as in the micrograph (*right*) made by Kiyoteru T. Tokuyasu of San Diego; the others fail to develop normally. In both micrographs a section through the head region of sperm bundle is enlarged about 17,000 diameters.



SPERM TAILS also fail to develop properly from half of the immature sperm cells in a fly carrying the *SD* chromosome, as is shown in micrographs made by Tokuyasu. Sections through a developing tail region are enlarged 10,000 diameters. In a normal fly (*left*) a tail fiber (wheel-like structure) develops in each individualizing sperm cell; in *SD* fly (*right*) half of the sperm cells do not individualize normally. They are sperms that do not carry *SD* chromosome. It "cheats" by inducing its homologous chromosome to cause sperm dysfunction.

trees near the airport in Madison: the strange results appeared during the winter. We were eager to collect more flies from the same place the next summer, and we were dismayed to find that the trees had been cut down and no drosophilas were to be found. Fortunately flies from elsewhere in nature (including the bit of nature that is my backyard) turned out to have the same peculiar chromosome; indeed, in almost every natural *Drosophila* population that has been studied, from many parts of the world, from 1 to 5 percent of the No. II chromosomes carry the *SD* trait. These chromosomes have one other peculiarity that should be mentioned. In virtually every case they carry at least one inversion, and usually more than one. Inversions (regions in which the genes run in an order that is the reverse of the normal order) are not uncommon in *Drosophila*, but it is unusual to find two or more on one chromosome. There must be a reason for the association of the *SD* trait and multiple inversions.

If natural *Drosophila* populations display this strange chromosome, why was it not discovered long ago? The main reason, I think, is simply that there is ordinarily no visible effect; the only way the phenomenon can be observed is by finding unusual inheritance ratios when crosses are made involving chromosomes marked with conspicuous mutant genes. Perhaps it actually was observed from time to time by investigators who attributed the bizarre results to some kind of experimental error.



MEIOSIS, the "reduction division" whereby a germ cell forms four sperms, is diagrammed for a hypothetical male germ cell (*1*) carrying only a single pair of homologous chromosomes, one from the male parent (*color*) and one from the female (*black*). The chromosomes replicate (*2*) and then, in the first division, one replicated chromosome goes to each of the two new germ cells (*3*). In the second division each strand of each replicated chromosome goes to a different spermatid (*4*); each resulting sperm (*5*) has one representative of the original chromosome pair. In a real germ cell containing many chromosomes (four pairs in *Drosophila*, 23 pairs in human beings) each pair "segregates" independently; the sperm receives a random mixture of paternal and maternal chromosomes. A similar process takes place in egg formation. When sperm and egg fuse (*6*), the fertilized egg has a random combination of chromosomes from each parent.

Several years before Hiraizumi's experiment, however, Laurence M. Sandler and Edward Novitski of the Oak Ridge National Laboratory had suggested on the basis of certain experimental indications that the rules of meiosis might sometimes be violated; they called such a process "meiotic drive" and considered some of its theoretical possibilities. By a fortunate coincidence Sandler had come to Wisconsin as a postdoctoral fellow. To our delight Hiraizumi's discovery provided a perfect example of the phenomenon Sandler and Novitski had discussed theoretically. This was the beginning of a close collaboration between Hiraizumi and Sandler, who were responsible for the term *Segregation Distorter* and for most of the early understanding of the phenomenon.

By now the *SD* system has been studied in a number of laboratories in the U.S., Japan, Australia and Italy. As I have mentioned, there is no visible evidence of the *SD* chromosome on a fly that carries it; the only noticeable effect is a distortion of the ratio of progeny types. In the elaborate experiments required to analyze the system the chromosomes are therefore always marked with conspicuous mutant genes such as the eye-color genes I have described, so that the chromosomes can be followed through a complicated series of matings. I shall omit the details about how the chromosomes were labeled in various experiments and simply report the results.

One thing that makes it hard to understand how the *SD* chromosome gives rise to such a fundamental change in meiotic behavior is the fact that the chromosome must somehow inactivate precisely those sperm cells that do not contain it. How can that happen?

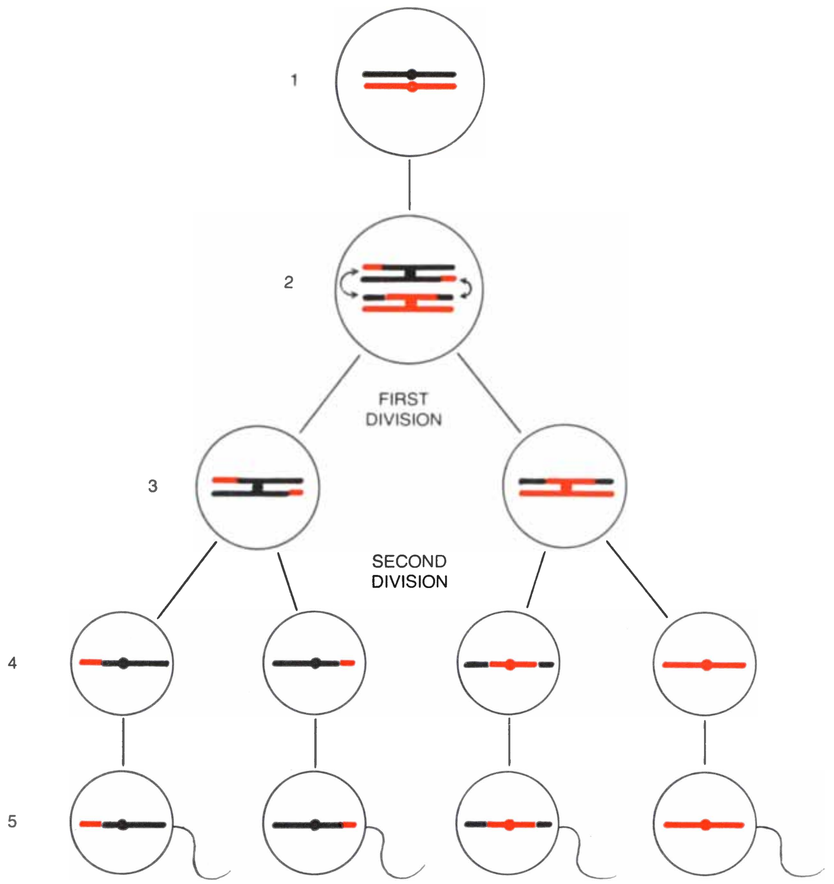
Sandler and Hiraizumi suggested one possibility quite early in the game: While the homologous chromosomes are still paired up during meiosis, the *SD* chromosome might do something to its normal partner (and rival) that later causes a dysfunction of the sperm receiving the normal chromosome. At first they suggested that *SD* might actually break the other chromosome. W. J. Peacock of the Commonwealth Scientific and Industrial Research Organization in Australia, working at the University of Oregon with a graduate student, John Erickson, was unable, however, to confirm the chromosome-breakage hypothesis. Microscopic observation showed that the chromosomes in *SD* heterozygotes (cells with one *SD* chromosome and one normal chromosome) went through meiosis unharmed, so that any disabling effect of *SD* would have to be subtler than outright breakage. Peacock and Erickson put forward a clever alternative idea. Some time earlier at Oak Ridge, Novitski and Iris Sandler had suggested on the basis of circumstantial evidence that only two of the four sperm cells produced in a single meiosis are functional—even in normal males. Peacock and Erickson thought the *SD* chromosomes might take advantage of this by somehow managing to get themselves included in the sperm cells that are destined to be functional.

It would seem to be a simple matter to distinguish between the two hypotheses, induced sperm dysfunction and preferential inclusion. If *SD* caused dysfunction of half of the sperm cells, then males with an *SD* chromosome should produce only half as many functional sperm cells as normal males, and the reduced sperm production might be reflected in reduced fertility. The trouble is that ordinarily the number of sperm cells produced by a male fly is much larger than the number required to fertilize all of a female's eggs; as a result the failure of even half of the sperm cells to function might not reduce fertility. Daniel L. Hartl, who was then a graduate student at Wisconsin, Hiraizumi and I found a way to get around the problem. By mating very young males (which produce fewer sperm cells than mature males) or by mating one male with many females over several days (so that the male's sperm supply became exhausted) we made sperm production the limiting factor determining the number of progeny. Quite independently and at almost exactly the same time Benedetto Nicoletti and Gianni Trippa of the University of Rome did similar experiments. The two groups reached the same conclusions. Males carrying an *SD* chromosome did indeed produce fewer progeny than normal males and not the same number, as would be the case if preferential inclusion were the mechanism. And the number of their progeny was reduced in just the proportion expected if the *SD* chromosome was causing a dysfunction of the sperm cells that received the normal, non-*SD* chromosome.

Clinching evidence for a dysfunction induced by *SD* came from electron-microscope studies of sperm maturation. Nicoletti reported that about half of the sperm cells in *SD* males had tails with an abnormal appearance. Then Peacock, Kiyoteru T. Tokuyasu and Robert W. Hardy of the University of California at San Diego confirmed Nicoletti's finding and revealed additional detail. They showed that whereas ordinarily each nucleus in a bundle of 64 immature sperm cells (derived by two meiotic divisions from a group of 16 germ cells) becomes small and dense to form a sperm head, only half of the nuclei condensed normally in flies carrying one *SD* chromosome and one normal chromosome. The effect was even more striking in the tail region of the sperm cells, which at first are held together in a single mass of protoplasm and then normally separate to form 64 individual tails. In *SD* males only half of the sperm tails were thus individualized; the rest of them remained massed in larger groups. The sperm cells that failed to develop properly were clearly those that did not contain the *SD* chromosome. This was established by running experiments in which the males' non-*SD* chromosome was one that is resistant to the effect of *SD*. In such males all 64 sperm cells in a bundle developed normally.

The most obvious deduction from these results would be that the *SD* chromosome somehow injures its non-*SD* partner and renders it unable to carry out its usual function. That, however, cannot be correct; the failure to develop cannot be simply the result of the failure to perform a normal function required for maturation because it has been known for a long time that a sperm cell's functioning does not depend on its chromosomal content. As long ago as 1927 H. J. Muller showed that a sperm cell can function normally even if many of its genes are missing. He produced drosophila strains in which some of the sperm cells lacked certain chromosomal material and some of the egg cells had a corresponding excess; when these sperm and egg cells were combined, "two wrongs made a right," as Muller put it, and a normal fly was hatched.

More recently Dan L. Lindsley of San Diego and Ellsworth H. Grell of Oak Ridge were able to produce sperm cells containing only the dotlike chromosome No. IV; such a sperm cell, when it was combined with an egg cell containing complementary extra chromosomes, gave rise to a normal fly. Since it was already known that the fourth chromosome is not needed for sperm-cell function, the experiment indicated that sperm function does not require any



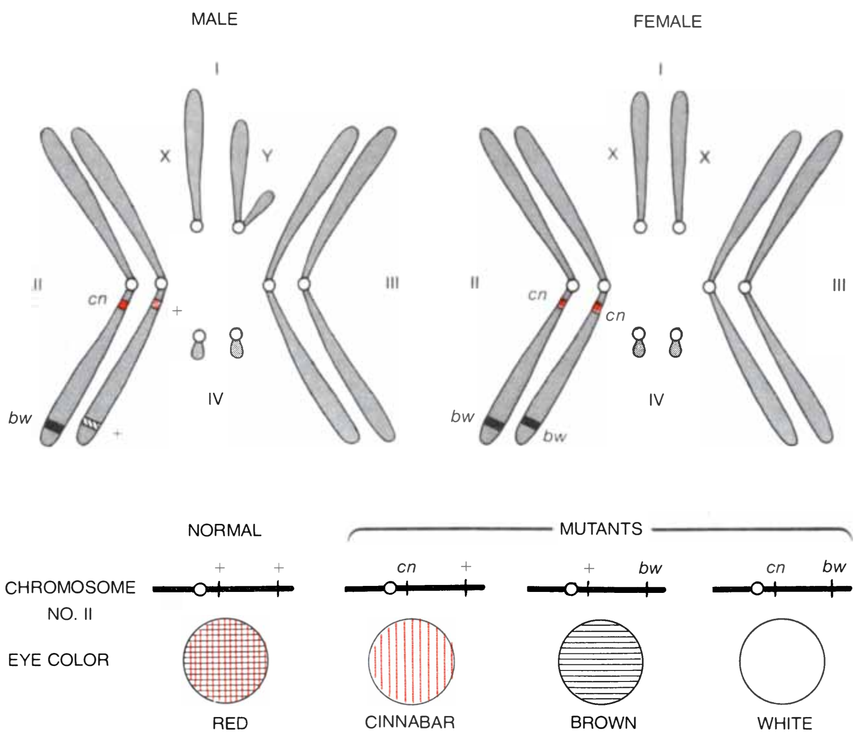
CROSSING-OVER multiplies the gene-shuffling process as is shown in this diagram of the same hypothetical meiosis as the one depicted in the illustration on the opposite page. In the course of replication the two members of a homologous pair may break at corresponding points. Homologous segments change places, so that some paternal genes (*color*) end up on maternal chromosome (*black*) and vice versa (*2*); individual genes become independent units in shuffle.

sperm chromosomes at all. The effect of the *SD* chromosome on its homologue cannot, then, be simply to inactivate some function, because no function is required. *SD* must somehow induce its partner to commit a positive act of sabotage.

Further evidence came from an analysis by Laurence Sandler and Adelaide Carpenter at the University of Washington. They reviewed the consequences of a rare error in sperm production by males with an *SD* chromosome and a normal one, as a result of which some sperm cells carried either both chromosomes or neither of them. The sperm cells with no chromosomes were functional but those with both *SD* and the normal chromosome were not, supporting the conclusion that *SD* does something to its homologue that causes the homologue in turn to produce sperm dysfunction. *SD* perpetuates itself by inducing its partner to destroy itself.

How this normal partner is instructed by *SD* to misbehave is not known, nor is the nature of the misbehavior. There is, however, one lead to the molecular details of what goes wrong in the aberrant sperm cell. During sperm-cell maturation there is normally a chemical change in the cell nucleus: lysine, one of the amino acids, is replaced by arginine. In *SD* flies this process fails, at least partly. The way is now open for a chemical study that could yield the details. At this stage one conclusion emerges clearly: The effect of *SD* is a complicated one requiring a number of active processes, not merely a failure of some normal sperm function.

Incidentally, it is a good thing (from the organism's point of view) that sperm-cell function does not require functional genes. If functional genes were required, it would be much easier for an *SD*-like system to get started, because then all the disruptive chromosome would have to do would be to injure its partner enough to prevent its functioning. As it is the *SD* chromosome must do more: it must cause its partner to become actively harmful to sperm-cell function. It is always easier to stop something than to start something new. Furthermore, if there were genes affecting sperm-cell function, there would be competition among sperm cells, and a gene that improved the ability to fertilize would increase in the population. If such a gene happened also to cause, say, malfunction of the liver, that would be just too bad; the gene would increase anyway, since selection for good health is much less effective than selection by competition among sperm cells. Whatever the evolutionary reason for the nonfunctioning of genes in sperm cells may be, the nonfunctioning makes it more difficult for harmful *SD*-like systems to arise and also prevents harmful, or at best irrelevant, competition among sperm cells carrying different genes.

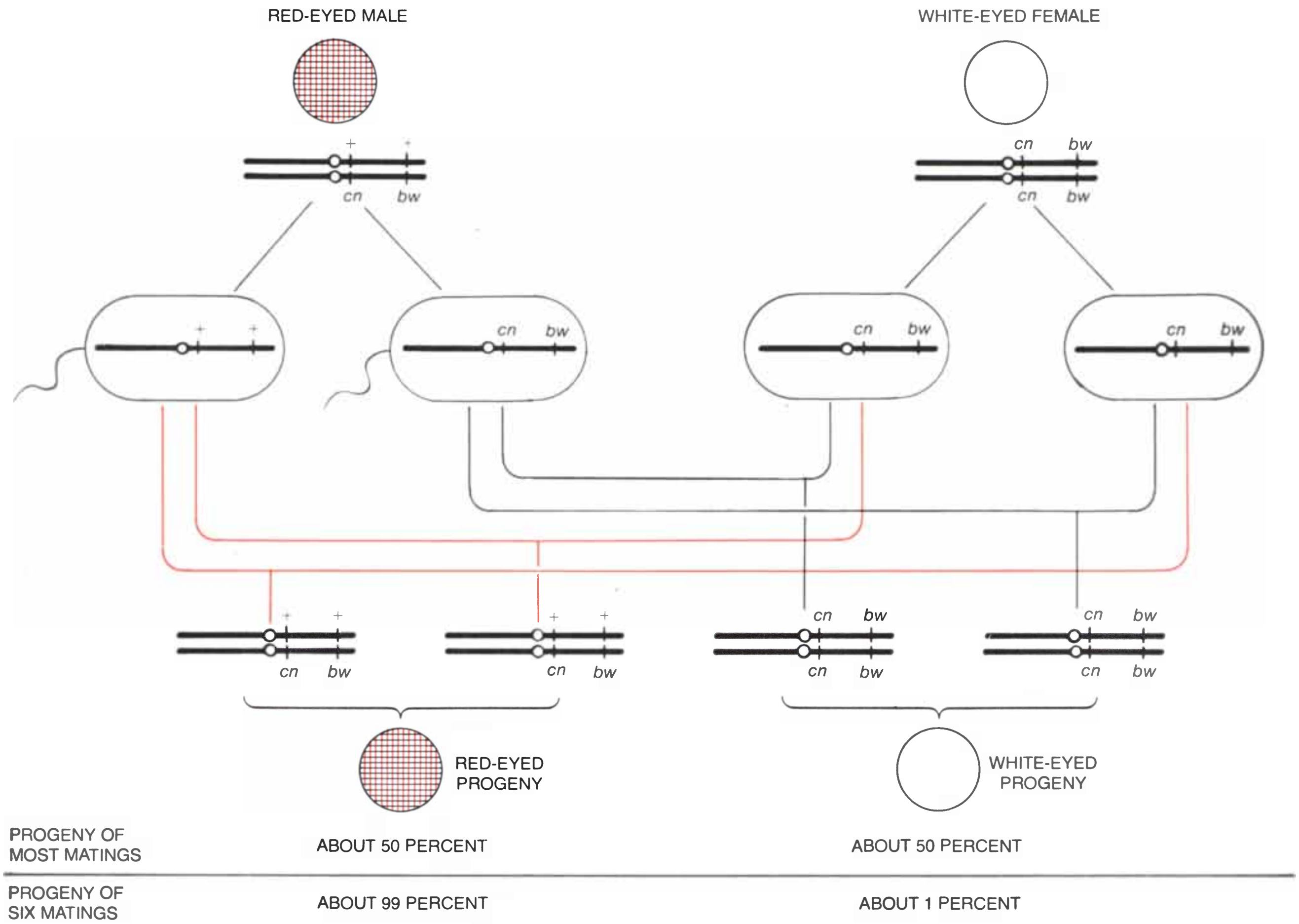


FRUIT FLY *D. melanogaster* has four pairs of chromosomes. Chromosome No. II is the site of eye-color mutations called *cn*, for *cinnabar* (*solid color*), and *bw*, for *brown* (*black segment*); a corresponding hatched segment indicates the corresponding normal (+) eye-color gene. The effects of the mutations are shown below. Two normal genes produce normal dark red eyes. The mutation *cinnabar* produces scarlet eyes, the mutation *brown* brown eyes; both mutations together produce white eyes. Female whose chromosomes are shown is white-eyed. Since normal genes are dominant they mask recessive mutant genes, and so male has red eyes.

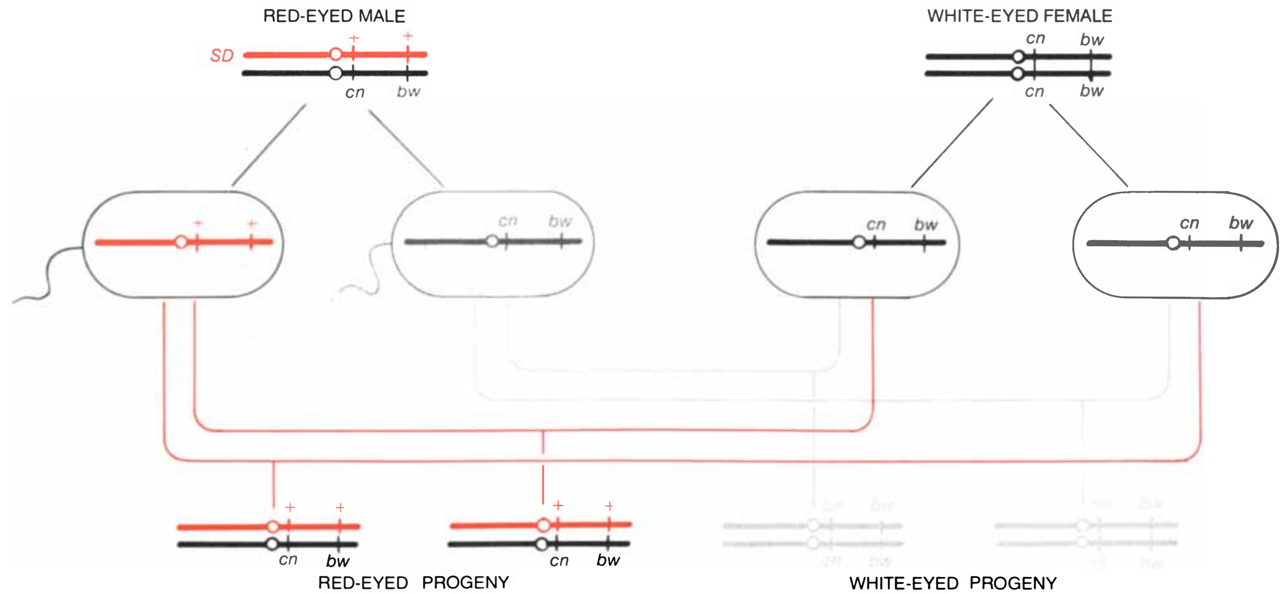
Genetic analysis of the *SD* chromosome to locate the relevant genes and learn how they interact has been difficult and time-consuming. As I have indicated, there is no way to recognize an *SD* fly by its appearance; *SD's* effect is on the ratios of the progeny, and distorted ratios can be recognized only when the chromosome is labeled with mutant marker genes. Moreover, experiments with various strains have sometimes produced inconsistent results. Finally, gene mapping by traditional crossing-over methods has been complicated by the inverted gene sequences I mentioned above, which tend to suppress crossing-over. Nevertheless, experiments in my own laboratory and in others have established that the main segregation-distortion effect is caused by two genes that are very close to each other but that straddle the centromere, the point where the fibers that pull sister chromosomes apart during cell division are attached. Here I shall designate the genes *S*, for *segregation distorter*, and *R*, for *responder*. With a superscript plus sign to indicate the corresponding normal genes, there are then four kinds of chromosome: *S*-*R*, *S*-R^+, S^+-*R* and S^+-R^+. The first is the *SD* chromosome, with its two components; the last is the normal chromosome.

The results obtained by mating males with certain combinations of these chromosomes are particularly revealing [*see illustration on page 141*]. The first row of the table shows once again the high degree of segregation distortion caused by the *SD* chromosome. The second and third rows show that neither component has an effect by itself; both *S* and *R* are required. The fourth row brings a surprise: when *R* is only on the partner chromosome, *S* is changed from a distorter gene to a suicide gene. The fifth row shows that the S^+-*R* chromosome is immune to the distorting effect of *S*-*R*. The *S* gene must direct the synthesis of some product that influences the R^+ gene on the homologous chromosome (or, in the suicide case, on the same chromosome) to prevent that chromosome's sperm cell from maturing properly. As I have emphasized, the effect must be a positive act on the part of the R^+ gene.

Barry S. Ganetzky, a graduate student of Laurence Sandler's at Washington, added significant details. He prepared chromosomes in which small pieces had been deleted by X rays. When the *S* gene was deleted, the distorting effect was lost; the chromosome behaved as though the normal gene were present. And so the *S* gene appears to be doing something that was not being done by any gene on a normal chromosome; the S^+ gene does nothing (or does not exist).



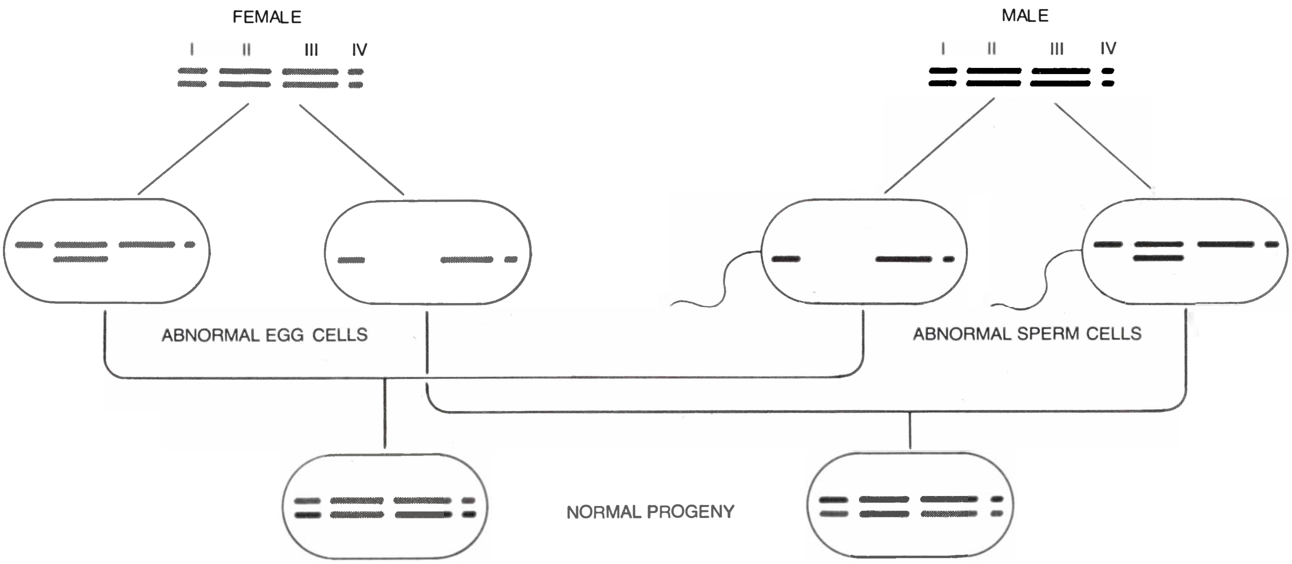
SD CHROMOSOME was discovered when red-eyed males having one No. II chromosome from a natural population and one from a white-eyed laboratory stock were mated to females having two laboratory chromosomes. The possible combinations of sperm cells and egg cells are shown. According to the rules of Mendelian inheritance, about half of the progeny should be red-eyed and half white-eyed. The progeny of each of some 200 matings did show the expected 50 : 50 ratio, but progeny of six matings were almost all red-eyed.



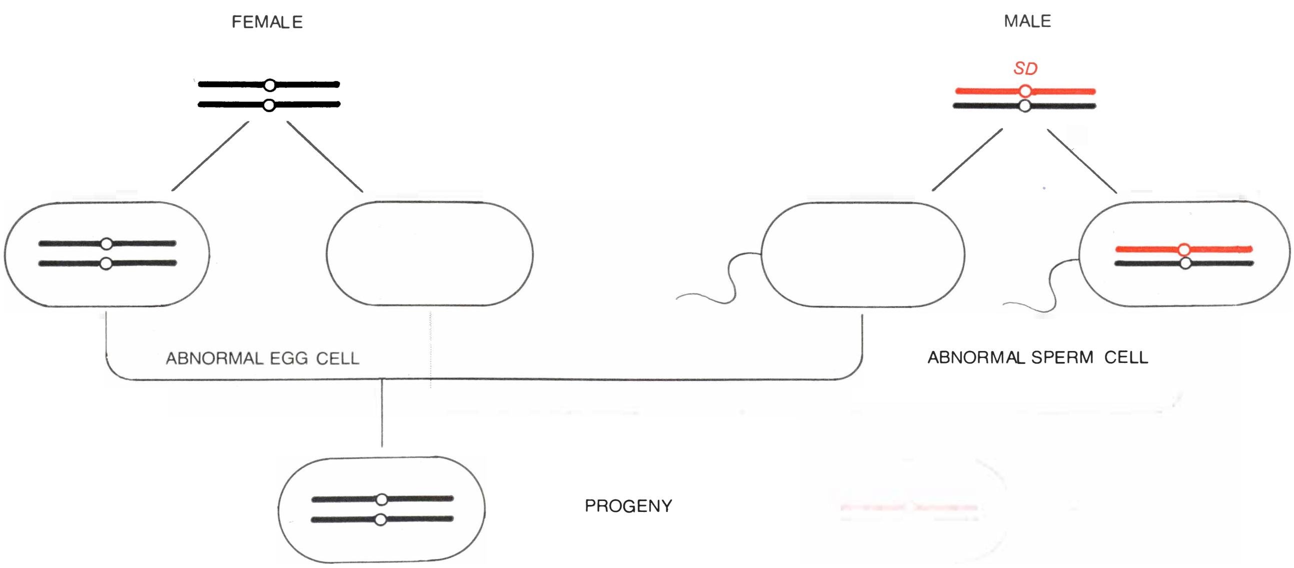
UNUSUAL RESULTS of the six matings are interpreted here. The eye-color ratio (almost 100 percent red-eyed) shows that the No. II chromosome from the natural population, bearing normal eye-color genes, is viable: it is combining with an egg's mutation-bearing chromosome and producing the expected red-eyed progeny. The homologous male chromosome, however, is apparently not combining successfully with a similar egg chromosome to produce white-eyed progeny. The chromosome from the natural population is distorting the results of segregation. It is a "segregation distorter" (*SD*) that favors its own survival by inactivating sperm cells containing its homologue.

When the R or R^+ gene was deleted, the chromosome behaved as if it carried an R gene; in other words, the R gene is not doing anything. To recapitulate: The S gene produces something that acts directly on the normal R^+ gene; when R^+ mutates to R or is deleted, it is no longer affected by S.

When does this effect of S on R^+ take place? The last time the homologous chromosomes are in the same nucleus and can interact easily is during early meiosis. That is eight or nine days before the time of sperm maturation, when the actual damage appears. Evidence supporting the time lag came from another experiment. If *SD* males are grown at a temperature of 19 degrees Celsius rather than the customary 25 degrees, the amount of distortion is substantially reduced. This temperature sensitivity provided an experimental handle for Elaine Mange, who was then a graduate student at Wisconsin. She lowered the temperature at which some flies were maintained from 25 degrees to 19 for just a brief period and mated the males at specified times after the temperature reduction. It turned out that the progeny of males mated between eight and nine days after the low-temperature pulse displayed less distorted ratios. Since it takes about that time for a cell to progress from early meiosis to the mature-sperm stage, this showed that the temperature-sensitive period is during early meiosis. These experiments were repeated carefully and extensively by Yukiko K. Hihara of Tokyo Metropolitan University, who got the same results. It appears, then, that the S gene communicates with or somehow influences the



TWO WRONGS MAKE A RIGHT when abnormal egg (or sperm) cells containing excess chromosomal material happen to combine with sperm (or egg) cells having a corresponding deficiency. Here an egg has two No. II chromosomes and a sperm lacks that chromosome, or vice versa, and yet progeny are normal. If sperm cells can function without a chromosome, *SD* cannot merely inactivate its homologue.



EITHER COMBINATION of an abnormal egg and sperm cell containing a deficiency and an excess in the No. II chromosome should produce progeny, according to the results shown in the upper illustration on this page. That is not the case when the two chromosomes in the male are the *SD* chromosome and a normal No. II chromosome. Although sperm cells with neither the *SD* nor the normal chromosome produce progeny, those with both the *SD* and the normal chromosome do not. Apparently, then, *SD* has some positive effect that causes its homologous chromosome to produce sperm-cell dysfunction, whether or not the *SD* chromosome is present in same sperm.

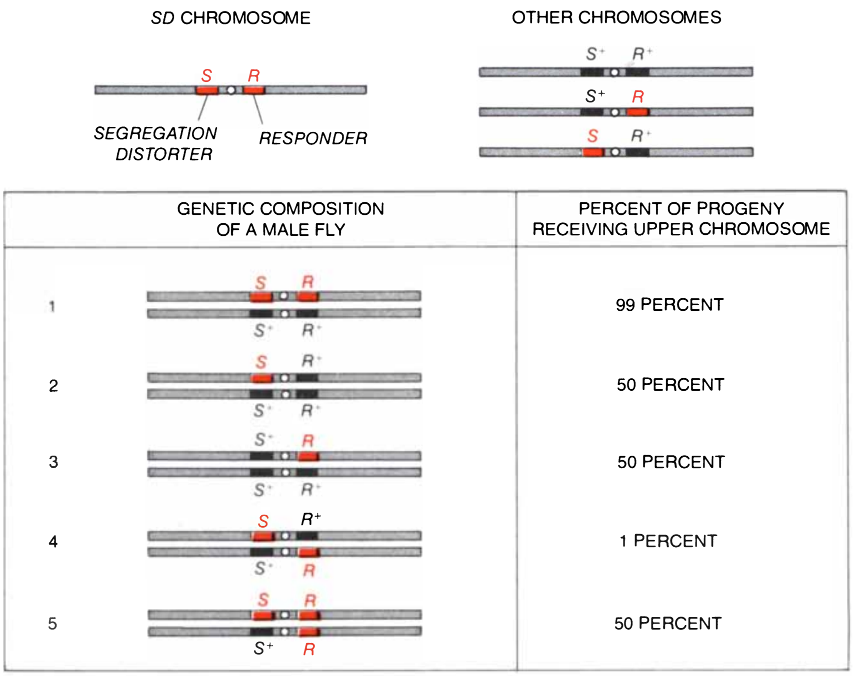
R^+ gene on the homologous chromosome while the two partners are paired early in meiosis.

Ganetzky also discovered a third gene close to the centromere, designated *En* for *enhancer*, that intensifies the distorting effect. The full effect (a progeny ratio of 99 percent or more) of an *SD* chromosome requires the presence of all three genes. A multiple-component system such as this must have evolved in stages over a long time. It could have a disastrous effect on the population, but it does not seem to. We have been able to learn quite a lot about its effect on a population and can make some plausible conjectures as to its evolution.

Because an *SD* chromosome is transmitted to almost all the progeny rather than to just half of them it ought to spread through a population like wildfire and quickly replace its homologue. It is clear enough why the *SD* chromosome does not in fact reach 100 percent. When a male is homozygous for *SD* (has two *SD* chromosomes), it is nearly sterile; in some strains such flies do not even survive. In a fly population the tendency of the *SD* heterozygote to increase through segregation distortion is countered by the tendency of the *SD* homozygote to be sterile or to die. One can easily compute what the frequency of *SD* chromosomes should be in such a system. The calculations predict a high frequency: more than 50 percent. And yet we actually find an *SD* frequency of less than 5 percent in natural populations. Something is holding the frequency down.

One brake on *SD* is provided by the presence of various "modifying genes" that lower the degree of distortion; such genes are scattered along all four drosophila chromosomes and are found in almost all natural populations of the flies. The most important brake, however, was revealed by Hartl's recent finding that about half of the "normal" chromosomes in one natural population are not S^+-R^+; they are S^+-R, and such chromosomes are immune to distortion by *SD*. There has not yet been any extensive analysis of other populations, but probably they too have a high proportion of S^+-R chromosomes.

I think we can reconstruct what happened. Somehow, a long time ago, the *S*-*R* chromosome arose. At first it increased rapidly. The increase was offset, however, by an increase of normal chromosomes that are resistant, that is, of chromosomes with the composition S^+-R. A three-way competition ensued. The *S*-*R* chromosome's tendency to increase was held in check by S^+-R, which nullified the former's meiotic advantage. And of course the S^+-R chromosome had an advantage over the other normal one, S^+-R^+, which is hardly transmitted at all when it comes up against *S*-*R*.



TWO MUTANT GENES, designated *segregation distorter* (*S*) and *responder* (*R*), are primarily responsible for the *SD* chromosome's effect. The four possible combinations of mutant (*color*) and normal (+) genes are shown at the top. The table gives the results of some matings involving chromosomes carrying these combinations. Genes *S* and *R* together exert the full *SD* effect (*1*), but neither gene has an effect alone (*2, 3*). When *R* is only on the homologous chromosome, *S* becomes a suicide gene (*4*). The S^+-*R* chromosome is immune to the *SD* effect (*5*). Apparently the *S* gene exerts an effect on the R^+ gene such that the R^+ gene causes sperm dysfunction.

There have been a number of mathematical analyses of this system and similar ones, most recently by Hartl and Brian Charlesworth of the University of Sussex. The results of one of my own computer runs are presented in the diagram on the next page. There are three chromosome types of interest: *S*-*R*, S^+-R^+ and S^+-R. (The fourth, *S*-R^+, is rare, and I have ignored it. It has no competitive advantage; in fact, it commits suicide when it is combined with S^+-R.) The frequency of each of the three chromosome types at every 10th generation is plotted in triangular coordinates. The proportion of *S*-*R* chromosomes is indicated by the distance from the base of an equilateral triangle whose altitude is 1, the proportion of S^+-R^+ chromosomes by the perpendicular distance from the right side of the triangle and the proportion of S^+-R chromosomes by the perpendicular distance from the left side.

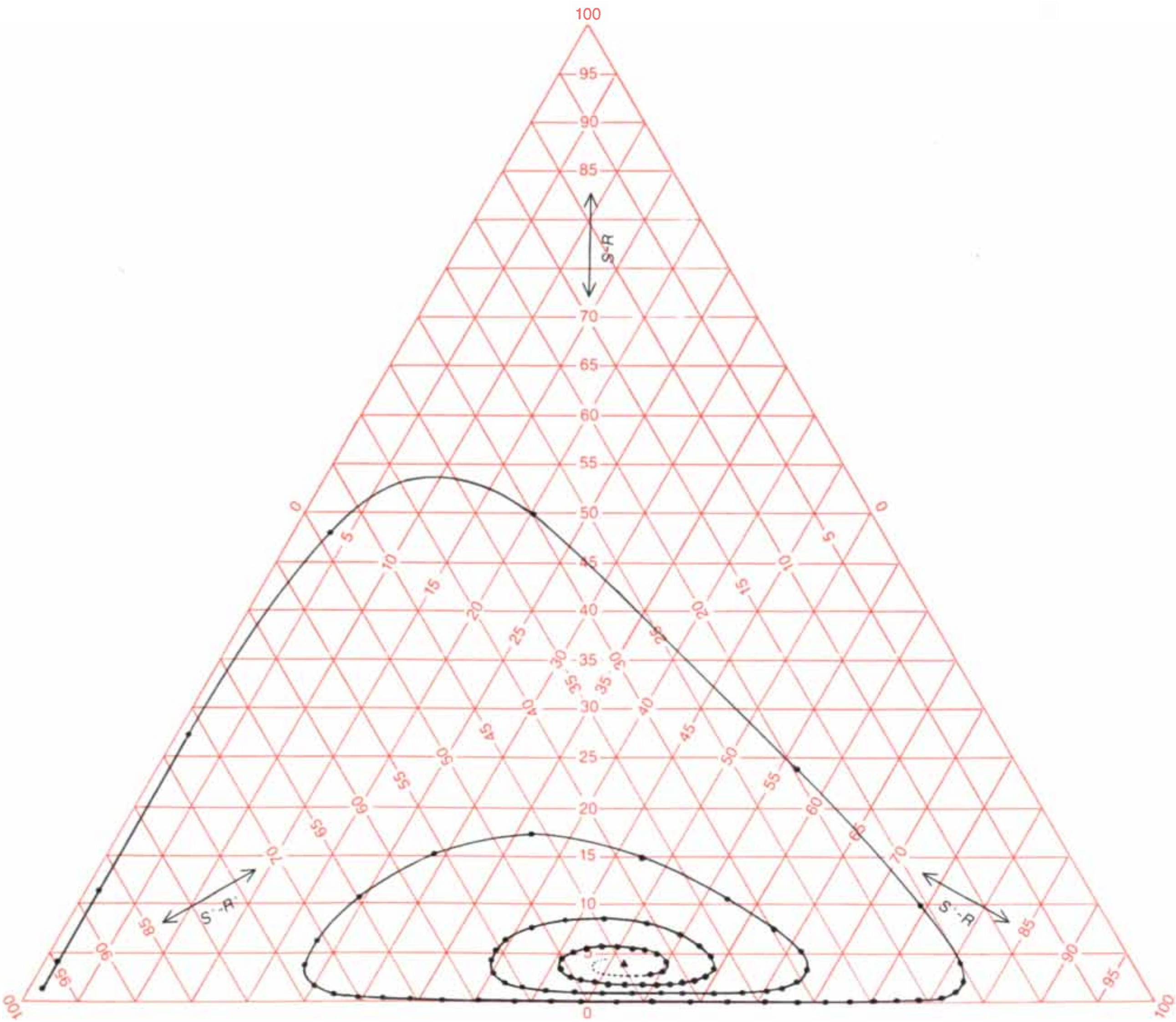
The original population must have been S^+-R^+. Then the *S*-*R* chromosome arose, presumably in two steps. Some S^+-R (and *S*-R^+) chromosomes must also have been present in low frequencies; they could arise by crossing-over between S^+-R^+ and *S*-*R* chromosomes. I assumed that the two rare types *S*-*R* and S^+-R began with a frequency of 1 percent each, as indicated by the starting point near the bottom left-hand corner. What I then did was to take plausible values for the degree of distortion and the relative viability and fertility of the various chromosome combinations. These values are not known exactly, since laboratory data may not reflect the true values in nature; the diagram reflects one set of plausible values that leads eventually to a point of equilibrium very close to what is actually found in a natural population.

At first the *S*-*R* chromosome increases rapidly; after about 45 generations more than half of the chromosomes are of this type. Because of its resistance to distortion caused by *S*-*R*, the S^+-R chromosome then starts to increase at the expense of both *S*-*R* and S^+-R^+, and the trajectory moves toward the bottom right-hand corner of the diagram. When the *S*-*R* type becomes rare, there is no longer any advantage for S^+-R, and because it has a somewhat reduced fertility it decreases in frequency and the trajectory moves to the left across the bottom of the diagram. When S^+-R becomes rare, *S*-*R* again increases because of its meiotic advantage over S^+-R^+. After about 325 generations one cycle has been completed and the process is repeated through a loop with a smaller amplitude. The cycle continues indefinitely, spiraling toward the center; only about the first 1,000 generations are plotted on the diagram. Eventually the population comes to equilibrium at the point indicated by the triangle at the center of the spiral, with about 4 percent of the chromosomes *S-R* and the remainder roughly half S^+-R^+ and half S^+-*R*. The diagram must represent a rough history of what went on among the three chromosomes competing for meiotic advantage. Exact calculations are complicated by the many modifying genes on other chromosomes, which shift the balance one way or another.

In the course of this three-way tug-of-war the *S-R* chromosome "tries" to hold on to any modifying genes that enhance its effect. *En* is one such gene and there are many minor modifying genes on the chromosome. Now the function of the multiple inversions on the *SD* chromosome becomes clear. By preventing crossing-over they keep the enhancing modifiers locked to the *S-R* complex; without the inversions they would become separated and soon would be found just as often on the normal chromosomes as on the *SD* chromosome.

The result of this complex interaction is that the *SD* chromosome seems to be effectively held in check by the presence of the S^+-*R* chromosome and the modifying genes, so that *SD*'s frequency in a population is not great enough to do much harm. The *SD* system and the other meiotic-drive systems found in nature are those to which the population has become adjusted in this way and similar ones. Populations in which such an adjustment was not worked out may simply have become extinct.

What would happen if a system such as *SD* were to arise on one of the chromosomes determining sex? For example, a normal male produces an equal number of sperm cells bearing X chromosomes and Y chromosomes, so that when the sperm cells fertilize X-bearing egg cells, XX females and XY males are produced in equal numbers. If an *SD* complex were to be located on the Y chromosome, nearly 100 percent of the progeny would be expected to be males. No such Y chromosome is known in *Drosophila*, but as a graduate student at Wisconsin, Terrence W. Lyttle, now at the University of Hawaii, contrived a way to produce one. By means of radiation he promoted an exchange of chro-



INTERPLAY OF THREE CHROMOSOMES in a fly population is traced by the spiraling curve, which gives the results of one of the author's computer runs. The proportion in the population of *S-R*, S^+-R^+ and S^+-*R* chromosomes is plotted, at every 10th generation (*dots*), in triangular coordinates; each chromosome's frequency in the population is measured along a different altitude of the equilateral triangle. At the beginning (*bottom left*) the population is 98 percent S^+-R^+. The distorting effect of *S-R* is countered by the resistance of S^+-*R*, which in turn has reduced fertility compared with S^+-R^+. Calculations based on plausible values for distortion, fertility and viability indicate that the population goes through a succession of cycles—the loops on the graph—that eventually bring it to a point of equilibrium (*black triangle*), at which some 4 percent of the chromosomes are *S-R* and remainder of them are roughly half S^+-R^+ and half S^+-*R*.

mosomal segments that tied the *SD* trait to the Y chromosome. Distortion by *SD* causes males carrying this translocation to produce an excess of sperm cells containing the Y chromosome, so that almost all their progeny are males. A population afflicted by such a system should become extinct within a few generations for lack of females.

Lyttle demonstrated the effect with an artificial population in the laboratory. The drosophilas were maintained for several generations in small cages, each one holding a few thousand flies. As expected, the population soon contained no females, and it died out. Then Lyttle did an experiment in which he started with a normal population and introduced a few males carrying the translocation. The distorting chromosomes from these males gradually replaced the normal Y chromosomes, and when they became prevalent enough, the proportion of females began to decrease. After a few generations this population too was extinct.

Clearly a meiotic-drive system, which is bad enough on one of the nonsex chromosomes, can be a disaster if it is on the Y chromosome. W. D. Hamilton of Michigan State University has suggested that one reason the Y chromosome is largely devoid of genes in many species is that the lack of genes prevents meiotic-drive mutants from arising and wiping out the population.

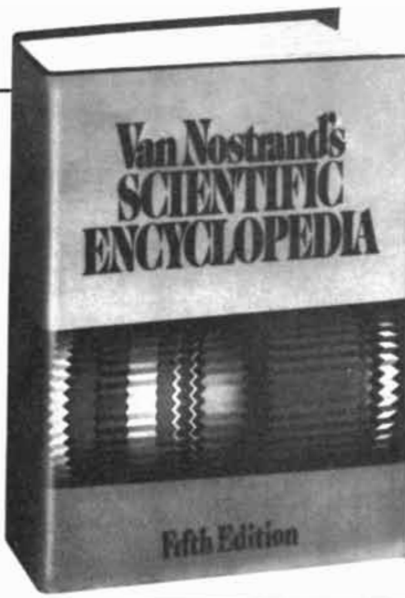
A distorting Y chromosome would seem to provide an ideal biological control of an insect population. An example of a meiotic-drive gene has already been discovered, by George B. Craig of the University of Notre Dame, near the male-determining gene of *Aëdes aegypti,* the yellow-fever mosquito. Unfortunately for any control application, there are so many modifying genes in the *A. aegypti* population that the distorting gene is rather ineffective outside the laboratory. If meiotic-drive systems are to be adapted for practical insect control, new mutant distorting genes will presumably have to be found in the laboratory for which there is no reservoir of distortion-reducing modifiers already in the population. It will probably be difficult to develop truly new mutants that have not arisen at some time in nature. It may nonetheless be possible, and in that case meiotic-drive genes could become a major new nonchemical technique in the continuing battle with insect pests.

Among other examples of cheating genes the best-known is the one that causes a variety of tail abnormalities in various populations of house mice. These mutants are often highly damaging or even lethal in homozygous mice, and it is clear that they are much too prevalent for the good of the population; they should be eliminated by natural selection. Instead they are main-

As your introduction to

The Library of Science

Choose either



this $67.50 Classic for only $3.95

A saving of 94%

Edited by Douglas M. Considine. Nearly 200 experts have contributed to this thoroughly revised and greatly expanded fifth edition of the most authoritative single-volume source of scientific knowledge ever assembled. Enormous 9¼" x 12" volume contains 2.2 million words, 2382 pages, 2500 photographs, drawings and charts, and 500 tables. 7200 articles cover everything from the earth and space sciences to energy technology, from the life sciences to mathematics, from information sciences to physics and chemistry. ". . . an amazing book . . . for both the general and scientific reader."—*The New York Times*

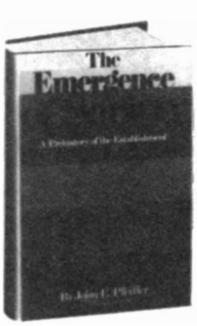
OR

any other 3 books for only $3.95

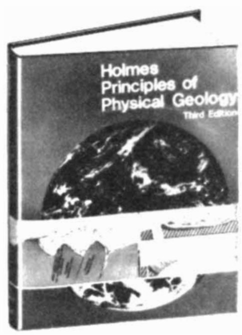
if you will join now for a trial period and agree to take 3 more books—at handsome discounts—over the next 12 months

(values to $71.40)

35740. THE BEGINNER'S GUIDE TO THE SKIES. *Clarence H. Cleminshaw.* Illustrated reference covering planet positions, constellations, paths of the sun, moon and stars. **$7.95**



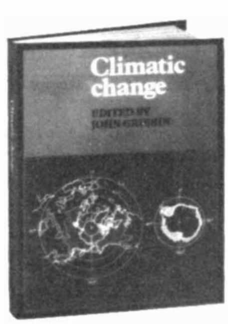
44330. THE EMERGENCE OF SOCIETY. *John Pfeiffer.* Offers a fascinating account of man's earliest religions, hierarchies, and cities. Profusely illustrated with 170 illustrations and drawings. **$15.00**



70195. PRINCIPLES OF PHYSICAL GEOLOGY. *Arthur Holmes.* New information on global mass movements and earth's energy dynamics. Profusely illustrated with over 800 photographs, maps, and diagrams. **$24.95**

36395. BLACK HOLES, QUASARS AND THE UNIVERSE. *Henry L. Shipman.* An absorbing exploration of the frontiers of astronomy. **$15.50**

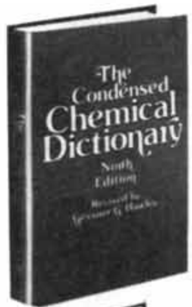
81403. SUN, MOON, AND STANDING STONES. *John Edwin Wood.* Includes the latest findings from researchers in archaeology, astronomy, climatology, and physics. Illustrated. **$14.95**



38362-2. CLIMATIC CHANGE. *Dr. John Gribbin.* Provides an A to Z report on the basic pieces and patterns of the earth's rich climatological puzzle. *Counts as 2 of your 3 books.* **$32.50**

68800. PHYSICS. Third Edition. *Resnick and Halliday.* Greatly expanded and rewritten reference includes new material on particle dynamics, fluid mechanics, reflection and refraction and polarization. **$24.95**

52130. HANDBOOK FOR CHEMICAL TECHNICIANS. *Strauss and Kaufman.* Contains over 200 illustrated laboratory problems with procedures of solutions laid out in a 1-2-3 fashion. 454 pages of direct data, problems and procedures, and 138 tables of information. **$19.50**



40167-2. THE CONDENSED CHEMICAL DICTIONARY. Updated to meet today's needs. Over 18,000 entries. *Counts as 2 of your 3 books.* **$32.50**

64145-2. A NEW DICTIONARY OF PHYSICS. *Gray and Isaacs.* Updated reference contains hundreds of illustrations, graphs, charts and tables. *Counts as 2 of your 3 books.* **$35.00**

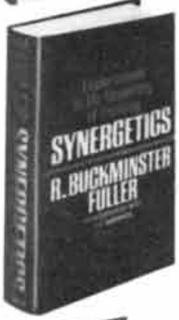
54995. THE ILLUSTRATED ENCYCLOPEDIA OF ARCHAEOLOGY. *Glyn Daniel.* The significant sites, cultures, civilizations, terms, tools, and techniques packed into one all-inclusive volume. 1000 entries. More than 100 illustrations. **$17.95**

50551. GRANTS: How to Find Out About Them and What to Do Next. *Virginia P. White.* A complete guide to "grantsmanship" and major sources of funding for anyone needing a grant. **$19.50**

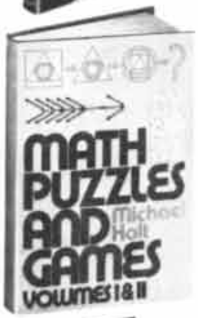
63340-2. MYSTERIES OF THE PAST. Beautifully illustrated volume explores the riddles of past civilizations. *Counts as 2 of your 3 books.* **$34.95**

50265. GEOMORPHOLOGY AND TIME. *Thornes and Brunsden.* Hundreds of figures, diagrams, and tables, it reviews new ideas, concepts and approaches in the field. **$12.95**

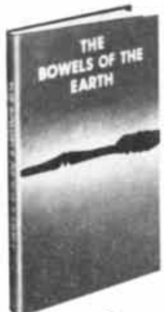
34730. THE ATLAS OF EARLY MAN. *Jacquetta Hawkes.* A panoramic overview of concurrent events in history through charts, graphs, and illustrations, from prehistory to the Roman Empire. **$20.00**



81840. SYNERGETICS. *R. Buckminster Fuller.* Mind-stretching work designed to teach us to understand the universe as a scenario in which we are the actors. **$25.00**

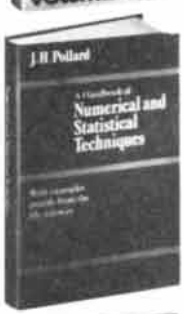


61100. MATH PUZZLES AND GAMES. *Michael Holt.* More than 300 mind-boggling brainteasers. Fully illustrated. **$13.95**



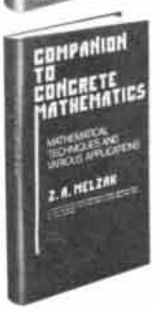
36785. THE BOWELS OF THE EARTH. *John Elder.* The first systematic text on the dynamics of the earth's interior, which joins together crustal geology and geophysics. **$14.00**

50560. GRAPH THEORY. *N.L. Briggs et al.* Complete sourcebook not only traces the history of graph theory but provides a thorough introduction to the science itself. **$22.50**

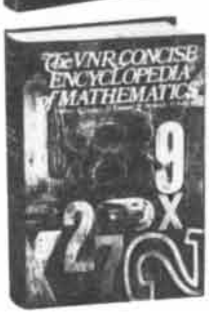


52292. A HANDBOOK OF NUMERICAL AND STATISTICAL TECHNIQUES. *J.H. Pollard.* Here are all the statistical and numerical methods you need to solve the mathematical problems on a programmable calculator, minicomputer or interactive terminal. **$24.95**

34140. APPLIED ANALYTICAL MATHEMATICS FOR PHYSICAL SCIENTISTS. *James T. Cushing.* This 650-page work for scientists deals with topics ranging from spectral analysis of linear operators to complex variables, to group theory. **$19.95**



39746-2. COMPANION TO CONCRETE MATHEMATICS. VOLUMES I AND II. *A. Melzak.* Mathematical techniques and various applications, with historical notes, solutions, on geometry, iteration, series and products, integrals, more. *The set counts as 2 of your 3 books.* **$46.45**



85865. THE VAN NOSTRAND CONCISE ENCYCLOPEDIA OF MATHEMATICS. *Edited by W. Gellert et al.* 800-page survey of mathematical ideas and methods lists hundreds of techniques, worked examples, and practical applications. 950 illustrations. **$18.95**

(Publishers' Prices shown)

If the reply card has been removed, please write to
The Library of Science
Dept. 2-B1A, Riverside, N.J. 08370
to obtain membership information and an application

tained in the population by distorted sperm ratios similar to those of *SD.* Mice are not as amenable to genetic analysis as flies, and so the details of this system are less well known; superficially, at least, the two systems appear to be very similar.

In plants a number of instances are known in which extra chromosomes that are harmful to the population are maintained by some kind of tricky behavior. One such chromosome is transmitted in excess of its rightful amount by getting itself included preferentially in the pollen-tube nucleus that fertilizes the egg. Other chromosomes take advantage of the fact that in females only one of the four products of meiosis is fertilized, the others becoming nonfunctional polar bodies; a chromosome may cheat by managing too often to get into the egg nucleus that is destined to be fertilized. In addition to these and other examples known in nature a number of cheating genes have arisen in the laboratory, sometimes by accident and sometimes in experiments designed to produce such genes.

Is it possible that meiotic-drive genes are commoner in nature than has been assumed? The examples that have been studied are all those with an extreme effect, such as *SD.* If *SD* produced a sperm ratio not of nearly 100 percent but of, say, 55 percent, it would probably never have been discovered. It is possible that mild cases of meiotic drive are rather prevalent in natural populations but have not been detected. Indeed, it has been suggested that the incidence of some human genetic diseases that are commoner than they ought to be may be explained by distorting genes, but there are other equally plausible explanations.

Mendelian inheritance is a marvelous device for making evolution by natural selection an efficient process. It would appear to be the best system that could be contrived (within the mechanical constraints imposed by the fact that genes are linked in chromosomes) for giving each gene a thorough test in combination with many other genes. The Mendelian system works with maximum efficiency only if it is scrupulously fair to all genes. It is in constant danger, however, of being upset by genes that subvert the meiotic process to their own advantage. If such genes have a harmful effect (as in the case of the sterility or lethality induced by homozygous *SD*), the population is weakened directly. Even if the cheating genes are not harmful in their own right, they inhibit the evolutionary process by reducing its efficiency. There are many refinements of meiosis and sperm formation whose purpose is apparently to render such cheating unlikely. And yet some genes have managed to beat the system.

The Medical Influence of the Stethoscope

The first instrument generally used for diagnosis, the stethoscope transformed the practice of medicine in the 19th century, altering the physician's picture of disease and his relation to the patient

by Stanley Joel Reiser

In 1816 a young French physician, René Théophile Hyacinthe Laënnec, was called to the bedside of a young woman suffering from the symptoms of heart disease. At that time most physicians arrived at a diagnosis by observing the patient's appearance and questioning the patient about symptoms. Faced with an unusually difficult diagnosis, Laënnec resorted to two little-used methods that involved physical contact with the patient: palpating her body with his hand to feel the underlying structures and rapping her chest with his fingers to generate sounds that would indicate the condition of her internal organs. Both of these efforts were thwarted by the patient's obesity, and the mores of the time prohibited Laënnec from undertaking immediate auscultation: pressing an ear to the chest to hear the sounds of the heart. Stymied, he recalled that sound is amplified when it is transmitted through certain solid bodies. He rolled a sheaf of paper into a cylinder and, placing one end on the patient's chest, put his ear to the other end. Through this makeshift instrument he heard more distinct sounds of the action of the heart than he had ever heard before.

Laënnec named his invention the stethoscope, from the Greek *stēthos*, "breast," and *skopein*, "to view." He soon abandoned the roll of paper in favor of what has come to be called the Laënnec stethoscope: a wood cylinder about a foot long and an inch and a half in diameter that could be disassembled into two parts for easy carrying. Running down the long axis of the cylinder was a narrow aperture; in experimenting with different designs Laënnec had discovered that such an aperture enhanced the sound-carrying properties of the instrument. The transmission of certain sounds, such as those of breathing, was further enhanced by a funnel-shaped cavity at the end of the cylinder placed on the patient's chest. If other sounds, such as those of the heart, were to be investigated, a small wood stopper was inserted into the cavity. Realizing that the stethoscope would enable him to hear the sounds generated by the motion of all the organs in the chest, Laënnec set to work to develop a systematic technique of diagnosis based on physical examination with the instrument. As a result of his clear and vigorous presentation of this technique, which he called mediate auscultation, the stethoscope became the first instrument of any kind to be widely used by physicians to diagnose illness.

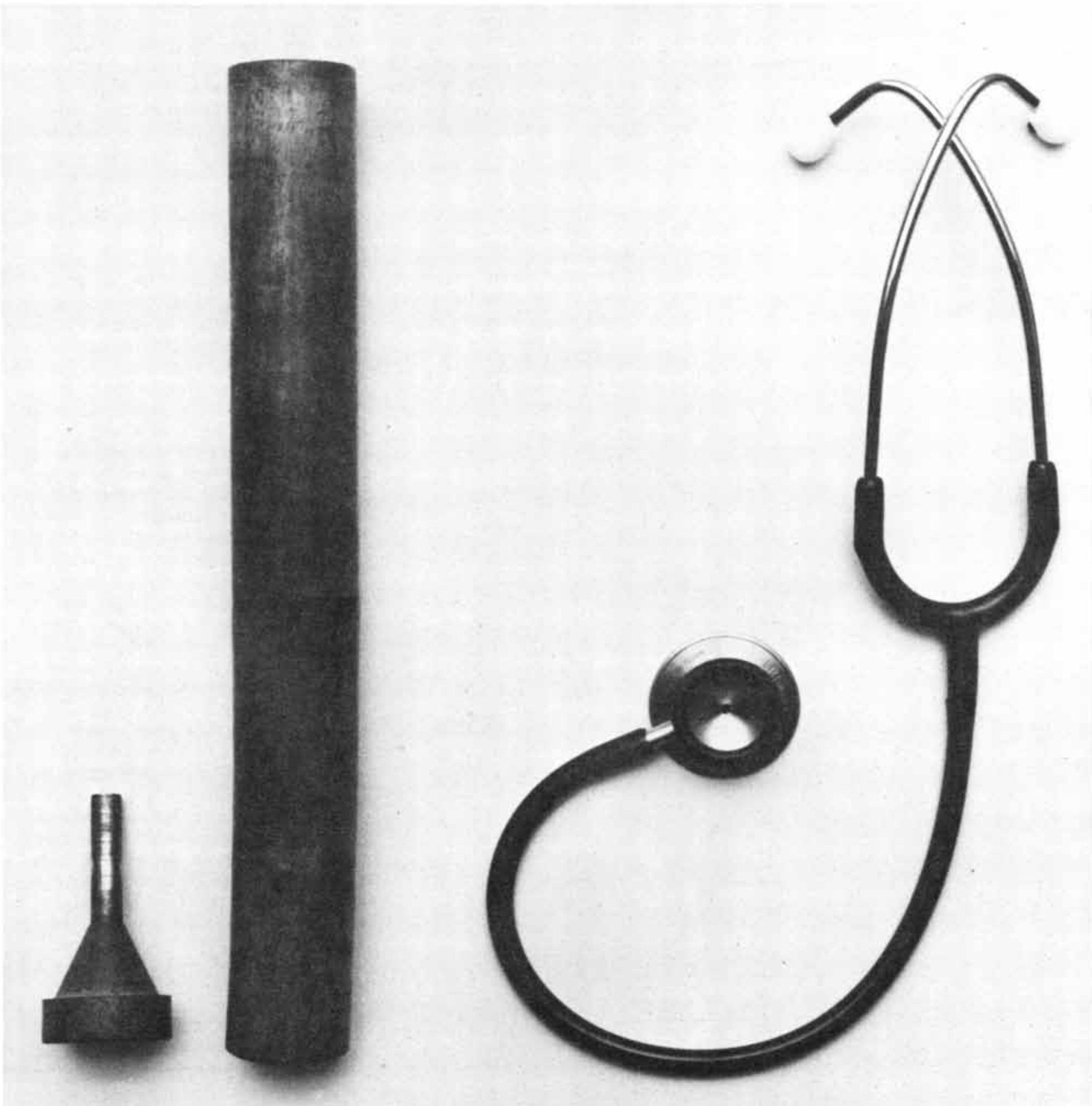
Diagnosis—the process of determining a patient's disorder from his symptoms—is clearly a crucial step in medical care. The acceptance of the stethoscope led to the establishment of physical examination as the keystone of diagnosis. It worked a profound transformation in the practice of medicine, altering both the physician's perceptions of disease and his relation to the patient. In particular the stethoscope drew the physician for the first time into a private world in which signs were directly communicated to him from the patient's body. It freed the physician to make objective diagnoses, but it also put him at a distance from the personal aspects of illness, opening a gap that is still widening. Indeed, today the stethoscope itself, perhaps the most familiar of all modern medical instruments, has fallen victim to the trend it started toward more objective types of diagnostic technology. To understand how this has come about, consider further the state of medical inquiry before the invention of the stethoscope.

At the beginning of the 19th century diagnoses were made principally on the basis of the patient's verbal account of his illness and the physician's observations (with his unaided senses) of the patient's skin color, tongue condition, breathing and so on. In the 18th century, however, a few manual techniques—methods of diagnostic exploration based on physical examination—had begun to attract attention. One of the most significant, in terms of the changing attitude toward clinical diagnosis, was the technique of percussion that Laënnec had used on his patient with the heart disorder: striking the body to generate sounds that would indicate the vitality of the internal organs.

Percussion was the invention of the Viennese physician Leopold Auenbrugger, who first described it in his *Inventum novum*, a small monograph published in 1761. Auenbrugger pointed out, as Laënnec would 58 years later, that the physician's observations of the external signs of illness were of little use in the diagnosis of chest disease. Patients with the same disorder often exhibited different symptoms, and patients with different disorders often exhibited the same symptoms. Auenbrugger was also critical of the patient's own account of his symptoms, which he considered variable and untrustworthy. For Auenbrugger it was the morbid sound he himself elicited by percussing a patient that was the most dependable index of the nature and course of chest disease, and in *Inventum novum* he proclaimed his ability to diagnose (to translate from the Latin) "from the testimony of my own senses."

Auenbrugger was concerned with finding objective evidence of disease, and he believed that the technique of percussion, which minimized the importance of the personality and physical appearance of the patient, would revolutionize the diagnosis of chest disorders, giving physicians a dependable alternative to their usual untrustworthy methods of diagnosis. Unfortunately Auenbrugger's technique of inferring the nature of an internal abnormality from the character of the sound created in striking the chest was in such strong opposition to reigning medical theory that in his lifetime percussion never gained the attention it deserved.

To appreciate percussion a physician had to think of disease in anatomical terms, explaining the genesis of symptoms in a patient in relation to structural changes in the tissues of the body. This picture of disease was highly unorthodox in the mid-18th century, when most physicians still held to the ancient Greek view that illness was caused by an imbalance in the basic body fluids known as humors. An interest in normal anatomy and its structural transformation through disease had been developing since the 16th century, when revolutionary treatises such as Vesalius' *De humani corporis fabrica* were published, but



STETHOSCOPE HAS EVOLVED from a rigid monaural instrument made of wood to a flexible binaural one generally made of rubber, plastic and metal. The wood instrument at the left is a Laënnec stethoscope. It is named for its inventor René Théophile Hyacinthe Laënnec, who in 1816 first used a rolled sheaf of paper to amplify and convey to his ear the sounds generated by the motion of the organs in the chest. The sound-carrying properties of the Laënnec stethoscope are enhanced by a small aperture that runs its length and widens into a funnel-shaped cavity at the end to be placed on the patient's chest. The small piece of wood at the lower left is inserted into the cavity to improve the transmission of certain sounds such as those made by the heart. An acoustically superior binaural stethoscope that was also flexible first came into wide use in the 1890's. A modern version of this instrument, the Littmann stethoscope, is at the right. On one side of its chest piece there is a bell-shaped metal cup rimmed with rubber for transmitting sounds of lower pitch; on the other side there is a wider, flatter cup covered with a thin, taut plastic diaphragm for transmitting sounds of higher pitch. The wood stethoscope shown here is now in the Warren Anatomical Museum of the Harvard Medical School; it may have belonged to Laënnec himself.

18th-century physicians still thought anatomical abnormalities were of secondary importance in gaining an understanding of the genesis of disease and the generation of symptoms. A work that would profoundly influence physicians to think in terms of anatomy, the Italian anatomist Giovanni Battista Morgagni's *De sedibus et causis morborum per anatomen indagatis* (*The Seats and Causes of Diseases Discovered by Anatomy*), was published the same year as Auenbrugger's *Inventum novum* and therefore had little effect on the early reactions to percussion. (By the time Laënnec began to promote mediate auscultation, however, efforts such as Auenbrugger's and Morgagni's had resulted in a somewhat more favorable attitude toward anatomical explanations of disease.)

The lack of enthusiasm for percussion also had to do with Auenbrugger's manner of presentation. The variations in percussive sounds by which diseases can be discriminated are subtle. To learn how to apply the method the physician needed extensive descriptions and many examples of these complex acoustic phenomena. In Auenbrugger's brief treatise he chose to give only terse, vague explanations, trusting that his discovery would be valued "by those who can firstly appreciate medical science." Indeed, he refused to elaborate on his technique for skeptics or critics, whom he judged to be "beset by envy, malice, detraction and calumny."

Auenbrugger's technique of percussion did have several failings. For example, it was not effective with patients whose disorders were located deep within the chest cavity. The comparison of healthy and morbid sounds in the same patient was an important part of diagnosis by percussion, and so the technique also did not work well with patients who had two diseased lungs. Modern physicians who employ percussion put the fingers of one hand on a patient's body to soften the percussive blow, but the form of percussion Auenbrugger advocated was harsher, requiring that the physician deliver a blow directly to the patient's body with the tips of all the fingers on one hand. This procedure could be quite distressing to the patient, and some physicians refused to employ it for that reason. Moreover, at that time any manual activity was generally considered to be beneath the dignity of the physician. In particular the physical intimacy required by percussion threatened to undermine the professional standing of the physician, even to place him in a class with the surgeon, over whom he affirmed both medical and social superiority.

In his time Laënnec was confronted with much the same kind of medical and social resistance to manual physical examination that Auenbrugger had faced in the mid-18th century. There had been one significant change in the medical climate in the time between the introduction of percussion and the introduction of mediate auscultation. By 1816 physicians were generally willing to believe specific anatomical defects were dependable hallmarks of certain disorders and could be exploited to differentiate among diseases. In particular the correlation of the symptoms in the living patient and the structural changes found in the body at autopsy was beginning to play an important part in the discovery and verification of medical facts (and in the training of physicians). Therefore Laënnec could demonstrate through autopsy that when sounds of a certain kind were heard in the chest, lesions of a certain kind were found in the cadaver.

Like Auenbrugger, Laënnec believed a set of audible signs could be the basis of a new and more dependable technique of diagnosis. He spent three years at the Necker Hospital in Paris auscultating healthy subjects and sick ones, listening to the murmurs, rales (crackling sounds), wheezes and burbles inside their chest and performing autopsies in order to identify and classify the sounds. In 1819 he published *Traité de la auscultation médiate* (*On Mediate Auscultation*), a comprehensive critique of the diagnostic methods of the time that culminated in an exhaustive description of the technique of mediate auscultation. The treatise is a meticulous exploration of the connections between the sounds generated within the body and specific anatomical defects and diseases. Some 10 times longer than Auenbrugger's monograph on percussion, Laënnec's work contains descriptions of sound that surpass in clarity most that have appeared since. "When the patient coughed or spoke, and still more during respiration," wrote Laënnec in one case study,

LAËNNEC, shown in an engraving made during his lifetime, also developed an innovative, systematic technique for applying the stethoscope in the diagnosis of illness. Before the invention of the stethoscope diagnoses were based primarily on the patient's account of his symptoms and the physician's observations of the patient; direct physical examination of the patient played little or no part in the diagnostic process. Laënnec's technique, which he called mediate auscultation, offered a more dependable approach to the evaluation of disease because it was based on objective analysis of aural signs detected in physical examination with stethoscope.

IT CAN RUN A MILE CHEAPER THAN YOU CAN.

The Rabbit Diesel runs a mile, and burns about 1.3 cents' worth of fuel.

Compared to that, you're a guzzler.

If you weigh 150 pounds, you'd burn around 90 calories per mile. Figure that as a mere fourth of a fast-food cheeseburger, and it comes to about 18 cents.

Fact is, if you were a car, you couldn't afford you.

So, don't walk. Run for a Volkswagen Rabbit Diesel. According to the 1979 EPA Mileage Guide, our Rabbit Diesel gets the highest mileage in America: A whopping 50 MPG on t e highway, 40 MPG in the city, using a 4-speed transmission.

(These estimates may vary depending on how and where you drive, optional equipment and your car's condition.)

In addition to giving you the best run for your money, the Rabbit Diesel doesn't require conventional tune-ups. There's nothing much to tune. No spark plugs, points, condensers, or carburetors.

Like all diesels, the Rabbit Diesel has a great reputation for reliability. But unlike many diesels, the Rabbit Diesel responds like a shot from a gun (0 to 50 MPH in 11.5). From its rack-and-pinion steering to its front-wheel drive, there's not a sluggish bolt in its body.

Obviously, all cars run on some kind of fuel. So do you. But what you save with a Rabbit Diesel, can fuel you with steak.

VOLKSWAGEN DOES IT AGAIN

Harris technology on the job

In information systems,
Harris produces data processing terminals, general-purpose computers, word processing systems and supervisory control systems.

Harris computer system at Shell Oil's Science and Information Center

Harris offers a complete family of high-performance, general-purpose computer systems. Providing virtual memory, data base management and a multitude of languages for scientific and commercial applications.

Harris technology works worldwide—in communication equipment, information systems, government systems, semiconductors and printing equipment. For information, write: Harris Corporation, Melbourne, Florida 32919.

HARRIS

COMMUNICATIONS AND INFORMATION HANDLING

"there was heard a tinkling like that of a small bell which has just stopped ringing or of a gnat buzzing within a porcelain vase." Such vivid descriptions had much to do with the success of mediate auscultation in gaining the attention and respect of a wide audience of physicians.

Initially physicians found a variety of reasons to criticize the technique. Some thought they would appear ludicrous, bent over the bodies of their patients listening for sounds through an instrument that resembled a divining rod. And there was still the concern that the use of instruments and manual techniques was beneath the station of the physician. The transporting of medical instruments was particularly disturbing, and some physicians resorted to such expedients as carrying the two disassembled parts of the stethoscope crossed inside their top hat. (This particular practice was not without its perils: on one occasion, when a snowball knocked off a medical student's top hat and revealed a stethoscope, the student was charged with carrying a concealed weapon.)

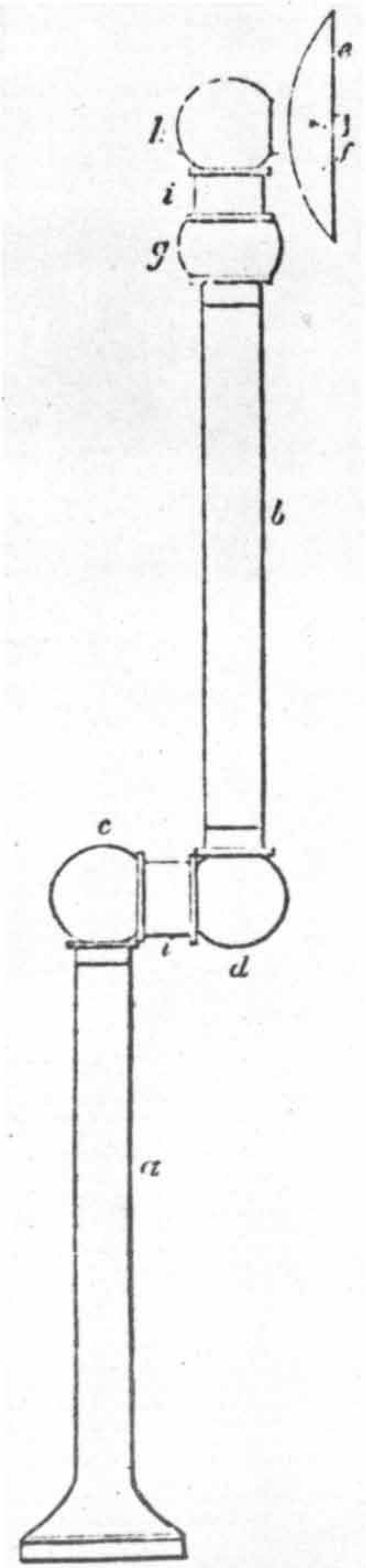
Furthermore, many physicians found the views of the early proponents of the stethoscope extreme. In *On Mediate Auscultation* Laënnec judged the standard methods—even the manual techniques—of the day harshly, finding palpation "too vague and uncertain to be of much benefit" and percussion only slightly more valuable. He was also critical of the ancient practice of pulse taking, explaining its popularity by pointing out that the procedure inconvenienced neither the physician nor the patient, and that any particular irregularity was subject to a wide range of interpretations. Laënnec and many of his followers maintained that authoritative diagnoses could be made principally—indeed, sometimes exclusively—on the basis of aural information. Other less partisan physicians believed the most reliable diagnosis could be arrived at by giving equal consideration to aural signs and information gained through traditional methods; they were understandably annoyed by the uncompromising views of Laënnec and his followers. Actually the excessive claims made for the stethoscope were due in part to Laënnec's extraordinary abilities. Having remarkably acute hearing and a precise knowledge of the pathological meaning of the most delicate variations in sound, he could distinguish diseases on the basis of auscultatory information with wonderful accuracy. Few of Laënnec's followers appreciated how unusual his skill was.

The reactions of patients to the stethoscope were generally favorable, although some were frightened by the device (associating instruments with surgeons) or embarrassed by the physical closeness of the physician required by auscultation. Some patients refused to be examined with the stethoscope because of the excessive force with which their physician pressed the instrument against them or because of the inconvenience of having to make frequent changes in position during such an examination. What patients feared most, however, was the accuracy of the stethoscope. Often it converted a last hope that an illness was remediable into a certainty that it was not.

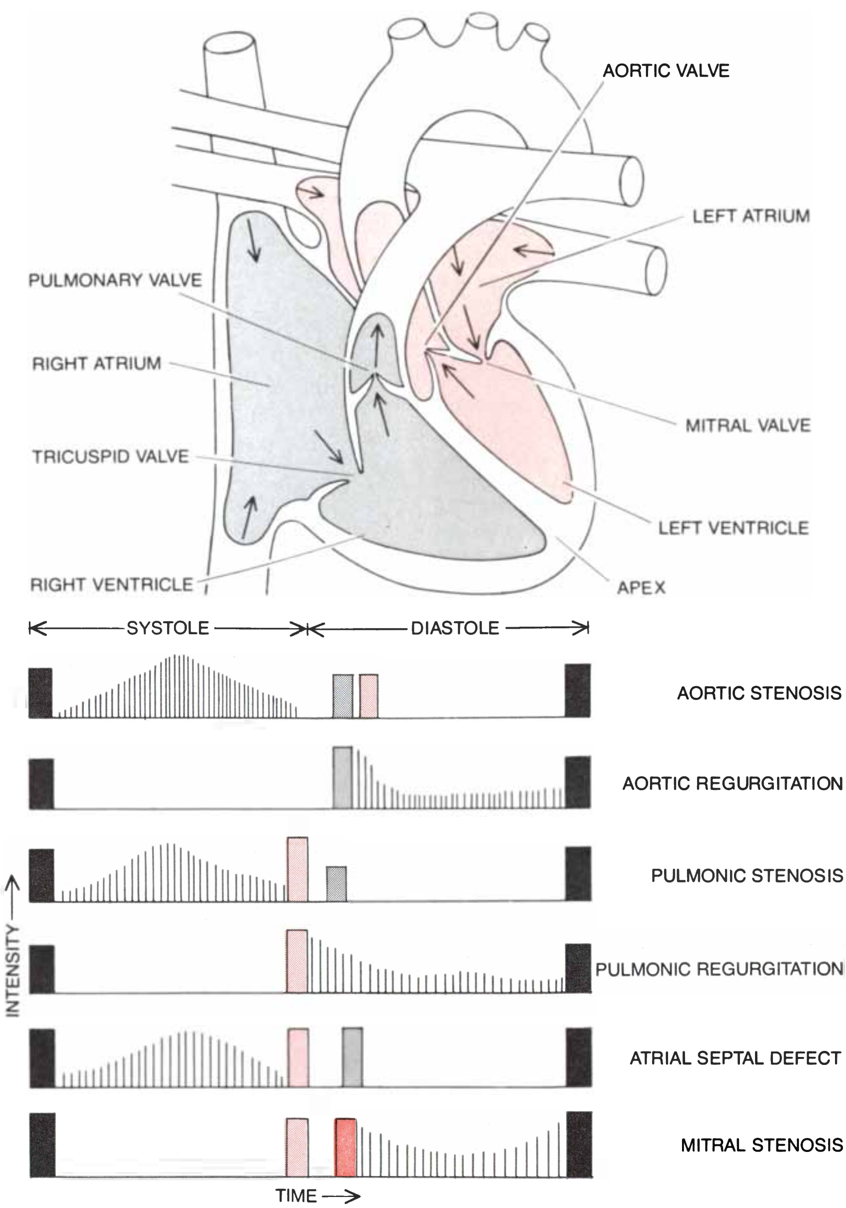
In the final analysis, of course, the great strengths of mediate auscultation were just this accuracy and dependability: the technique could convert diagnostic suspicions into diagnostic certainties. As Laënnec and his followers pointed out, having confidence in a diagnosis meant that diseases could be treated with new vigor. Therapies could be applied promptly, in some cases at early stages of illness. And even if the disease diagnosed was incurable, the patient could at least be spared useless and perhaps painful therapy.

The accuracy of diagnoses made with the stethoscope resulted in a growing public acceptance of mediate auscultation. In fact, within a decade of the publication of *On Mediate Auscultation* popular sentiment in favor of the technique had become so strong that a physician who did not employ the stethoscope jeopardized his professional reputation. Some physicians who did not know how to auscultate patients took to carrying stethoscopes with them for display purposes. Physical examination—through auscultation—was the new mode of diagnosis.

In the second half of the 19th century many improvements were made in the rigid, monaural Laënnec stethoscope. To accommodate physicians who resented carrying such a bulky object smaller stethoscopes were developed. Changes in the design of the instrument (for example the substitution of a parabolic cavity for the conical one at the end to be placed on the patient's chest) improved its acoustic properties. As early as 1829 the first stethoscope that was not completely rigid was built by an Edinburgh physician named Nicholas Comins. The instrument consisted of



PARTIALLY FLEXIBLE stethoscope was designed in 1829 by Nicholas Comins, an Edinburgh physician. Made of two rigid tubes (*a*, *b*) connected by a joint and attached to a movable earpiece (*e*), the instrument could be bent at any angle in a plane, so that the patient did not have to change his position as often during an examination and the physician did not have to bend as much. Eventually fully flexible stethoscopes made of pliable tubing were developed; these spared the patient the discomfort of having a rigid instrument pushed against his body and also reduced the need for close physical contact between the physician and the patient in the course of auscultation.



PATHOLOGICAL HEART MURMURS are blurred auscultatory heart sounds caused by structural defects in the tissue of the heart. The intensity and duration of several common sounds of this type (as they are heard at specific locations on the chest) are shown at the bottom; the flow of blood through the heart, at the top. Blood is pumped through the body by the contractions of the ventricles of the heart. The period during which the right and left ventricles are contracted and blood is forced into the pulmonary artery and the aorta is systole; the period during which the ventricles are relaxed is diastole. In a single cardiac cycle of a normal heart two sounds are heard: first the simultaneous closing snaps of the mitral and tricuspid valves, which prevent the blood from flowing back into the atriums as the ventricles contract, and second the not quite simultaneous closing snaps of the pulmonic and aortic valves, which prevent the blood from flowing back into the ventricles as they relax. In the graph a black rectangle represents the first sound in each cycle; a gray rectangle (the pulmonic component) and a light colored rectangle (the aortic component) represent the second sound. Sounds of lower pitch are indicated with more widely spaced lines. If the aortic valve is narrowed or obstructed (the defect known as aortic stenosis), then during systole, when blood is forced through the valve, an unusual amount of turbulence is generated and a rising and falling sound called an ejection murmur is created. In aortic regurgitation the aortic valve does not function, so that during diastole blood flows back into the ventricles, creating a fairly steady sound called a regurgitation murmur. If there is an atrial septal defect in the heart (a hole in the wall separating the atriums), then during diastole a large amount of blood flows through it, causing the aortic and pulmonic components of the second heart sound to be widely separated; during systole the right ventricle ejects the blood into the pulmonary artery, causing an ejection murmur. A different type of sound is created when the mitral valve is narrowed or obstructed (the defect known as mitral stenosis). In this case there is an audible opening snapping sound of the mitral valve (*dark colored rectangle*) at the beginning of diastole that is followed by a low-pitched rumbling sound as the blood flows into the left ventricle. The graphs shown in the illustration are adapted from the book *Bedside Diagnostic Examination*, by Elmer L. DeGowin and Richard L. DeGowin.

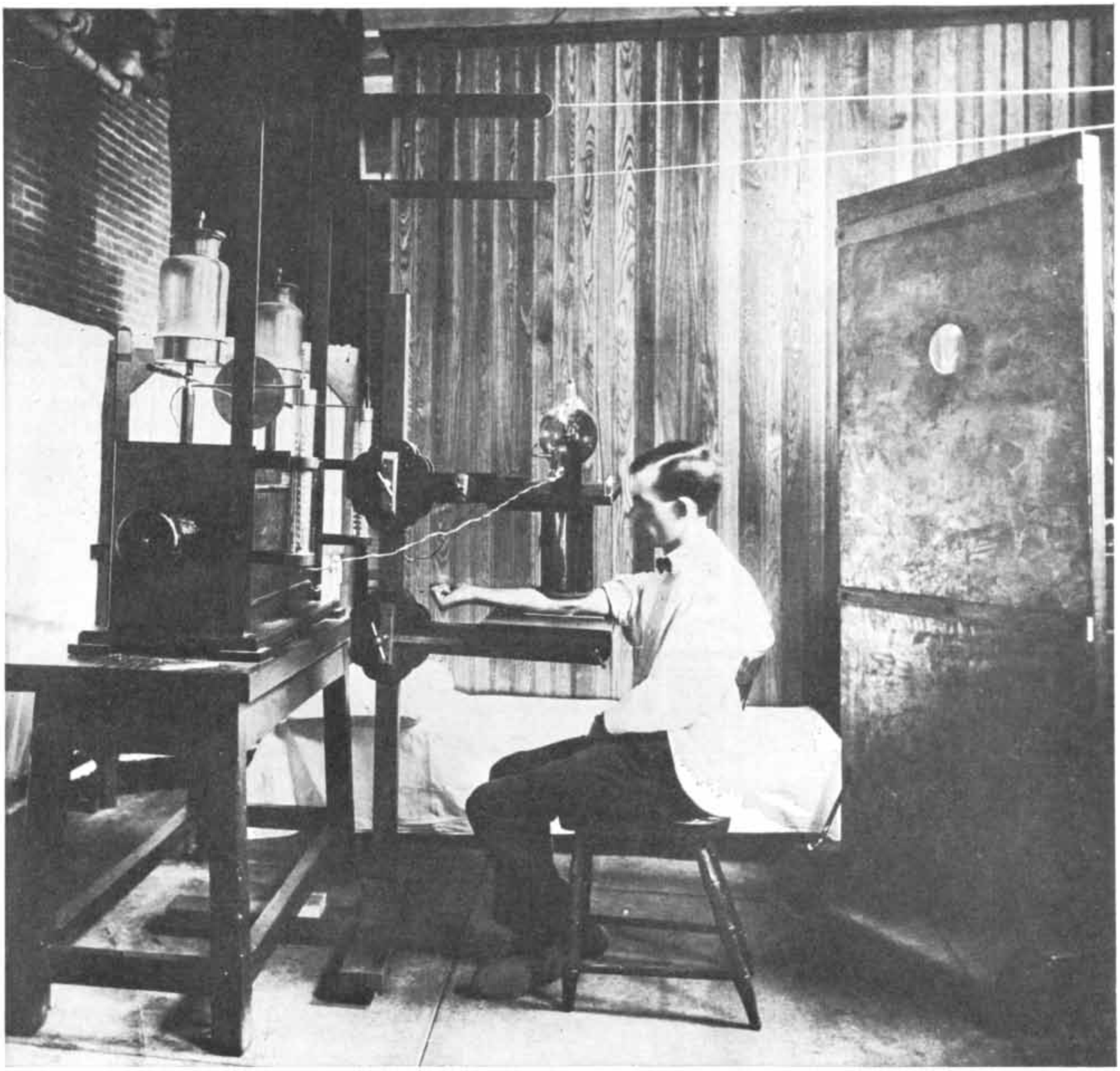
two rigid wood tubes, each one about seven inches long, connected by a joint so that they could be set at any angle; an earpiece attached to one of the tubes also rotated about a joint.

Fully flexible monaural stethoscopes made of pliable tubing first appeared in the 1830's. These instruments, which could be up to two and a half feet long, spared patients having to change position repeatedly during an examination and the discomfort of a rigid instrument pushed against them. Moreover, with a flexible stethoscope patients were no longer forced into such close contact with the physician; a woman could place the cup of the flexible stethoscope on her body herself, thereby keeping the physician at a proper distance. Similarly, physicians could avoid close contact with patients who had contagious diseases. They were also spared the discomfort of bending to examine patients (a particular boon for fat physicians), and they could even auscultate themselves.

Comins was probably also the first to suggest the binaural stethoscope: a flexible stethoscope with tubes going to both ears. It was not until the early 1850's, however, that the first satisfactory working models were introduced. In 1878, when the microphone was invented, an even greater amplification of body sounds seemed possible, but it was soon evident that the heart and lung sounds picked up by the microphone were not clear enough for diagnostic purposes.

The initial excitement and later disappointment over the microphone coincided with the introduction of a large number of different types of stethoscope, and much confusion resulted. Physicians began to worry that any amplification would alter the sounds to which they had become accustomed. Many chose to ignore the vast array of different stethoscopes, insisting that it was the learning of the physician that mattered in auscultation, not the instrument he wielded. As a result the acoustically superior (and handily carried) flexible binaural stethoscope was not generally adopted until the 1890's. In the early 1900's physicians realized that a diaphragm stretched over the mouth of the cup of the stethoscope would enhance the transmission of certain kinds of sounds, and so this feature was added to the instrument. Modern stethoscopes follow the same basic design and generally have two cups: one bell-shaped cup without a diaphragm for transmitting sounds of lower pitch and a wider, flatter cup with a diaphragm for transmitting sounds of higher pitch.

Paradoxically the acceptance of the binaural stethoscope coincided with the discovery of a device that would quick-



EARLY X-RAY MACHINE is demonstrated in this photograph made at the Massachusetts General Hospital in Boston in about 1910. X rays were discovered in 1895 (at about the same time that the flexible binaural stethoscope came into wide use) and were quickly applied to a variety of medical tasks, such as diagnosing bone fractures and locating kidney stones. Heart and lung disorders in particular could often be diagnosed earlier and more accurately with X rays than with a stethoscope. As a result the X-ray machine began to displace the stethoscope as the preeminent instrument in the analysis of chest disorders. In the 20th century the stethoscope has been increasingly ignored in both the practice of medicine and medical training in favor of more sophisticated modern diagnostic technologies.

ly displace the stethoscope as the preeminent tool in the diagnosis of chest disease. In 1895 W. K. Röntgen discovered that a cathode-ray tube emitted radiation that could pass through a solid object and fix the image of the object's internal structure on a photographic plate. News of Röntgen's discovery of these X rays, a "new kind of light which penetrated and photographed everything," spread quickly. The X-ray picture he had made of his wife's hand was reproduced in publications all over the world, and newspapers competed with one another in publishing exaggerated accounts of the potential uses of the discovery. The general public avidly bought X-ray images of their loved ones' hand and of ordinary objects such as keys taken through an opaque container such as a purse. (The queen of Portugal had X-ray pictures made of all the members of her court; the distorted rib cages disclosed by the X rays led some of the female members to discard their tightly laced corsets.)

Physicians were immensely enthusiastic about Röntgen's discovery and quickly applied X rays to the diagnosis of bone fractures and kidney stones and to the location of foreign objects such as bullets embedded in body tissue. Within a month of the discovery an apparatus was in use that projected X rays through a patient's body onto a fluorescent screen so that the physician could even directly observe lung and heart motions. As X-ray techniques improved it became possible to analyze disorders within the softer tissues of the body such as the lungs and the heart.

With X rays physicians could often

discern lung and heart disorders earlier and define their character more accurately than they could with a stethoscope. In 1896 the pioneering radiologist Francis H. Williams expressed the sentiments of many other physicians when he wrote: "We may now look where we have previously only been able to listen and sometimes to hear but imperfectly." To many the transformation of auscultatory sounds into images in the mind of the physician seemed less reliable than the direct analysis of defects displayed in X-ray pictures. Patients too had more confidence in a diagnosis based on an anatomical defect they could see in a picture. The use of the stethoscope began to decline, a trend that was accelerated in the 20th century by the development of more sophisticated types of diagnostic technology such as the electrocardiograph. Indeed, as physicians became increasingly captivated with diagnostic technology they became less concerned with physical examination in general and with the techniques developed in the 19th century for accomplishing it.

Today the stethoscope is the old warrior of medicine. Although it cannot compete with the array of elaborate and expensive technologies for which it paved the way, it clings tenaciously, resisting retirement. Its staying power in modern times is based in part on its giving both physicians and patients a sense of continuity with the past. Identified with dependable diagnosis, the familiar object evokes confidence. Most important, it provides those physicians who still know how to use it with good, immediate and low-cost information that can eliminate the need for complicated diagnostic tests.

It is, of course, easy for the modern physician to refer his patient to a clinical laboratory or X-ray department for diagnostic tests instead of conducting an intensive examination with a stethoscope. Such tests are usually covered by insurance policies, so that there is little immediate cost to the patient and the physician saves time. Indeed, as fewer physicians are trained in methods such as auscultation, choosing the high-technology route is only prudent. What is lost is primarily the self-reliance of the physician. When the 19th-century physician chose to make diagnoses less on patients' verbal accounts of their symptoms and more on the physical signs of illness that in many cases he alone detected, he was obliged to make up his own mind about illness. As a medical era the 20th century must be characterized as a time when physicians have come to rely less on themselves and more on specialists, technicians and machines to collect and evaluate the evidence of disease.

THE AMATEUR SCIENTIST

Strange to relate, smokestacks and pencil points break in the same way

by Jearl Walker

The connection between a falling chimney and a breaking pencil point is far from obvious, but in most of what follows I shall undertake to make the connection clear. Then I shall take up a quite unrelated matter: an ingenious device that couples a telescope and a pocket calculator in such a way that changes in the angle at which the telescope is pointed are recorded by the calculator and can be directly read off the calculator's display panel.

If you ever have an opportunity to watch as an old chimney (the tall, freestanding and basically cylindrical kind) is demolished, look carefully at one side of it as it begins to fall. In the usual demolition procedure a section of the base is knocked out by dynamite or a bulldozer, forcing the chimney to fall to one side. By the time it has reached a tilt of about 45 degrees it will probably show a lateral crack near its middle. As a result the top part of the chimney begins to fall more slowly than the bottom part, and the two parts form a broad *V*. The reason for the break is that as the chimney begins to fall the top part necessarily has to accelerate downward more rapidly than the bottom part. The chimney is usually not strong enough to withstand the bending stresses and the acceleration, and so it cracks.

Ernest L. Madsen of the University of Wisconsin Medical School has recently examined the mechanics of a falling chimney. He first considers the forces on the entire chimney once it is falling. It is subjected to three forces: its own weight (acting through the center of mass), an upward reaction force from the ground and friction (horizontally) from the ground. The combination of these forces makes the chimney accelerate toward the ground and rotate vertically at the base.

Madsen next considers a lower section of the chimney extending from the base to some arbitrary point along its length. This section has an angular acceleration around the base throughout the fall because of three torques acting on it. One torque arises from the section's own weight. The other two torques arise because the lower section must drag the upper section around in the rotation. One of these torques is due to the shearing force resulting when the upper section attempts to slide over the top of the lower section. Finally, the lower section is subjected to a bending torque because the upper section tends to lag behind in the rotation and so to bend backward.

Except for the exact bottom and top of the chimney any theoretical choice for the length of the lower section will include a bending torque acting to sever the lower section from the upper one. According to calculations, the bending torque is greatest at a point a third of the way up the chimney. As the chimney begins to fall and to be subjected to an angular acceleration around the base, the bending torques along its entire length increase with time. Eventually the bending torque at the point a third of the way up the chimney is enough to rupture it, so that the upper two-thirds breaks off and begins to lag behind the lower third.

Madsen's results pertain to a chimney that is uniformly cylindrical. In a chimney that is tapered or has some other form the maximum bending torque will be at some other point along its length. If you are able to watch a chimney being toppled, you might try to photograph the break. It would be particularly interesting to photograph the action in slow motion so that you can see the crack begin in the lower section and then propagate across the width of the chimney. (Stay alert. When I watched a falling chimney recently, I had to dodge bricks that were thrown into the air when the chimney hit the ground.)

The rupture does not travel directly across the width of the chimney. In 1940 Francis P. Bundy explained the rupture pattern as being due not only to the bending stress but also to the accompanying compression of the chimney on the trailing side. The bending apparently forces the line of maximum compression downward as the rupture travels to that side.

Some of Bundy's calculations suggest that if a tall chimney does not break during its fall, something strange may happen: its base may hop up into the air near the end of the fall, apparently because the chimney is rotating about its center of mass. This result is not predicted by Madsen's calculations. You might want to watch the base of a chimney that falls without breaking to see if it does in fact hop.

A falling chimney can display two other interesting features. One is another break point that may develop close to the base because of shearing as the top part of a tall, heavy chimney attempts to slide over the lower part. Some photographs of falling chimneys clearly reveal the extra breaking: the chimney breaks in two places.

The other interesting feature is that the base of the chimney may slide during the fall. I would have guessed that such a slide would be in the direction opposite to the direction of fall, but normally it is not. The base moves the way the chimney does, being dragged along in the direction in which the chimney has acquired some horizontal momentum. According to Madsen's work, the base is most likely to slide in this way when the chimney has fallen through at least 50 degrees. Beyond that angle the amount of friction needed to hold the base in place becomes impractically large.

The falling chimney can be modeled somewhat imperfectly with an unsharpened pencil that is placed vertically on a level surface, eraser end up, and then allowed to fall. The lower end does come to rest displaced in the direction of the fall, but the motion is more than just a simple slide. The end first moves a short distance in a direction opposite to the direction of fall, but as the pencil bounces on the surface, the end hops a greater distance in the direction of fall. The static friction on the lower end of the pencil is apparently not enough to keep that end from sliding along the surface as it tends to rotate about the falling center of mass. The later motion in the direction of fall results from the lower end's being dragged in the direction in which the falling pencil has acquired horizontal momentum.

Madsen has suggested to me a simple arrangement of Tinker Toy components with which you can simulate the bending of a falling chimney. A Tinker Toy set includes a squat wood cylinder with a central hole through it. You need a stack of about 30 of them to simulate the chimney. A Tinker Toy stick fits snugly through the bottom cylinder. Sharpen the lower end of the stick so that the stack will be held in place when you push the point into a small piece of sponge. Run an elastic band, preferably a latex one, around the sharpened stick and up through the central holes in the cylinders. You will probably need several of these bands looped together to

Get a head start on your work in 1979.

Special offers make this the time to buy a Texas Instruments programmable.

Buy a TI-58 and choose two free software Specialty Pakettes from TI, a $20.00 value. Select from: Electronic Engineering. Civil Engineering. Fluid Dynamics. Blackbody Radiation. Oil/Gas/Energy. 3-D Graphics. Mathematics. Statistical Testing. Marketing/Sales. Production Planning. Astrology.

TI Specialty Pakettes are convenient notebooks containing ready-to-use programs written by professionals in many fields. Each pakette includes step-by-step program listings, applications notes and sample problems. Just key in the program you need and you can put it to work right away.†

Specialty Pakette programs are easy to use with the powerful TI-58 key programmable. Up to 480 program steps or up to 60 memories. Over 170 functions and operations. TI-58, only $125.00*.

Buy a TI-59 and get a $10.00 rebate from TI. The TI-59 is a computer-like programmable with up to 960 program steps or up to 100 memories. Over 170 functions and operations. Plus magnetic card read/write capability. TI-59, $300.00*.

Ready-to-use programs for both the 58 and 59 are available in Solid State Software™ libraries featuring plug-in micro-memory modules. The Master Library Module is included, with 25 programs in math, statistics, finance and conversions.

Twelve optional libraries are available in fields like engineering, investments, business and statistics.

See your dealer to choose the TI programmable that's right for you. Use the coupon below to take advantage of special savings during these limited time offers.

†Specialty Pakettes do not include plug-in modules or magnetic cards.
*U.S. suggested retail price.

SA-279

☐ I've bought my TI-58, please send me these two free TI pakettes:
1. __________ 2. __________

☐ I've bought my TI-59, please send me my $10.00 rebate.

Texas Instruments will fulfill the offer you have selected above when you: (1) return this completed coupon, including serial number, (2) along with your completed customer information card (packed in box), and (3) a dated copy of proof of your purchase verifying purchase of a TI Programmable 58 or 59 (whichever is applicable) between January 1 and February 28, 1979. **Items must be postmarked on or before March 7, 1979, to qualify for this special offer.**

Send to: TI-58/59 Special Offer, P.O. Box 53, Lubbock, Texas 79408.

Name __________

Address __________

City __________

State __________ Zip __________

Calculator Serial Number __________
(from back of calculator)

Please allow 30 days for delivery. Offer void where prohibited by law. Offer good in U.S. only.

Texas Instruments technology—bringing affordable electronics to your fingertips.

TEXAS INSTRUMENTS
INCORPORATED

45632

reach the top of the stack. Place across the hole of the top cylinder a smaller stick around which the top band is looped. The cylinders should be aligned to make a straight "chimney," and the bands should not be tangled inside the cylinders. The tension should be adjusted so that the chimney bends if it is leaned to one side.

The pointed end of the lower stick is stuck into the sponge, which is glued or clamped to a large board held stationary on a tabletop. When the chimney is pushed gently to one side and falls, it bends backward much as a real chimney does. The base does not slip away from the direction of fall because of the friction that arises between the stick and the sponge.

Since the fall is rather quick, a photograph of the falling stack is more revealing than direct observation. Madsen suggests that motion pictures made at 36 frames per second or faster are best. If you lack the equipment for that, you can make a snapshot of the chimney with a strobe light. Turn off the room lights, open the aperture of the camera, cause the chimney to start falling, flash the strobe light during the fall and then close the aperture before turning on the lights again.

A common demonstration apparatus in physics classes is related to the falling chimney and illustrates the large acceleration at the top of a chimney that does not crack. A board is hinged at one end to a base that sits on a tabletop. The other end of the board is propped up with a stick so that the board makes an angle of about 35 degrees with the tabletop. A small paper cup is fastened to the upper side of the board near the free end. Still closer to the free end is a holder for a small metal ball. The holder

A chimney breaks as it falls

A Landmark Work on our Ecological Crisis

ECOSCIENCE

Population, Resources, Environment

Paul R. Ehrlich, Anne H. Ehrlich, *Stanford University* John P. Holdren, *University of California, Berkeley*

Ecoscience presents extensive, current information on virtually every aspect of the environmental and resource sciences. The authors also suggest strategies for solving some of our most pressing environmental problems.

"The real question, for those concerned about realistic solutions, is whether scholars and decision-makers of all varieties can devise ways to bring human behavior into harmony with physical reality in time."—The authors

"The ultimate book on the future of our planet."
—Maurice Strong, First Director of the United Nations Environmental Programme

"A blockbuster classic. . . . Its 1,050 pages, 370-plus illustrations and more than 1,000 references contain the stuff of several books. . . . Even those areas in which I have focused my own professional activities are covered with material that I found comprehensive, up-to-date, and thought-provoking." —Dennis L. Meadows, *Sierra*

"It is unlikely to be superseded for decades."
—L. C. Birch, *Search* Magazine

"*Ecoscience* combines the timely and the timeless, hard science with social science, into a clearly readable discourse. . . . It is a truly awesome achievement."
—Stephen H. Schneider, *National Center for Atmospheric Research*

CONTENTS

15% Discount with this Coupon

Please send me the copies of *Ecoscience* indicated below, at 15% publisher's discount. Full payment enclosed, publisher pays postage and handling. This coupon must accompany order; discount not available through retail bookstores.

______copies, cloth, regularly $39.95, **Now $33.95**
______copies, paper, regularly $19.95, **Now $16.95**

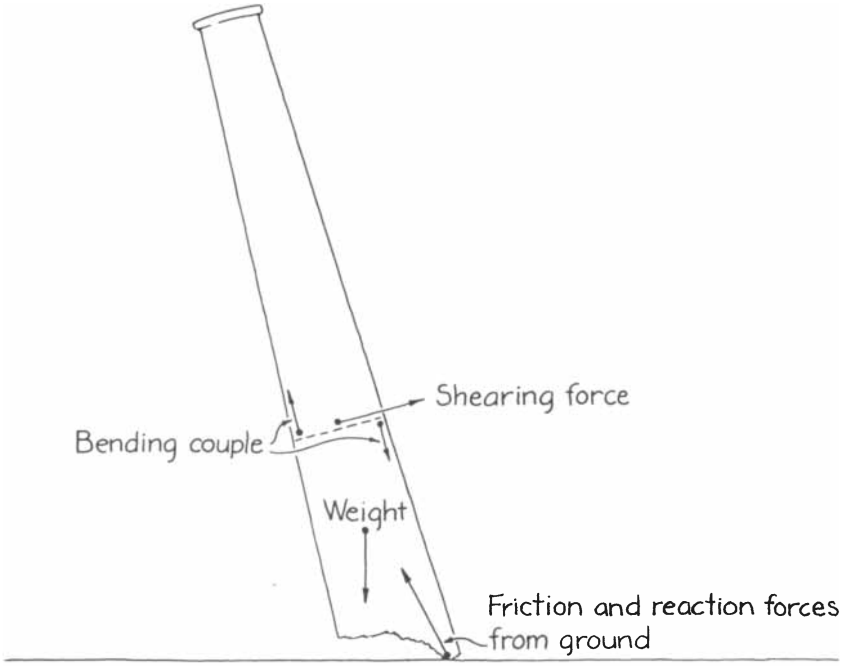
W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, CA 94104

Name ____________________

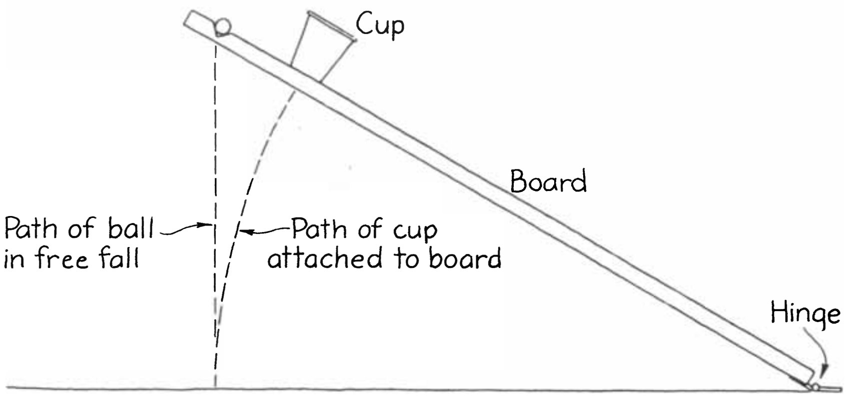
Address ____________________

City/State ________________ Zip ________

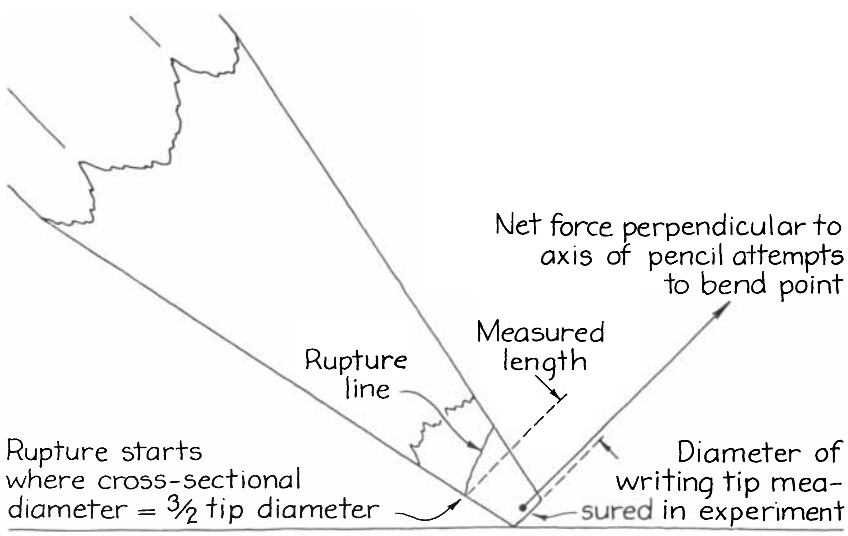
California residents add 6% sales tax; 6½% in BART counties X 2933



Forces on the lower section of a falling chimney



Apparatus to demonstrate an effect of acceleration in falling



Forces that tend to make a pencil point break

could be just a notch in the board. The positioning of these items is such that when the stick is knocked away, the ball separates from the board (because the board accelerates downward faster) and falls into the cup.

One might expect the ball and the board to fall at the same rate, in which case they would reach the tabletop at the same time and the ball would never end up in the paper cup. Once the stick is knocked away the ball is in free fall; it accelerates downward with the normal gravitational acceleration of about 9.8 meters per second per second. The board is subjected to an angular acceleration around the hinge because the weight of the board creates a torque that acts on the center of mass. This motion causes the upper end of the board to accelerate downward at a rate exceeding the acceleration caused by gravity. The rate does not depend on the length of the board (because the center of mass will still lie at the midpoint of any length and the torque there will give rise to the same acceleration at the upper end), but it does depend on the angle between the board and the table. As the board falls and the angle decreases, the downward acceleration of the upper end increases to a value of 1.5 times the normal acceleration. As a result the board and the cup reach the tabletop before the ball does. If everything is positioned correctly at the outset, the ball will drop neatly into the cup.

Albert A. Bartlett of the University of Colorado at Boulder has recently described a modification of this design. He fastened a weight at the uppermost end of the board, just above the notch for the ball. I would have thought that the additional weight would cause the board to accelerate downward even faster, but actually it has just the opposite effect. The additional mass on the board slows the acceleration because the torque cannot accelerate both the board and the additional mass as much as it does the board alone. If enough extra weight is added, the board cannot reach the table ahead of the ball and the ball therefore does not drop into the cup. Under some circumstances the acceleration of the board is so slow that the ball remains in contact with the board at the beginning of the fall and acquires a horizontal velocity before it finally separates from it. The ball then overshoots the cup instead of landing in it.

Now for the pencil. As you are quite aware, the point of a pencil is likely to break if you push down too hard while you are writing. The point cracks because of bending stresses, in much the same way a falling chimney does.

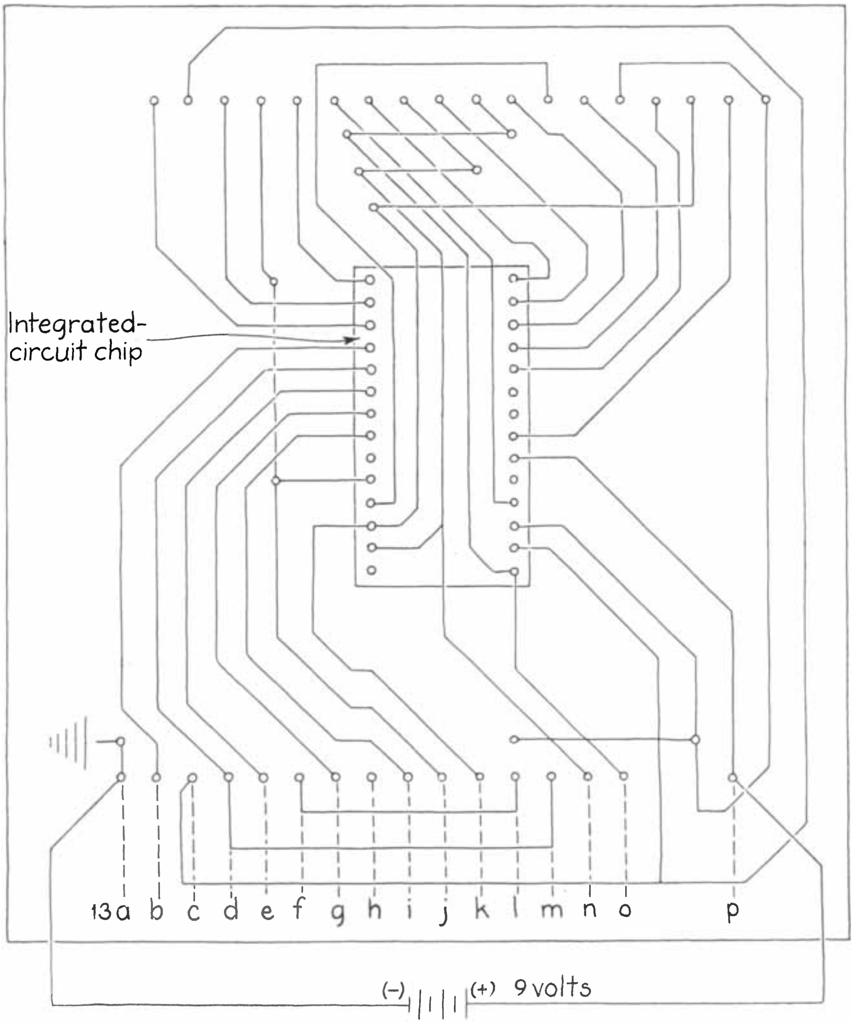
Have you ever noticed that all broken pencil points are about equal in length? Donald H. Cronquist of San Jose has observed this consistency and has devised a simple model to explain it. He first ruled out defects of manufacturing and damage from sharpening as the primary cause of the uniform breaks. Manufacturing defects would be too random from pencil to pencil, and damage from sharpening could be detected by looking closely at the point. It seemed more likely that the point ruptured on its bottom side at a place where the bending stress exceeded the tensile strength of the lead. Tensile strength is a measure of the maximum stress (the force per unit of area) a material can withstand without bending as the stress acts to elongate it. When the stress exceeds that maximum value, the material ruptures. (If, on the other hand, the material is compressed, the stress ultimately exceeds the compressive strength of the material and the material collapses.)

When you write with a pencil, the forces on the tip from the paper and the tabletop set up stresses along the point. In particular a net force perpendicular to the axis of the pencil acts to bend the point, compressing the top side and elongating the bottom side. Stresses arise across a cross section of the point. In a brittle material such as pencil lead the maximum possible stress in elongation (the tensile strength) is usually less than the maximum possible stress in compression (the compressive strength). Hence when a stress limit is reached, it is most likely to be at a place on the bottom side of the point.

When a pencil is sharpened, the tip is normally not made perfectly sharp and so the point does not form a complete cone. A certain length is missing. Cronquist generalized his calculations by measuring the distance from the actual writing tip upward as a ratio of the missing length. His calculations indicate that the bending stress is at a maximum at a place distant from the writing tip by half the missing length. Since the sharpener grinds the point in a conical shape, this result means that the bending stress is at a maximum at a place where the cross-sectional diameter is 1.5 times that of the writing tip (a ratio of 3 to 2). An idealized pencil is most likely to rupture at that place.

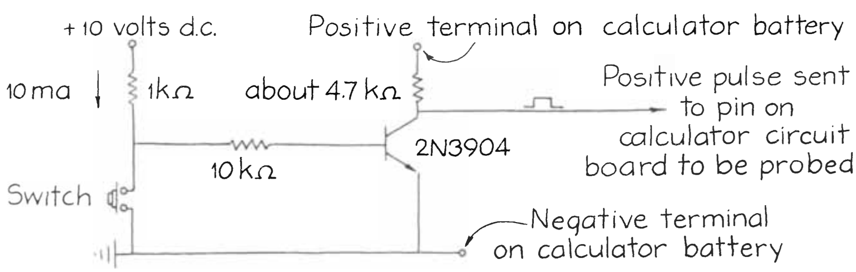
For a blunter tip the broken-off point is longer, because with a larger diameter for the writing tip the place of maximum bending stress is farther up the cone formed by the point. Since the point characteristically breaks soon after the pencil is sharpened, however, the broken-off points will usually be short and of about equal length. The actual force and stress required for breaking will vary from pencil to pencil, depending on the hardness of the lead and the angle of the cone on the pencil point, but the maximum bending stress should still occur at the theoretical place determined by the 3 : 2 ratio of diameters.

The analysis does not extend to severely blunted pencils. In such a pencil



Pins (pairs)	Function	Pins (pairs)	Function
13 e and f	0	13 g and n	9
13 c and d	1	13 a	negative terminal
13 c and e	2	13 p	positive terminal
13 c and g	3	13 e and k	+/− (change sign)
13 d and o	4	13 k and n	clear
13 e and o	5	13 c and i	+
13 g and o	6	13 i and o	−
13 m and n	7	13 g and h	decimal point
13 e and n	8	13 h and i	=

Circuit board and pin connections of the Concept V calculator



Circuitry for probing the pins of a calculator the breaking point would theoretically be up on the wood shaft. Moreover, the 3 : 2 ratio of diameters representing the most likely place of breaking may not exist. The analysis also does not allow for variations that actually are due to manufacturing and sharpening and does not consider the variety of forces on the pencil during use. Finally, the analysis ignores the net force parallel to the axis of the pencil, concentrating instead on the net force perpendicular to the axis. Nevertheless, the formula does predict the lengths of the broken-off points fairly well.

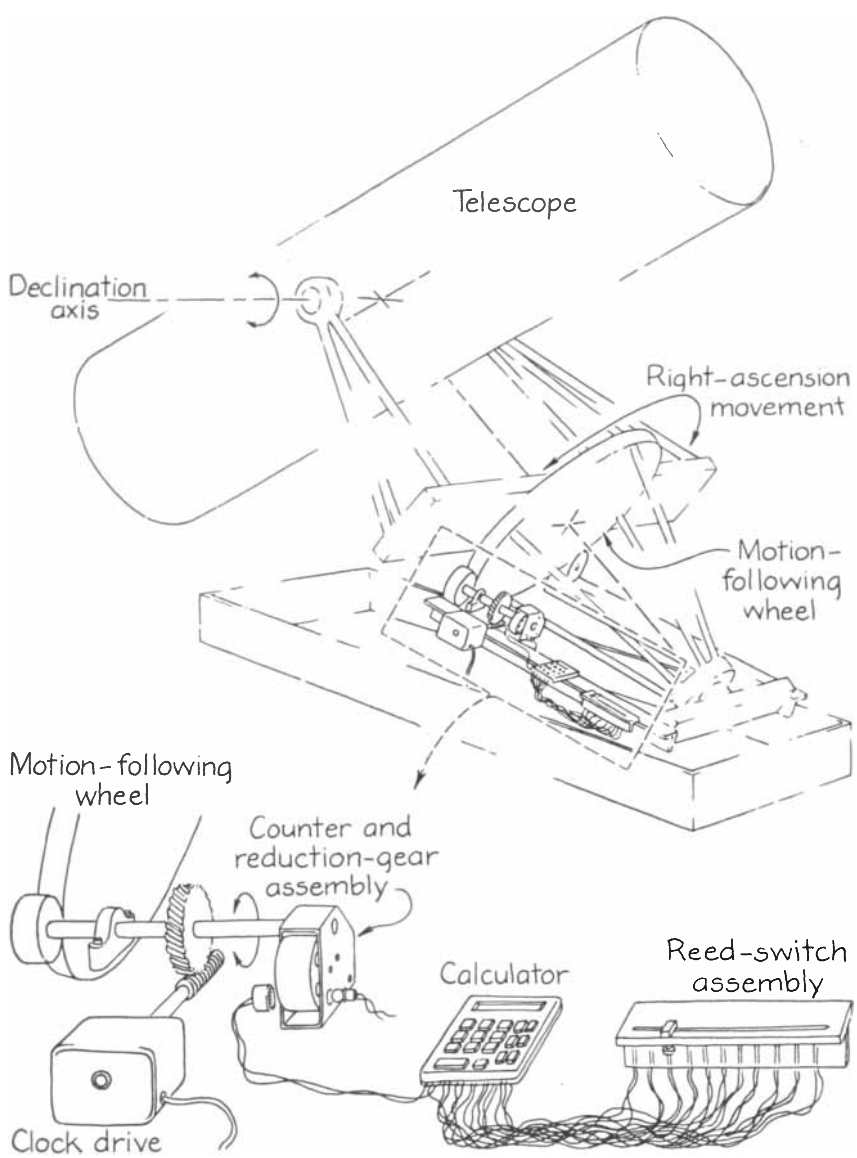
I checked the analysis by systematically sharpening and breaking the points of several No. 2 pencils. The diameters and lengths of the broken-off points (measured from the writing tip to the place where the rupture began on the bottom side) were measured with a micrometer. At first I made no attempt to standardize the breaking. The results were inconclusive because the lengths of the broken-off points varied considerably even when the tip diameters were the same.

Next I tried to make the breaking more consistent. When I sharpened a pencil, I rotated it as I rotated the handle of the sharpener in order to make the grinding uniform around the pencil point. Then I held the pencil so that its tip was at a certain place on the table and my hand was poised just above a rubber stopper I used as a reference height. In this way I kept the angle of the pencil with the tabletop roughly the same (about 45 degrees) throughout. With a pencil so positioned I pressed down until the point snapped off. (Take care that the point does not fly into your eye.) If the rupture occurred up on the wood casing around the lead, I discarded the datum.

The points nearly always broke in the same way. The rupture began on the bottom side of the point and propagated upward and away from the writing tip until it reached the top side. Neither Cronquist nor I know exactly why the rupture propagates in this manner, although the pattern is similar to the rupture patterns of falling chimneys.

I plotted the lengths of the broken-off points against the diameters of the writing tips. To fit a straight line through the data points I used a pocket calculator that could do a linear regression. With both axes on the graph having the same scale, the slope of the line was 2.5. With my sharpener, which produces pencil points whose apex angle is about 12 degrees, the slope of the line should be about 2.2 according to Cronquist's theory. Considering the variation in the construction of pencils the experimental results were surprisingly close to the theoretical ones.

You might continue collecting data on the lengths of broken-off points by investigating pencils of differing hardness and diameter. Is Cronquist's model correct in predicting that neither factor plays a major role in the lengths of the broken-off points? You might also see how the length varies as the angle between the pencil and the table is varied. For any such variation you need to collect a lot of data, plot them and then determine the best fit of a straight line through them. You might find that the theory does not work as well when the pencil tip is quite sharp or quite dull and that something other than a straight line fits the data better.

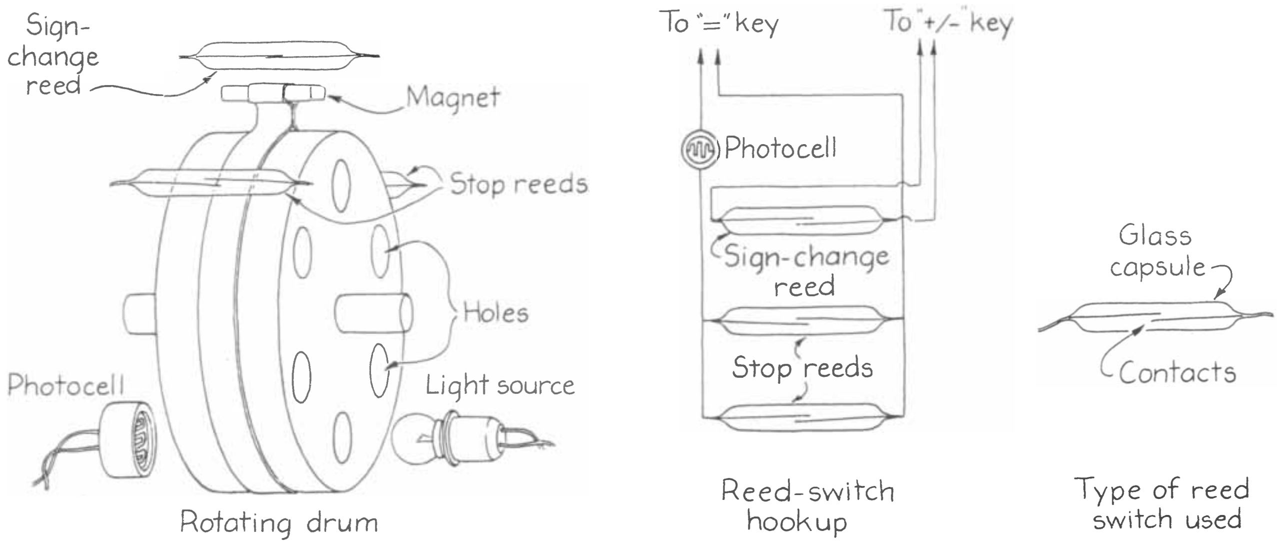


How John and Dave Guerra set up a calculator to record a telescope's lateral motion

Mechanisms for indicating the position of a telescope are sometimes awkward and difficult to read in the dark. John and Dave Guerra of Ludlow, Mass., have designed a position indicator employing the integrated circuit of a pocket calculator. The device is inexpensive, highly accurate and easy to read in the dark.

Not all calculators are suitable. You can find an appropriate one by the following procedure. Turn on the calculator and press first the "plus" key and then the "1." Next press the "equals" key several times. The calculator should add 1 to a running sum each time you push the equals key. Now press the "plus-minus" key once (this key is the one that causes the calculator to change sign) and the equals key again several times. The calculator should subtract 1 from the sum with every push of the equals key. If the calculator does all of this properly, it is suitable for the Guerra design.

The Guerras' idea was to couple the rotating shaft of the telescope to the digital display of the calculator so that the display gives the orientation of the telescope. The coupling is accomplished in several steps. The Guerras' telescope has a polar driveshaft with a gear ratio of 10 : 1. They attached a clock reduction mechanism with a ratio of 120 : 1, thereby providing a net output rotation



Details of the rotating drum

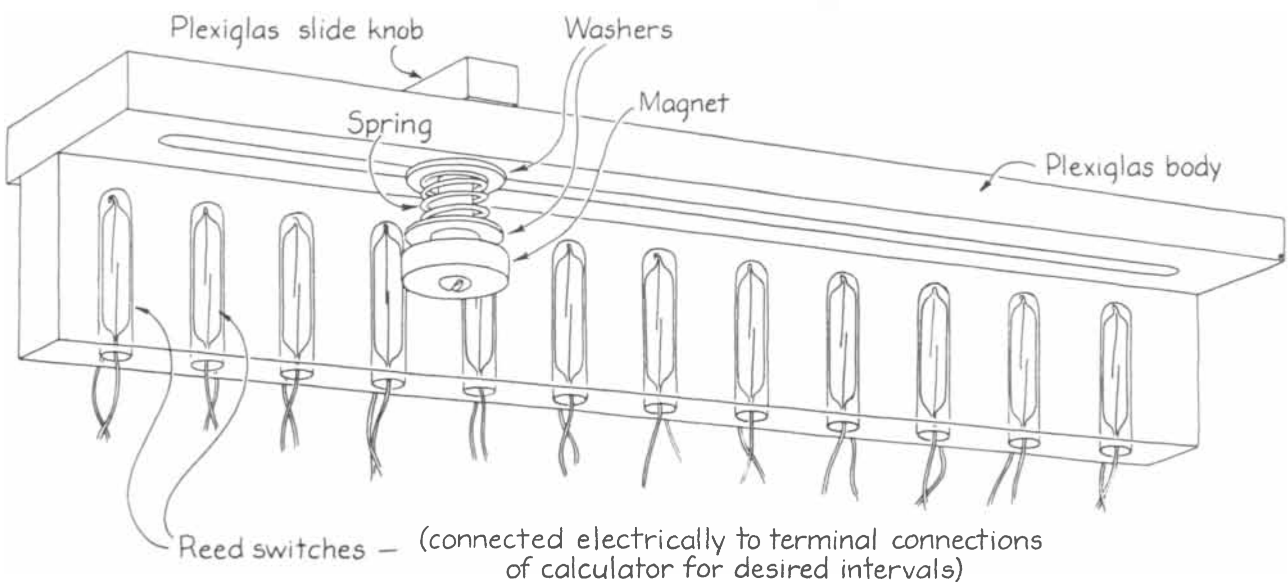
1,200 times the rotation of the telescope. On the final drum of the clockwork they drilled six holes arranged to rotate through a light beam directed on a photocell. Each time the final drum of the clockwork rotated one revolution the beam fell on the photocell six times. One complete polar rotation of the telescope therefore resulted in 7,200 light pulses (6 × 1,200). The photocell was wired to the equals key on the calculator so that each light pulse actuated that button and added another 1 (or any other chosen quantity) to the running sum. The Guerras built two of these devices, one for the right-ascension axis of the telescope and one for the declination axis.

To use the devices the Guerras first orient the telescope toward a star of known position, setting their calculator at the star's coordinates as obtained from a star atlas. Then they turn the telescope about its rotational axes to another star whose position they want to determine. The rotation triggers the photocells, which in turn actuates the calculators. The new coordinates are then read from the displays.

The interval in a calculator (how much the calculator advances with each light pulse on the photocell) can be preset. Suppose you make it .2. Then each pulse increases the display by .2. A full rotation of the telescope, which corresponds to 24 hours, or 1,440 minutes, reads out as 1,440. In other words, the readout is in minutes of time, which would be suitable for right ascension. If one prefers minutes of arc (suitable for declination), a preset quantity of 3 would be appropriate.

The Guerras used a Concept V calculator in their setup. Other brands may suit you better. If you use another calculator, you will have to repeat the Guerras' procedure for determining how the pins of the circuit are coupled to the keys on the keyboard. For the Concept V the Guerras were able to determine by visual inspection how the switches behind the individual keys connected to the leads on the integrated-circuit board and how the leads in turn were connected to the pins on the back of the board.

Visual inspection may be too difficult with some calculators. Then you will have to probe the pins with a device that consists of an inexpensive *n-p-n* transistor, some resistors and a push-button switch. Choose the resistor associated with the switch so that about 10 milliamperes goes through the resistor when the switch is closed. For example, if 10 volts is put through the switch (as in the bottom illustration on page 163), the resistor should be 1,000 ohms. When



Arrangement of the slide switch

the switch is pressed, a signal is sent to the transistor, causing a positive logic signal to be fed from the battery of the calculator to the pin under investigation. If you have a calculator that responds to a negative logic signal, interchange the positions of the push-button switch and its resistor. Some calculators, such as the SR-40 made by Texas Instruments Incorporated, work differently. Instead of having a common ground on one pin, the calculator is designed so that when a key is pressed, it short-circuits two pins to actuate the desired process. With any calculator you should do your probing carefully in order to avoid damaging the instrument.

When the telescope is rotated in the direction opposite to what is considered the positive sense, the calculator has to be told to subtract each pulse it receives from the photocell. The Guerras devised a simple mechanism that triggered the plus-minus key so that the proper operation would be performed. A thin metal band (acting as a friction clutch) around the rotating drum holds a small magnet. Positioned around the magnet are three reed switches. When the telescope is rotated in the positive sense, the magnet rests against one of the reed switches, with the result that each light pulse causes the calculator to add. When the telescope changes direction, friction causes the metal band to rotate with the drum, so that the magnet passes the central reed switch and comes to rest on the third switch. When the magnet passes the central switch, the magnetic field momentarily closes the switch by forcing together the two leads in the switch. During that brief closing the switch sends a pulse to the calculator to switch signs. When the magnet comes to rest again, this time on the third switch, light pulses actuate the photocell and the equals key again, but now the calculator subtracts the preset interval from the preceding sum. The two outer reeds are in electric parallel with each other and together are in electric series with the photocell to prevent erroneous pulse counts during the change of direction.

The Guerras also constructed a turn-on mechanism that takes care of the initial operations of actuating the plus key and presetting the desired interval. It consists of a sliding, spring-loaded bar of Plexiglas fitted with a small magnet. Below the path of the magnet is another sheet of Plexiglas containing several reed switches. Each switch is wired to the calculator keys that would normally have to be pressed to prepare the calculator for use. To prepare the calculator one slides the bar over the reed switches. The last switch, which acts as a stop for the magnet, actuates the penlight cells and a three-volt flashlight bulb that shines on the photocell. If the operator wants instead to use the calculator just as a calculator, the magnet is pulled back short of this last reed switch.

Habitat–the Bottom Line

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

THE SIEVE OF ERATOSTHENES AND THE MÖBIUS STRIP. H. T. McAdams in *The American Mathematical Monthly,* Vol. 55, No. 5, pages 308–309; May, 1948.

TILING WITH INCOMPARABLE RECTANGLES. Edward M. Reingold, Andrew C. C. Yao and Bill Sands in *Journal of Recreational Mathematics,* Vol. 8, No. 2, pages 112–119; 1975–76.

THE VERIFICATION OF THE SALT II AGREEMENT

RECONNAISSANCE AND ARMS CONTROL. Ted Greenwood in *Scientific American,* Vol. 228, No. 2, pages 14–25; February, 1973.

THE VERIFICATION OF ARMS CONTROL AGREEMENTS. George W. Rathjens in *Arms Control Today,* Vol. 7, No. 7/8, pages 1–4; July/August, 1977.

SALT II AND AMERICAN SECURITY. Jan M. Lodal in *Foreign Affairs,* Vol. 57, No. 2, pages 245–268; Winter, 1978–79.

HOW BIRD EGGS BREATHE

DIFFUSION OF GASES ACROSS THE SHELL OF THE HEN'S EGG. O. Douglas Wangensteen, Donald Wilson and Hermann Rahn in *Respiration Physiology,* Vol. 11, pages 16–30; 1970–71.

RESPIRATORY GAS EXCHANGE BY THE AVIAN EMBRYO. O. Douglas Wangensteen and Hermann Rahn in *Respiration Physiology,* Vol. 11, pages 31–45; 1970–71.

RESPIRATORY PROPERTIES OF CHICKEN EMBRYONIC BLOOD DURING DEVELOPMENT. Hiroshi Tazawa, Tomohisa Mikami and Chiyoshi Yoshimoto in *Respiration Physiology,* Vol. 13, pages 160–170; 1971.

THE AVIAN EGG: AIR-CELL GAS TENSION, METABOLISM AND INCUBATION TIME. H. Rahn, C. V. Paganelli and A. Ar in *Respiration Physiology,* Vol. 22, pages 297–309; 1974.

THE AVIAN EGG: INCUBATION TIME AND WATER LOSS. H. Rahn and A. Ar in *The Condor,* Vol. 76, No. 2, pages 147–152; 1974.

LAKE-BOTTOM ARCHAEOLOGY

THE LAKE DWELLINGS OF EUROPE. Robert Munro. Cassell & Co., London, 1890.

LES FOUILLES PRÉHISTORIQUES: TECHNIQUES ET MÉTHODES. André Leroi-Gourhan. A. et J. Picard et cie., 1950.

L'ÉTAT DE LA DENDROCHRONOLOGIE EN EUROPE OCCIDENTALE ET LES RAPPORTS ENTRE DENDROCHRONOLOGIE ET ARCHÉOLOGIE EN SUISSE. Georges Lambert and Christian Orcel in *Archives suisses d'anthropologie générale,* Tome 41, Fascicule 2, pages 73–97; 1977.

SILVICULTURE

THE PRACTICE OF SILVICULTURE. David M. Smith. John Wiley & Sons, Inc., 1962.

FOREST ECOLOGY. Stephen H. Spurr and Burton V. Barnes. John Wiley & Sons, Inc., 1973.

TIMBER: BIOLOGICAL AND ECONOMIC POTENTIAL. Stephen H. Spurr and Henry J. Vaux in *Science,* Vol. 191, No. 4227, pages 752–756; February 20, 1976.

SOLITONS

MAGNETIC MONOPOLES IN UNIFIED GAUGE THEORIES. G. 't Hooft in *Nuclear Physics B,* Vol. B79, No. 2, pages 276–284; September 18, 1974.

PSEUDOPARTICLE SOLUTIONS OF THE YANG-MILLS EQUATIONS. A. A. Belavin, A. M. Polyakov, A. S. Schwartz and Yu. S. Tyupkin in *Physics Letters,* Vol. 59B, No. 1, pages 85–87; October 13, 1975.

FERMION-BOSON PUZZLE IN A GAUGE THEORY. P. Hasenfratz and G. 't Hooft in *Physical Review Letters,* Vol. 36, No. 19, pages 1119–1122; May 10, 1976.

SPIN FROM ISOSPIN IN A GAUGE THEORY. R. Jackiw and C. Rebbi in *Physical Review Letters,* Vol. 36, No. 19, pages 1116–1119; May 10, 1976.

SOLITONS WITH FERMION NUMBER 1/2. R. Jackiw and C. Rebbi in *Physical Review D,* Vol. 13, No. 12, pages 3398–3409; June 15, 1976.

SYMMETRY BREAKING THROUGH BELL-JACKIW ANOMALIES. G. 't Hooft in *Physical Review Letters,* Vol. 37, No. 1, pages 8–11; July 5, 1976.

VACUUM PERIODICITY IN A YANG-MILLS QUANTUM THEORY. R. Jackiw and C. Rebbi in *Physical Review Letters,* Vol. 37, No. 3, pages 172–175; July 19, 1976.

QUANTUM MEANING OF CLASSICAL FIELD THEORY. R. Jackiw in *Reviews of Modern Physics,* Vol. 49, No. 3, pages 681–706; July, 1977.

THE SOURCE OF THE EARTH'S MAGNETIC FIELD

THEORIES OF THE GEOMAGNETIC AND SOLAR DYNAMOS. David Gubbins in *Reviews of Geophysics and Space Physics,* Vol. 12, No. 2, pages 137–154; May, 1974.

THE EARTH'S CORE. J. A. Jacobs. Academic Press, 1975.

LABORATORY SIMULATION OF THERMAL CONVECTION IN ROTATING PLANETS AND STARS. F. H. Busse and C. R. Carrigan in *Science,* Vol. 191, No. 4222, pages 81–83; January 9, 1976.

MAGNETOHYDRODYNAMICS OF THE EARTH'S DYNAMO. F. H. Busse in *Annual Review of Fluid Mechanics,* Vol. 10, pages 435–462; 1978.

GENES THAT VIOLATE MENDEL'S RULES

MECHANISMS OF MEIOTIC DRIVE. S. Zimmering, L. Sandler and B. Nicoletti in *Annual Review of Genetics,* Vol. 4, pages 409–436; 1970.

SEGREGATION DISTORTION AFTER FIFTEEN YEARS. D. L. Hartl and Y. Hiraizumi in *The Genetics of Drosophila melanogaster,* edited by E. Novitski and M. Ashburner. Academic Press, 1976.

DYNAMICS OF SPERMIOGENESIS IN *DROSOPHILA MELANOGASTER,* VII: EFFECTS OF SEGREGATION DISTORTER (*SD*) CHROMOSOME. K. T. Tokuyasu, W. J. Peacock and R. W. Hardy in *Journal of Ultrastructure Research,* Vol. 68, No. 1, pages 96–107; January, 1977.

EXPERIMENTAL POPULATION GENETICS OF MEIOTIC DRIVE SYSTEMS, I: PSEUDO-*Y* CHROMOSOMAL DRIVE AS A MEANS OF ELIMINATING CAGE POPULATIONS OF *DROSOPHILA MELANOGASTER*. Terrence W. Lyttle in *Genetics,* Vol. 86, pages 413–445; June, 1977.

ON THE COMPONENTS OF SEGREGATION DISTORTION IN *DROSOPHILA MELANOGASTER*. Barry Ganetzky in *Genetics,* Vol. 86, pages 321–355; June, 1977.

THE MEDICAL INFLUENCE OF THE STETHOSCOPE

IN DEFENSE OF THE STETHOSCOPE. James B. Herrick in *Annals of Internal Medicine,* Vol. 4, No. 2, pages 113–116; August, 1930.

DISCUSSION ON THE STETHOSCOPE VERSUS X-RAYS. James Maxwell, Peter Kerley and L. G. Blair in *Proceedings of the Royal Society of Medicine,* Vol. 39, pages 355–357; 1946.

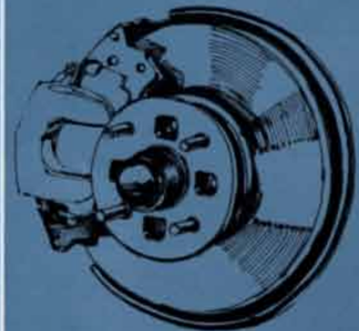
THE AMATEUR SCIENTIST

STRESSES IN FREELY FALLING CHIMNEYS AND COLUMNS. Francis P. Bundy in *Journal of Applied Physics,* Vol. 11, No. 2, pages 112–123; February, 1940.

THEORY OF THE CHIMNEY BREAKING WHILE FALLING. Ernest L. Madsen in *American Journal of Physics,* Vol. 45, No. 2, pages 182–184; February, 1977.

THE NEWEST, NEW CORONA.
THE CAR WITH THE GRAND OPENING.

There's good news and *good* news. First, the good news. The new 1979 Toyota Corona has been redesigned to deliver even more style, comfort and value than you've come to expect of this class car. Now, the *good* news. There's a new sportier Corona with a large rear opening hatch. As you can see by the illustration, this opens up a whole new world of convenience and versatility. The Corona Liftback is the newest of the new. A grand addition to a great car.



The news gets better. The European influenced styling you can see, but wait until you experience the comfort, convenience and Toyota quality. Fully adjustable/reclining bucket seats, tinted glass, and plush pile carpeting. It's all there. What you may not see is the newly redesigned front suspension which gives the '79 Corona a really smooth ride. It's beauty. It's comfort. It's Toyota.

The best news isn't news. The new Corona is a Toyota. Built in a tradition of toughness. Welded unitized body construction helps eliminate rattles and squeaks. Power assisted front disc brakes instill a feeling of confidence. Extensive rust preventative measures help ensure a car that will last. All in all, the new Corona does a lot to further the premise. "If you can find a better built small car than a Toyota...buy it." The new Toyota Corona.

YOU GOT IT.

THE NEW 5-DOOR CORONA **TOYOTA**