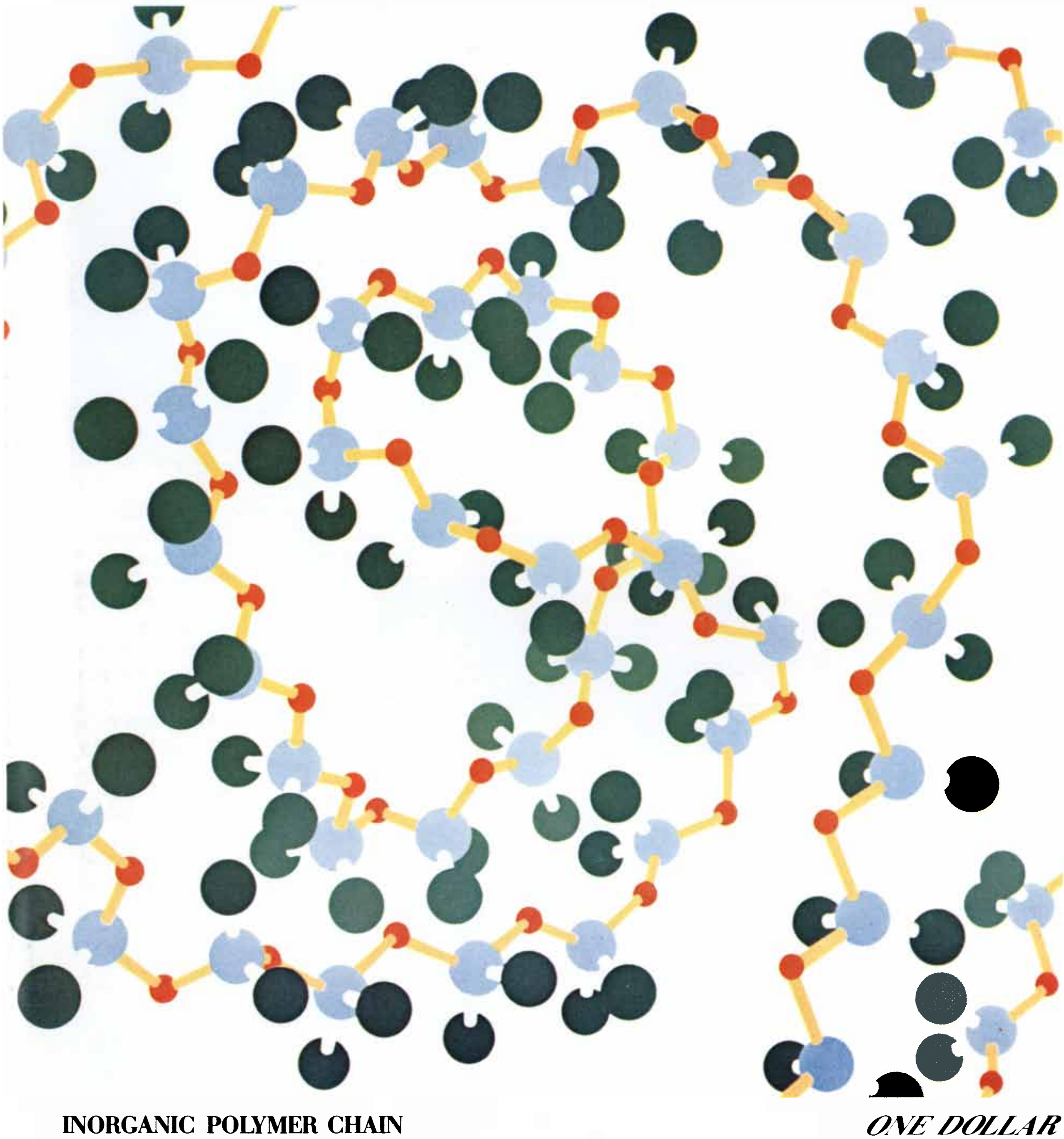
SCIENTIFIC AMERICAN



INORGANIC POLYMER CHAIN

ONE DOLLAR

March 1974

© 1974 SCIENTIFIC AMERICAN, INC

Introducing Plymouth Valiant Brougham.

One of life's little luxuries.
We built the new Valiant Brougham for people who want the comfort of a luxury car, and the economy and size of a compact car.

The Car.
Valiant Brougham is designed to have plenty of room for six passengers. But it's not too big. It's a tough little Plymouth compact with the kind of luxuries you'd never expect to find in a compact.

The luxuries.
Imagine settling comfortably into Valiant Brougham's parchment colored, crushed velour bucket seats with center armrest. (Or, if you prefer, there's a crushed velour bench seat).

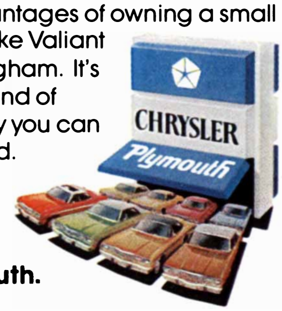
There is an instrument panel with a wood-grained look, complete with radio; there's also cut pile carpeting, map pockets behind the front seat, and assist straps left and right. The total impression is that of many European luxury cars. Standard features like this assure you, Valiant Brougham is not an ordinary compact.


Crushed velour bucket seats.

The engineering.
You'll find such engineering features as torsion-bar suspension. Unibody construction. And an Electronic Ignition System that has no points or condenser to replace. You'll also find such standard equipment as: power steering, disc brakes, and automatic transmission.

The gas mileage.
This is the one time Valiant Brougham doesn't act like a luxury car. You may order it with a standard six cylinder engine. After all, saving gas is stylish, too.

See your Chrysler-Plymouth dealer. He'll tell you about the advantages of owning a small car like Valiant Brougham. It's the kind of luxury you can afford.



Valiant Brougham.
One of eight great small car buys from Plymouth.

Tune in NHL Hockey on NBC-TV. Brought to you by Chrysler Corporation.

Extra care in engineering...it makes a difference.

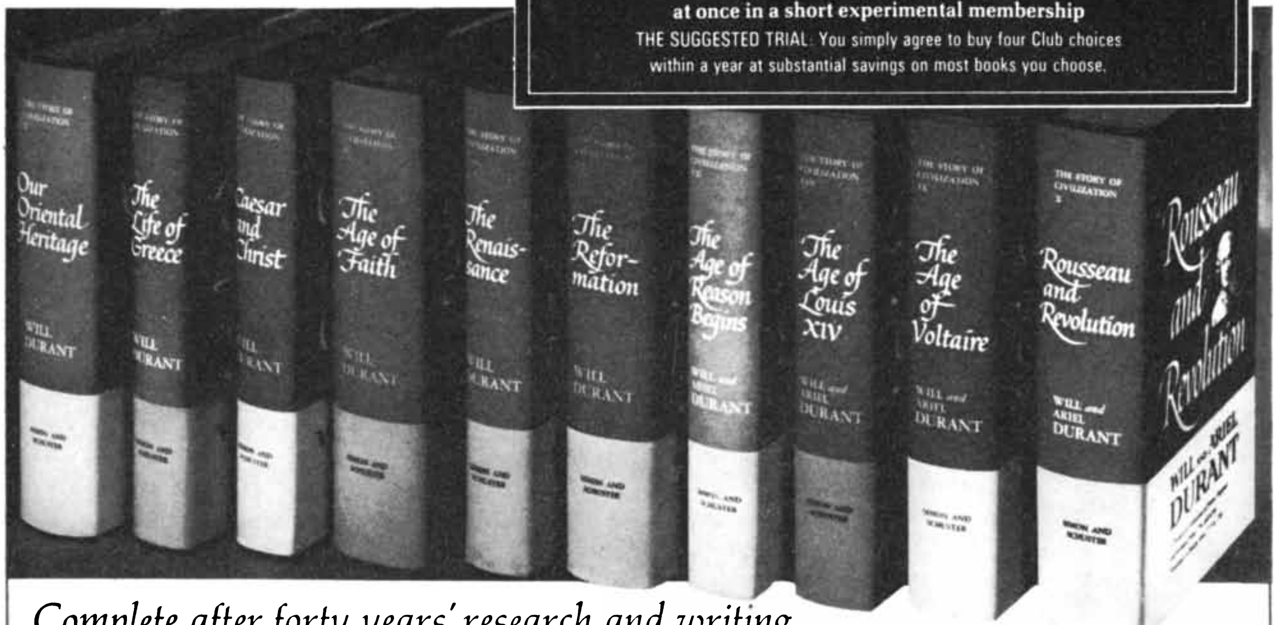
© 1974 SCIENTIFIC AMERICAN, INC

As your introduction to membership in the BOOK-OF-THE-MONTH CLUB®

THIS GREAT $150 TEN-VOLUME HISTORY FOR ONLY $1.50 A VOLUME

at once in a short experimental membership

THE SUGGESTED TRIAL: You simply agree to buy four Club choices within a year at substantial savings on most books you choose.



Complete after forty years' research and writing

The Story of Civilization

by WILL and ARIEL DURANT

I. Our Oriental Heritage. Ancient Egypt and the Near East. The Far East, early to modern times.

II. The Life of Greece. Explores all facets of Greek life from prehistoric times to the Roman conquest.

III. Caesar and Christ. The rise of Rome and Christianity and the collapse of classic civilization.

IV. The Age of Faith. Christian, Islamic and Judaic civilizations, 325 to 1300, including the Crusades.

V. The Renaissance. Italy's golden age, 1304 to 1576. A turbulent world of intrigue and great art.

VI. The Reformation. Europe's religious conflicts, from two centuries before Luther to Calvin.

VII. The Age of Reason Begins. Europe, 1558-1646. The age of Shakespeare, Rembrandt, Galileo.

VIII. The Age of Louis XIV. The brilliant era of the "Sun King," Milton, Cromwell, Peter the Great.

IX. The Age of Voltaire. Europe from 1715 to 1756. The world of Frederick the Great, Wesley, Bach.

X. Rousseau and Revolution. Europe from the Seven Years' War to the storming of the Bastille.

EACH OF THE TEN VOLUMES IN THE SET LISTS FOR $15

THE TRIAL suggested here will prove, *by your own actual experience,* how effectually membership in the Book-of-the-Month Club can **keep you from missing,** through oversight, books you fully intend to read.

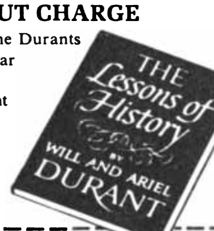
As long as you remain a member, you will be sent approximately every three and a half weeks the Book-of-the-Month Club *News,* a literary magazine published by the Club fifteen times a year, which describes the Selection as well as scores of other books, most of which are available at substantial discounts – up to **40%** on more expensive volumes. All these books **are identical to the publishers' editions in content, format, size and quality.** If you wish to purchase the Selection, do nothing and it will be shipped to you automatically. However, there is no obligation to purchase any particular volume. If you do not want the Selection – or you would like one of the Alternates or no book at all – simply indicate your decision on the reply form always enclosed with the *News* and mail it so we receive it by the date specified. If, because of late mail delivery of the *News,* you should receive a Selection without having had 10 days to decide whether you want it, that Selection may be returned at Club expense.

BOOK-DIVIDENDS: If you continue after this experimental membership, you will earn, *for every Club Selection or Alternate you buy,* at least one Book-Dividend Credit. Each Credit, upon payment of a nominal sum, often only **$1.50** – somewhat more for unusually expensive volumes or sets – will entitle you to a Book-Dividend® which you may choose from over 100 fine library volumes available over the year. This unique library-building system enables members to save **70% or more** of what they would otherwise have to pay for books they want to read and own.

INCLUDED WITHOUT CHARGE

A brilliant summary of all the Durants learned during their forty-year study of world history . . . and an invaluable supplement to *The Story of Civilization.*

A copy will be included, *free,* with each set sent to new members who enroll at this time. **Pub price $5.95**



BOOK-OF-THE-MONTH CLUB, INC.
280 Park Ave., New York, N.Y. 10017 **4-A170-3**

Please enroll me as a member of the Book-of-the-Month Club and send me *The Story of Civilization* in 10 volumes, billing me $1.50 per volume. I agree to purchase at least four additional Selections or Alternates during the first year I am a member, paying in most cases special members' prices. My membership is cancelable any time after I buy these four books. A shipping charge is added to all shipments.

MR. / MRS. / MISS ....
(Please print plainly)

Address 35

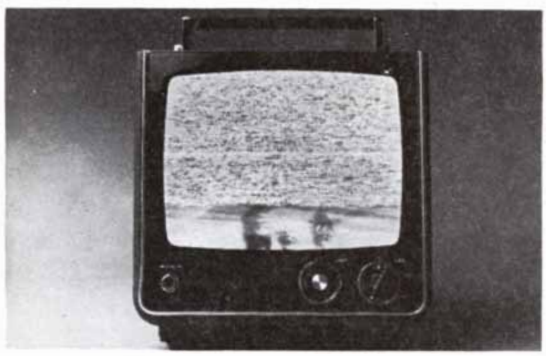
City ....

State .... Zip ....

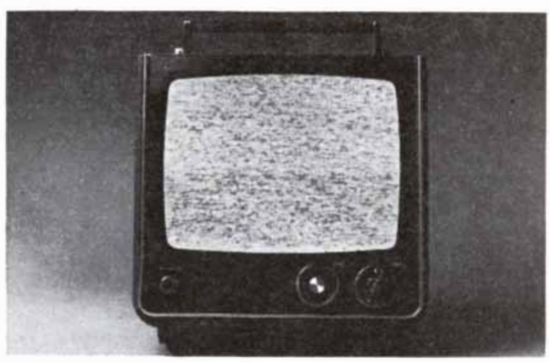
© 1974 SCIENTIFIC AMERICAN, INC

The real
The news media, both broadcast and print, continue to be the center of controversy. Have they gone too far or aren't they going far enough?

The very heart of a democracy is an open free-flow of information. Without it, our system of government cannot exist. A non-informed electorate is no electorate at all. Without freedom to investigate and report the news from every angle, the members of the press become corporate and government spokesmen. We must keep the channels of communication open to all, even when it hurts



The ideal
An open society in which information is passed freely back and forth.
The more we know, the better off we are



AtlanticRichfieldCompany

© 1974 SCIENTIFIC AMERICAN, INC

Established 1845 **SCIENTIFIC AMERICAN** March 1974 Volume 230 Number 3

ARTICLES

DEPARTMENTS

BOARD OF EDITORS Gerard Piel (Publisher), Dennis Flanagan (Editor), Francis Bello (Associate Editor), Philip Morrison (Book Editor), Trudy E. Bell, Brian P. Hayes, Jonathan B. Piel, David Popoff, John Purcell, James T. Rogers, Armand Schwab, Jr., C. L. Stong, Joseph Wisnovsky
ART DEPARTMENT Samuel L. Howard (Art Director), Ilil Arbel, Edward Bell
PRODUCTION DEPARTMENT Richard Sasso (Production Manager), Leo J. Petruzzi and Carol Eisler (Assistant Production Managers), Maureen McKiernan
COPY DEPARTMENT Sally Porter Jenks (Copy Chief), Dorothy Patterson, Candace E. Trunzo, Julio E. Xavier
GENERAL MANAGER Donald H. Miller, Jr.
ADVERTISING DIRECTOR Harry T. Morris
ASSISTANTS TO THE PUBLISHER George S. Conn, Stephen M. Fischer
CIRCULATION MANAGER William H. Yokel

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1974 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. 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: $12 PER YEAR, U.S., POSSESSIONS AND CANADA; $15 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.

© 1974 SCIENTIFIC AMERICAN, INC

The fully-equipped Land-Rover. $4599* Mystique and all.

The 4wd Land-Rover has been climbing mountains, fording rivers, crossing deserts, exploring jungles—and starring in movies—for more than 25 years.

That's why Land-Rover is considered indispensable equipment for every kind of exploration. From a camp-out on the Cape to a summer safari across the Sahara.

The base price of a Land-Rover is $4599. And at $4599, it's fully-equipped. Not stripped-down.

Land-Rover has standard equipment that some others provide as extra-cost options, if at all. Like 4-speed transmission, power-assisted brakes, corrosion resistant all-aluminum body, hard-top, rear door, 3 power take-off points, all-purpose tires and a deluxe interior with fully upholstered seats for 7.

Go test price a Land-Rover today. You'll discover more basics than you bargained for.

For the name of your nearest dealer call: 800-447-4700. In Illinois call: 800-322-4400.

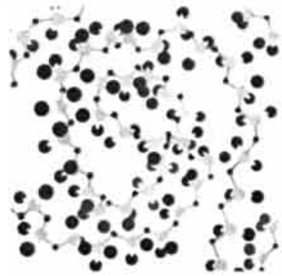
*Manufacturer's suggested retail price. P.O.E. Does *not* include inland transportation, local taxes or preparation charges.

Land-Rover
British Leyland Motors Inc.
Leonia, New Jersey 07605

BRITISH LEYLAND

STANDARD EQUIPMENT: On the new Series III Land-Rover we offer an all synchromesh gearbox (4 forward speeds and reverse, plus a 2-speed transfer box, equals 8 forward, 2 reverse); alternator; larger capacity heater; improved power brakes plus new facia and other styling advances.

As always: Four wheel drive. Full length metal top. Body of corrosion resistant aluminum alloy with sliding side windows. Side-hinged rear door. Seats for seven. Windshield ventilators with built in fly screens. Dual braking system. Windshield washers. Back-up lights. Fresh air heater and defrosters.



THE COVER

The design on the cover represents the long-chain molecule of poly(dichlorophosphazene), an inorganic polymer that is being investigated for its possible utility as an intermediate in the preparation of films, water-soluble plastics and elastomers (see "Inorganic Polymers," page 66). Whereas most commercial polymers have carbon atoms in their backbone, polymers with backbones made up of other atoms promise a new range of properties. The backbone of the polymer on the cover is traced by the yellow sticks. The blue balls in the backbone are phosphorus atoms, the smaller red balls nitrogen atoms. The green balls connected to each phosphorus atom by white sticks are chlorine atoms. This particular inorganic polymer has the drawback of being subject to degradation by moisture. The possibility of overcoming this drawback by the replacement of the chlorine atoms with more stable organic side groups is currently under intensive investigation.

THE ILLUSTRATIONS

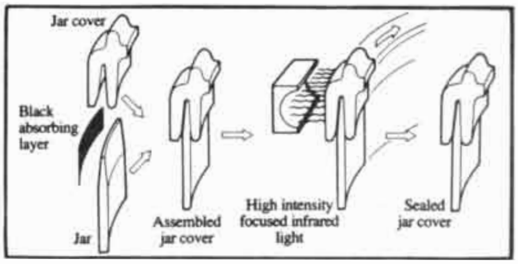
Cover illustration by George V. Kelvin

Page	Source	Page	Source
21–25	Dan Todd	85	Tom Prentiss
26	Roderick A. Capaldi, University of Oregon	86	Tom Prentiss (*top*); Peter F. Ostwald, University of California School of Medicine at San Francisco (*bottom*)
28–31	George V. Kelvin	87	Tom Prentiss
32	Garth Nicolson, Salk Institute for Biological Studies	88	Tom Prentiss (*top*); Peter F. Ostwald and Philip Peltzman, University of California School of Medicine at San Francisco (*bottom*)
33	George V. Kelvin	89	Tom Prentiss
35–42	Bunji Tagawa	90	Peter F. Ostwald and Philip Peltzman, University of California School of Medicine at San Francisco
50	Ames Research Center, National Aeronautics and Space Administration	92–93	Deutsches Museum, Munich
52–59	Ilil Arbel	94–100	Gabor Kiss
65	Ames Research Center, National Aeronautics and Space Administration	103–108	Ilil Arbel
67–73	Enid Kotschnig	111–114	Jerome Kuhl
74	Ben Rose		
77–78	Federal Aviation Administration		
80	Alan D. Iselin		
81	Federal Aviation Administration		
82	Alan D. Iselin		

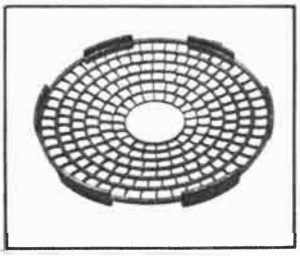
© 1974 SCIENTIFIC AMERICAN, INC

WESTERN ELECTRIC REPORTS

A cutaway view of the new lead-acid battery. For use in the Bell System, four types—each with a different ampere-hour capacity—will replace the 60 configurations currently in use over the same capacity range.



In the sealing process, focused infra-red light is absorbed in a carbon black coating at the jar-cover interface, causing localized melting of the plastic.



The positive grids are designed so that as corrosive growth occurs the space between hoops remains constant. Thus contact with the paste is maintained and electrical capacity actually increases with age as corrosion produces additional lead-dioxide material.

Developing a new lead-acid battery.

Every year, Bell System telephone companies spend over $30 million to buy and maintain the lead-acid batteries they use as intermediate sources of standby power during emergencies.

So they know just how susceptible all lead-acid batteries are to problems caused by corrosion. Problems such as gradual loss of capacity, short- circuits and cracking that could result in acid leaks and occasional fires.

That's why Bell Labs and Western Electric engineers recently undertook the first major improvements on what is essentially a 100-year-old design.

The result: a revolutionary, cylindrical lead-acid battery with a jar and cover fabricated from an improved flame-retardant, impact-resistant polyvinylchloride. The bond between jar and cover is leakproof due to a new infra-red sealing process.

Inside the battery are circular, cone-shaped grids cast of pure lead rather than a lead alloy, then stacked horizontally in a self-supporting structure. Positive grids are cast with large grain-size to minimize corrosion. They're then filled with a paste (tetrabasic lead sulfate) whose rod-like particles interlock for maximum mechanical stability.

These new features required new manufacturing techniques. For example, how could potential suppliers best mass-produce positive plates of the required grain-size and paste the grids rapidly and efficiently, given their conical shape and the new oxide material's crystal structure?

Western Electric's Purchased Product Engineering organization and Bell Labs set up a design capability line at a company subsidiary, Nassau Smelting & Refining.

Using machinery developed at Western Electric's Kearny Works, they refined production methods and materials that made it possible for a supplier to produce the new battery economically, in commercial quantities and to Bell System specifications.

And Western Electric plans to achieve still further savings through a continuing cost-reduction program.

Conclusion: Close cooperation between Bell Labs and Western Electric has resulted in the creation of a superior lead-acid battery. Its expected useful lifetime is at least 30 years—double that of even its best predecessors. It lowers maintenance costs substantially. And its unusual design virtually eliminates the hazard of fire due to mechanical failure.

Western Electric

We make things that bring people closer.

© 1974 SCIENTIFIC AMERICAN, INC

In this instance, the versatile HP 2802A thermometer measures with the repeatability of a hundredth of a degree, the transition temperature of crystal specimens.

Now, you can get a dual range, high resolution digital thermometer for $700.*

For almost 10 years our Model 2801 Digital Quartz Thermometer has been well-known to those scientists whose needs for super-performance (0.0001°C) could justify the price (about $3500).*

Now we offer another great performer, Model 2802A. It is a digital platinum resistance thermometer with 4½ digit display in dual ranges: −200° to +600°C and −100° to +200°C. Within these ranges it does a better job and costs less than other high-performance instruments.

Accuracy is always within 0.5°C ±0.25% of reading (including instruments, probe and display working together) at any ambient from 0° to 55°C.

Resolution is 0.1°C from −200° to +600°C. Then push a button, and you've got extra sensitive 0.01°C resolution in the most commonly used range, −100° to +200°C. It doesn't suffer reference junction woes as do thermocouple types, and its built-in linear analog output lets you record your measurements directly on any voltage recorder. Its solid-state circuitry and display make it rugged enough for field use.

Three interchangeable probes are available: a general-purpose probe, a high-temperature probe and a fast-response probe, as well as a sensor cartridge. Prices range from $105 to $180.*

Also consider its "snap-in" performance options: a rechargeable battery module, a BCD module, even a complete series of modules for voltage, current and resistance measurements.

Write for the brochure on our new, low-cost thermometer.

© 1974 SCIENTIFIC AMERICAN, INC

And a computer system that's more accessible to people.

If your problem of getting to the computer is often times greater than the problem you're trying to solve, then our new HP 3000 can improve your lot.

Here's how: First, it offers access to a number of users, at any time, wherever it's needed (even at your desk). Second, it offers access to the appropriate computer language for the job. Third, it offers access to this capability for $170,000.*

We believe that the HP 3000 offers these features more effectively than any existing computer system, at a cost no other system can match.

Because of its multi-programming capabilities, it provides a needed level of computational power to all users concurrently, whether their application requires batch processing or creative work at an interactive terminal. You get direct, on-line response. No more long waits.

Because it's multi-lingual, it allows each user to employ whatever language is most appropriate to his application, whether in science, engineering, production or finance . . . and even to use several languages in the same application. Four languages are available: FORTRAN, BASIC, COBOL and the new SPL which combines the power of a compiler with the efficiency of an assembler.

To repeat, the price of the HP 3000 starts at $170,000,* about ⅓ less than the cost of comparable systems.

Fill in the coupon for access to our comprehensive HP 3000 Systems brochure.

For more information on all of the above write to us. Hewlett-Packard, 1502 Page Mill Road, Palo Alto, California 94304.

*Domestic USA prices only. 00406

Mail to: Hewlett-Packard, 1502 Page Mill Road, Palo Alto, CA 94304
Please send me information on:
( ) HP 2802A Digital Thermometer
( ) HP 3000 Computer Systems

Name________ Title________
Company________
Address________
City________ State________ Zip________

HEWLETT *hp* PACKARD

Sales and service from 172 offices in 65 countries.
Palo Alto, California 94304

© 1974 SCIENTIFIC AMERICAN, INC

LETTERS

Sirs:

The article "Flywheels," by Richard F. Post and Stephen F. Post [SCIENTIFIC AMERICAN, December, 1973], was intensely interesting, as it is one of the best proposals I've seen that might ease the "energy crunch."

A couple of questions did pop into my mind as I read the article, so I did a spot of research in an attempt to clarify the points. The article's claim that 30 kilowatt-hours of stored energy would suffice for reasonable performance by a small car appeared awfully optimistic. Thirty kilowatt-hours is equivalent to 40.23 horsepower-hours, which is rather low for a 3.33-hour trip at 60 miles per hour.

Using the formula h.p. $= 0.0000087\ AV^3$ (V in m.p.h.) and allowing h.p. $= 12.08$ (available for a given hour) and solving for A, we find that the power available will permit moving a flat plate of 6.43 square feet at 60 m.p.h. Allowing for this figure to be the effective flat-plate area for a well-streamlined small car, it would still be quite small for a space enclosing four passengers. Unfortunately, even if we build the car to the required size we have no power left to haul the payload and the dead weight of the structure. Another 30 kwh just might do the trick; at least it conforms to an old horsepower-required plot I made for a Bonneville streamliner some years back, which indicated that 22 h.p. would produce a speed of 60 m.p.h. in an 1,800-pound vehicle.

My other point is more of a question. How in hell am I going to cram 30 kwh into my flywheel in five minutes? I was hoping that I could simply plug into my household 115-volt circuit with the blessings of the Pacific Gas and Electric Company, but unless I poked a few wrong keys on my number piano, I do not think they'd approve of me trying to suck 3,130 amperes out of their measly service drop even if I went to the trouble of installing heavy bus bars in my garage.

Put in other terms, it takes 482.76 h.p. acting over five minutes to store 40.23 horsepower-hours. That is one hell of a motor to try to fit inside a 600-pound envelope. Seems to me that the last 400-horse electric motor I saw weighed a bit over 3,000 pounds. It was also run at 550 volts—to save weight.

We still have to put in a bit more than we take out, don't we?

JOHN R. MORRIS

Greenfield, Calif.

Sirs:

It is interesting to read of advances in the technique of storing energy in flywheels, but the Messrs. Post have made some highly inaccurate statements in their discussion of flywheel-powered vehicles. For example:

1. The overall efficiency of the flywheel system is compared with the 40 percent of the best steam-electric plants against the fuel-to-rear-wheel efficiency of the automobile under the worst conditions of operation, with the statement that the flywheel system is five times better. This comparison entirely omits the losses in the flywheel-powered vehicle as well as those in the transmission of electric power. The efficiency of the electric-power system overall is about 30 percent to the user [see "The Flow of Energy in an Industrial Society," by Earl Cook; SCIENTIFIC AMERICAN, September, 1971]. The efficiency of the vehicle system, assumed to be 95 percent, certainly cannot be expected to exceed 75 percent. The Electrovair II is the most advanced electric vehicle yet produced (E. A. Rishavy *et al.* in *SAE Transactions*, Volume 76, Paper No. 670175, 1967); it reports an energy loss of 15 percent in the control, motor and transmission system ("The Automobile and Air Pollution," National Technical Information Service of the Department of Commerce Report PB-176–884, page 81, 1967). To this must be added the losses in the motor-generator unit in the flywheel, which occur twice, once in the charging process and again in vehicle operation, giving a total energy loss of 25 percent. The resulting overall efficiency to the rear wheels of 22.5 percent is about twice that of the gasoline automobile under the worst conditions, and it is about equal to the gasoline automobile's efficiency under its most favorable conditions.

2. A car that has a 600-pound power plant will have an unloaded weight of not less than 2,400 pounds, or 3,000 pounds loaded. Such a car will not travel 200 miles at 60 m.p.h. on 30 kwh of energy. About 20 kw will be required at the rear wheels for steady-state level road 60 m.p.h. ("The Automobile and Air Pollution," page 51), needing 25 kw from the energy source. This gives 1¼ hours of operation at 60 m.p.h., or 75 miles. Urban-cycle driving can be expected to give about half this range (*ibid.*, page 53).

Regenerative braking will add nothing to the range at steady speed, since no braking is done, and offers a negligible gain in highway driving. Under no circumstances possible in any normal driving cycle is more than about 10 percent energy recovery potentially possible (*ibid.*, page 85), and the 25-to-50-percent potential gain claimed in the article is clearly not available.

Regenerative braking is difficult to achieve in practice because of the control difficulty of matching the propulsion unit as a source to the source as a receptor, and the propulsion system limits the possible deceleration that can be achieved with regeneration, so that the energy dissipated at the higher levels of deceleration is not fully recoverable. The increased duty-cycle of the propulsion system requires heavier motors and cooling systems, and the control problem adds to the weight and complexity of the control system, so that in practice, regenerative braking is seldom consid-

Scientific American, March, 1974; Vol. 230, No. 3. 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, treasurer; 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 Harry T. Morris, 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 25 cents.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 3578.

☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $12 ☐ 2 years, $22 ☐ 3 years, $30
All other countries:
☐ 1 year, $15 ☐ 2 years, $27 ☐ 3 years, $37

NAME

NEW ADDRESS

OLD ADDRESS

© 1974 SCIENTIFIC AMERICAN, INC

ered to be justified by the small potential gains (Rishavy *et al.*, page 990).

3. To charge the flywheel with 30 kwh "in five minutes from an electrical outlet" would require the internal motor-generator to exert an average power of 483 h.p. and draw about 400 kw from the outlet. If one allows that the motor-generator is capable of 90 h.p., the charging process would take half an hour and draw about 75 kw from the "outlet." Limited to the 10 kw available from the most powerful of today's ordinary domestic outlets, the process would require about 3½ hours.

4. The internal energy of 30 kwh stored in the 130-pound flywheel is about the same as that in 30 pounds of dynamite. The instantaneous release of this amount of energy by a mechanical failure could be expected to have about the same effect as the explosion of the dynamite, no matter how small the particles into which the wheel would disintegrate.

FREDERICK J. HOOVEN

Thayer School of Engineering
Dartmouth College
Hanover, N.H.

Sirs:

Our article, which was intended to stimulate interest in new possibilities in inertial energy storage, was necessarily abbreviated. We have been overwhelmed by the response to the article and are pleased to have this opportunity to amplify some of the points made. The numerical examples given were based either on deductions from published information or on design calculations. Readers' questions were mainly concerned with these quantitative aspects. The answers here, also abbreviated, summarize the design studies; additional material is available on request (c/o *Scientific American*).

1. The projected stored energy requirements for small cars clearly depend on the drag/weight ratios assumed; these vary by as much as a factor of two among existing cars. At a total drag force (aerodynamic, rolling friction, drive-train-gear and bearing losses) of 70 pounds at 60 m.p.h., 30 kwh converted with 0.9 drive-system efficiency (discussed below) yields a range of approximately 200 miles. Examples are found in the literature of small (2,200 pounds test weight) internal-combustion-engine cars with this drag figure (*Road & Track,* March, 1966, page 98); further reductions in drag of at least several percent would be expected with modern radial tires and with the elimination of aerodynamic drag associated with air flow through the cooling system and engine compartment, as is possible in electric cars. Since for cars of the characteristics we have mentioned the internal-combustion-engine drive train (engine transmission, cooling system, exhaust system and gasoline tank) weighs about as much as the projected weight of the inertia-electric drive system, little if any weight-related drag increase would be expected.

2. We believe that achieving a combined flywheel-system–vehicle-drive-motor efficiency of about 0.9 for inertia-electric drive is possible, based on published data on high-performance alternating-current electric-drive motors (General Motors drive; 50–90 horsepower; 0.92 to 0.94 typical efficiency; *IEEE Transactions on Power Apparatus and Systems,* Volume PAS-88, page 86, 1969), together with an average efficiency of 0.99 for solid-state "cycloconverter" controllers (*Machine Design,* June 23, 1966; page 150) and on the projected high efficiency of the flywheel generator-motor (discussed below). For the Environmental Protection Agency's "seven mode" simulated urban-driving cycle, using published drag-v.-speed data and urban gas mileage averages for a particular small car (*Road & Track,* November, 1970; page 98) the calculated engine-transmission conversion efficiency with internal-combustion power is about 10 percent. This result is not in disagreement with the 8.3 percent overall efficiency for a "reference" internal-combustion-engine car calculated by Jalal L. Salihi over the Federal Driving Cycle (*IEEE Transactions on Industry Applications,* Volume IA-9, page 516, September-October, 1973). With an inertia-electric drive using regenerative braking the calculated stored energy requirement for our example car in the seven-mode cycle is 0.123 kwh per mile. (About 45 percent more stored energy would be required for this driving cycle without regenerative braking.) Taking .33 (utility-supplied figure) as the average efficiency for charging energy delivered to customers in major population centers from installed fossil-fuel-fired (coal, residual oil or crude oil) central generating plants, and factoring in 85 percent refinery energy efficiency (Salihi, *op. cit.*), the calculated energy-conservation factor (inverse of the relative primary energy inputs) for inertia-electric v. internal-combustion drive is 5.2 to one; at 0.85 (maximum) crude-oil-to-gasoline

Are you playing your records or ruining them?

If you're like most music listeners, you never think about your records after putting them on your record player.

You just sit back and enjoy the music.

Chances are you'd be less relaxed if you knew that your records might be losing something with every play.

Like the high notes.

When the stylus touches down in the record groove it is violently tossed up, down and sideways, thousands of times a second.

These motions produce either beautiful sounds or expensive memories. It all depends on the tonearm. Which is why so many serious music lovers won't play their precious records on anything but a Dual turntable.

If you would like to know why, mail the coupon. We'll send you complete reprints of independent lab reports. Plus an article from a leading music magazine that tells you what to look for in record playing equipment.

Better yet, ask your franchised United Audio dealer for a Dual demonstration. You'll never have to worry about your records again.

United Audio Products, Inc.
120 So. Columbus Ave.
Mt. Vernon, N.Y. 10553 Dept. SA

Please send me your free literature on turntables, including test report reprints.

Name ______
Address ______
City ______
State ______ Zip ______

United Audio is exclusive U.S. Distribution Agency for Dual

© 1974 SCIENTIFIC AMERICAN, INC

refinery yield the projected crude oil savings would be six to one in favor of inertia-electric drive. In freeway driving, where regenerative braking would be less important, the calculated energy-conservation factor is about 2.8; averaged over U.S. driving patterns (urban and freeway), an energy-conservation factor of about four might be expected. Increases in the energy-conservation factor based on lowered drag-weight coefficients, improved drive-motor efficiencies and the trend toward higher generating-plant efficiencies do not seem out of the question. If we assume, on the other hand, that no such improvements are possible and that drive system efficiencies of only 0.8 are achieved, the calculated energy-conservation factor is still nearly four to one in the seven-mode cycle, certainly a worthwhile gain.

3. Since both the power capacity and the efficiency of electromechanical converters increase as the relative velocity of field and armature increases (exploited in the G.M. drive), both of these factors should be unusually high in a high-speed fiber-composite flywheel system. The proposed inverted permanent-magnet-excited three-phase generator-motor design with centrifugal forces on the field magnets restrained by a fiber composite ring represents one way to exploit this possibility. In a design-study example a 30-kwh system consisting of two 15-kwh flywheel units, each weighing 220 pounds (weight of fused-silica-composite flywheel plus its integral generator-motor assembly; rotation speed 75,000 r.p.m. maximum, down to 25,000 r.p.m. at discharge) constant-current charging in five minutes gives a calculated conductor adiabatic temperature rise of 60 degrees Celsius (600 watts copper losses during charging, corresponding to an averaged energy loss of .34 percent). Peak charging power per flywheel was 270 kw. This could be provided at "charging stations" equipped with flywheel energy-storage units, solid-state frequency converters (for synchronous charging) and charging cables of the gauge used in portable arc welders or in airport auxiliary aircraft power units. Home charging would of course be slower; the power from an electric-dryer-sized outlet would be sufficient for "overnight" (four-to-five-hour) charging. A similar design study was made assuming Kevlar fibers; here the flywheel weights would have to be increased substantially, or else a reduced range accepted. Otherwise the results were similar.

4. Achievable rundown times (with the car parked) are predicted to be several months. In an inertia-electric-drive flywheel of the proposed integral design windage losses and core losses (hysteresis and eddy-current) are expected to dominate. With a low-loss ferrite core and oriented ceramic permanent magnets the above design example gives calculated total no-load core losses of about three watts at full r.p.m., corresponding to rundown times of about 200 days. Calculated windage rundown times are comparably long if the pressure of the chamber-filling gas (hydrogen or helium) is below about one millitorr. A magnetic suspension, backed up by low-friction mechanical bearings, is proposed. J. W. Beams (C. E. Williams and J. W. Beams; *1961 Transactions of the Eighth National Vacuum Symposium: Volume I,* Pergamon Press, page 295, 1962) reports rundown times measured in years with magnetically supported (axis vertical) ultrahigh-speed rotors coasting in vacuo; low-friction suspensions (with the car parked) should also be feasible for inertia-electric flywheels with integral generator-motors.

5. The safe dissipation of flywheel energy in case of mechanical failure or vehicle accident is clearly imperative. In a vehicle inertially stored energy and latent chemical energy (as in gasoline) are not equivalent from a hazard standpoint, either in magnitude or in potential destructive effect. A stored energy of 30 kwh is equal to the energy in 0.8 gallon of gasoline. Unlike gasoline (or dynamite), however, a multishell composite flywheel cannot "explode" in the chemical sense of the word; in failing mechanically the rate of delivery of energy will be limited by the low shear and flexural strength of the composite (in contrast to steel), and by the subdivided nature of the flywheel itself. Furthermore, some of the kinetic energy will be dissipated in fragmenting the composite. In flywheel-rupture tests at Johns Hopkins (with the "superflywheel" geometry) it was found that only a few percent of the inertially stored energy appeared as energy impulsively delivered to the protective housing, indicating that this effect can be pronounced (see bibliography for our article). Thus whereas a stout vented enclosure (made of a fiber composite, of course!), in part composed of structural elements of the car body, would have to be provided, it is our belief that the inertia-electric drive system can be designed to have a substantially lower hazard potential from explosive energy release or fire than gasoline-powered vehicles.

6. Achieving the high stored energy densities we postulate depends on attaining sufficiently high stress levels in the fiber composites. Usable long-term safe operating stress levels for high-strength fibers, particularly those of glass or fused silica, are known to depend greatly on the fiber's history (that is, on its operating environment and on the conditions under which it is drawn and fabricated into fiber composite). Glass and fused-silica fibers are particularly sensitive to the presence of water vapor. In vacuo virgin fibers show maximum strength (up to two million pounds per square inch for fused silica) and little if any strength deterioration with time (B. A. Proctor *et al.* in *Proceedings of the Royal Society,* Series A, Volume 297, page 534, 1967). Carbon fiber and Kevlar are apparently less sensitive to operating environment, with stressed Kevlar fibers projected to drop to about 65 percent of original strength over a 10-year period under normal humidity conditions (T. T. Chiao *et al., Proceedings of Third A.S.T.M. Symposium on Composite Materials: Testing and Design,* March 21–22, 1973). Carbon fibers show an even flatter slope (T. T. Chiao, private communication). The stress-rupture behavior in vacuo is apparently not known for either but is expected to be improved. All such issues having to do with the fabrication of fiber-composite flywheels, however, can admittedly be resolved only through further testing and development.

7. Gimbaling (spin-axis vertical) to minimize gyroscopic effects is proposed, with energy being coupled out of the flywheel unit through flexible electrical leads, as illustrated in our article. This technique was employed in the Swiss "Gyrobus." (See Paper No. 711A, by Robert C. Clerk, S.A.E. International Summer Meeting, June, 1963; this paper also discusses flywheel safety tests.)

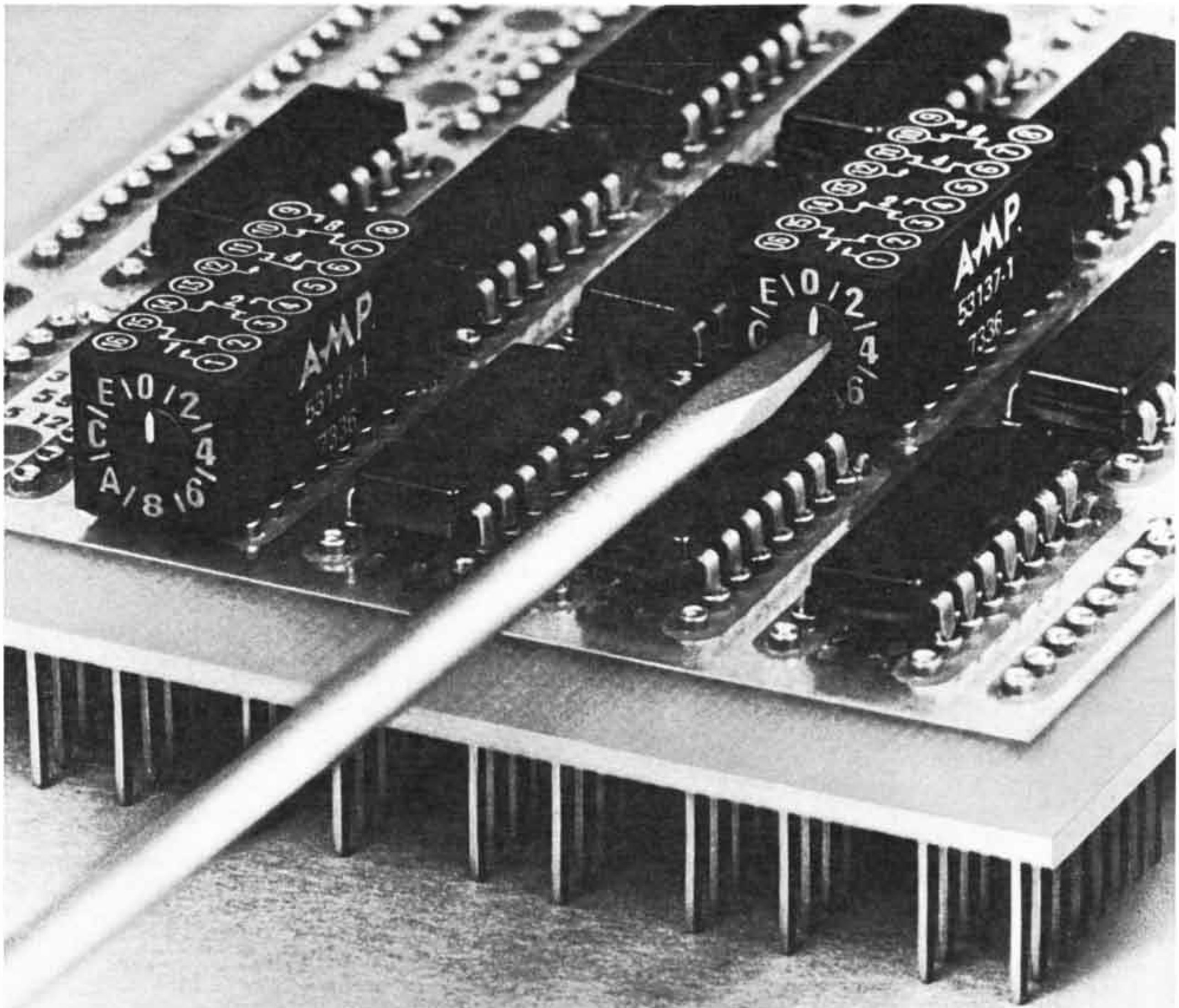
We appreciate the interest in inertia-energy storage shown by the many letters we have received, and we hope that our article will help to stimulate the further necessary development that will lead to flywheel automobiles and stationary energy-storage units. Although our calculations have been checked and reviewed by others, we realize that final answers to some of the questions that have been raised must await such development.

RICHARD F. POST

STEPHEN F. POST

Walnut Creek, Calif.

© 1974 SCIENTIFIC AMERICAN, INC



Introducing our new member of the board.

Meet the pluggable AMP Hexadecimal Rotary Switch, a fast, economical 16-pin DIP type performer. It's made to be programmed manually—on the pc board itself.

Put a screwdriver in the slot. Turn the bi-directional binary encoded cam to the desired, easy-to-read setting. You're now in business.

This AMP innovation, housing four Form "C" switches, is completely compatible with current packaging techniques.

Input/output pins are spaced on .100″ x .300″ centers. So you can mount it by any method normally used for 16-lead ICs.

What about environmental damage to the surfaces of the gold plated phosphor bronze contacts? The fully enclosed design provides all the protection you need.

And what about seeing our board member in action? Just call this number: (717) 564-0101. Or write AMP Incorporated, Harrisburg, Pa. 17105.

AMP is a trademark of AMP Incorporated.

AMP
INCORPORATED

© 1974 SCIENTIFIC AMERICAN, INC

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1924: "The angular diameter of a star has been measured for the first time with an apparatus devised by A. A. Michelson. A 20-foot interferometer constructed at the Mount Wilson Observatory was turned on the star Betelgeuse, and the measurement revealed that this star has a disk one-twentieth of a second of arc in diameter. The distance of Betelgeuse is known roughly, so that we can convert this apparent size into approximate actual size. Betelgeuse is not less than 200 million miles in diameter, and 10 million suns would make one Betelgeuse. This leaves open the question of whether, in order to obtain such a giant star, we should take the material of 10 million suns rolled into one, or whether we should take the material of the sun and inflate it to 10 million times its present size. There is no doubt that the latter answer is nearer the truth. Betelgeuse contains more matter than the sun (perhaps 50 times as much), but in the main its bulk is due to the diffuseness with which this material is spread out. It is a great balloon of low density, much more tenuous than air, whereas in the sun the material is compressed to a density greater than that of water."

"There is every reason to believe that the practice of analyzing metals by means of X rays will soon bring great advances in determining the crystallization and therefore the properties of metals. The microscope has rendered valuable service largely because it enables the form and arrangement of the crystalline grains to be studied. The X ray carries the same form of inquiry into a region 10,000 times more minute, so that it is now possible to see the atoms and the molecules and the way they form crystals. Every crystal has its characteristic X-ray spectrum and can be identified thereby even when the individual crystals are beyond the resolving power of the microscope."

"The fish filet has brought a revolution in marketing. It began experimentally in Boston about two years ago, when haddock filets were wrapped in parchment paper and shipped without ice to New England towns during the cold weather. Summer brought the shipment to an end, for a very large part of the difference between the seaboard and inland prices of fish is freight not on food but on ice and offal. One fish dealer's faith in the idea was so strong, however, that he developed a compact box in which a tin container holding the wrapped filet could be shipped with a moderate quantity of ice. Success was immediate. Inland grocers and butchers found that people who knew ocean fish only by reputation became customers for these fish filets."

"In Reykjavik, Iceland, plans are being made to make use of the neighboring geysers for heating the city. The geysers are situated within a mile of the city and have long been used for laundry purposes. Now it is planned to make use of this never failing source of heat by bringing the water into the city in wooden pipes."

"For the first time in Great Britain an installation of neon lighting is now being applied outside the London Coliseum. The apparatus is one of the most interesting evolved since the introduction of electric lighting. Neon lighting is carried out by means of glass tubes from which the air has been exhausted and replaced by a small quantity of pure neon gas at a pressure very much below ordinary atmospheric pressure. This neon gas is rendered luminescent by the passage of a high-voltage alternating current, the color of the gas being a peculiarly rich flame color."

SCIENTIFIC AMERICAN

MARCH, 1874: "About 800 miles west of Omaha the Union Pacific Railroad crosses the Green River, and the approach to the river is for a considerable distance through a cutting made in rock. During the construction of the road some workmen piled together a few pieces of the excavated rock as protection for a dinner fire and soon observed that the stone itself ignited. The place thereafter became known as Burning Stone Cut. The general superintendent of the railroad, Mr. T. E. Sickels, has caused analyses and experiments to be made with this substance, which proves to be a shale rock rich in mineral oils. The oil can be produced by distillation in abundant quantities, say 35 gallons to the ton of rock. The oil thus obtained is of excellent quality."

"Professor Tyndall, in the course of a recent investigation into the performance of the signals that serve to warn vessels approaching dangerous coasts during foggy weather, has been led to the determination of some important facts regarding the acoustic transparency and opacity of the atmosphere. On one day the distance at which a sound could be heard was 5½ miles, on the next day 10 miles. Professor Tyndall resolved this paradox by analogy to a snowball, in which two transparent substances—ice and air—produce a substance nearly as impervious to light as a really opaque one. A snowball is an aggregate of grains of ice, and the light that falls on the snow is reflected at the limiting surfaces of the granules. Similarly, vapor produced by evaporation streaks and mottles the atmosphere with spaces in which the air is in different degrees saturated. At the limiting surfaces of these spaces, though invisible, we have the conditions necessary to the production of partial echoes and the consequent waste of sound."

"In no branch of manufacture has automatic machinery proved such a thorough success as in the production of watches. Twenty years ago America was supplied with her better class of watches almost wholly by Coventry and Liverpool, the demand for a common article being met by a large importation of movements of Swiss and French make. Today these latter countries still supply the enormous demand for cheap work, but more than 90 percent of the good watches are now of American make. The main portion of the success of the American watch companies must be attributed to their machinery. The American system apportions a machine to the perfection of almost every operation, leaving not more than 10 percent of the work to the skilled workmen."

"M. Naquet has lately been studying the physiological action of hashish. The extract of hemp seed (*Cannabis indica*) administered to various persons produces a great exuberance of ideation; it is not new ideas but the exaggeration, amplification and combination of ideas that pre-existed in the person's mind. Hashish produces one curious effect (which is also observed in acute mania); this is a singular inclination to make puns and plays on words."

© 1974 SCIENTIFIC AMERICAN, INC

The McDonnell Douglas DC-10...

In today's real world of fuel shortages, increased loads, and rising costs, the DC-10 has the built-in efficiencies that airlines must have to move people reliably and economically.

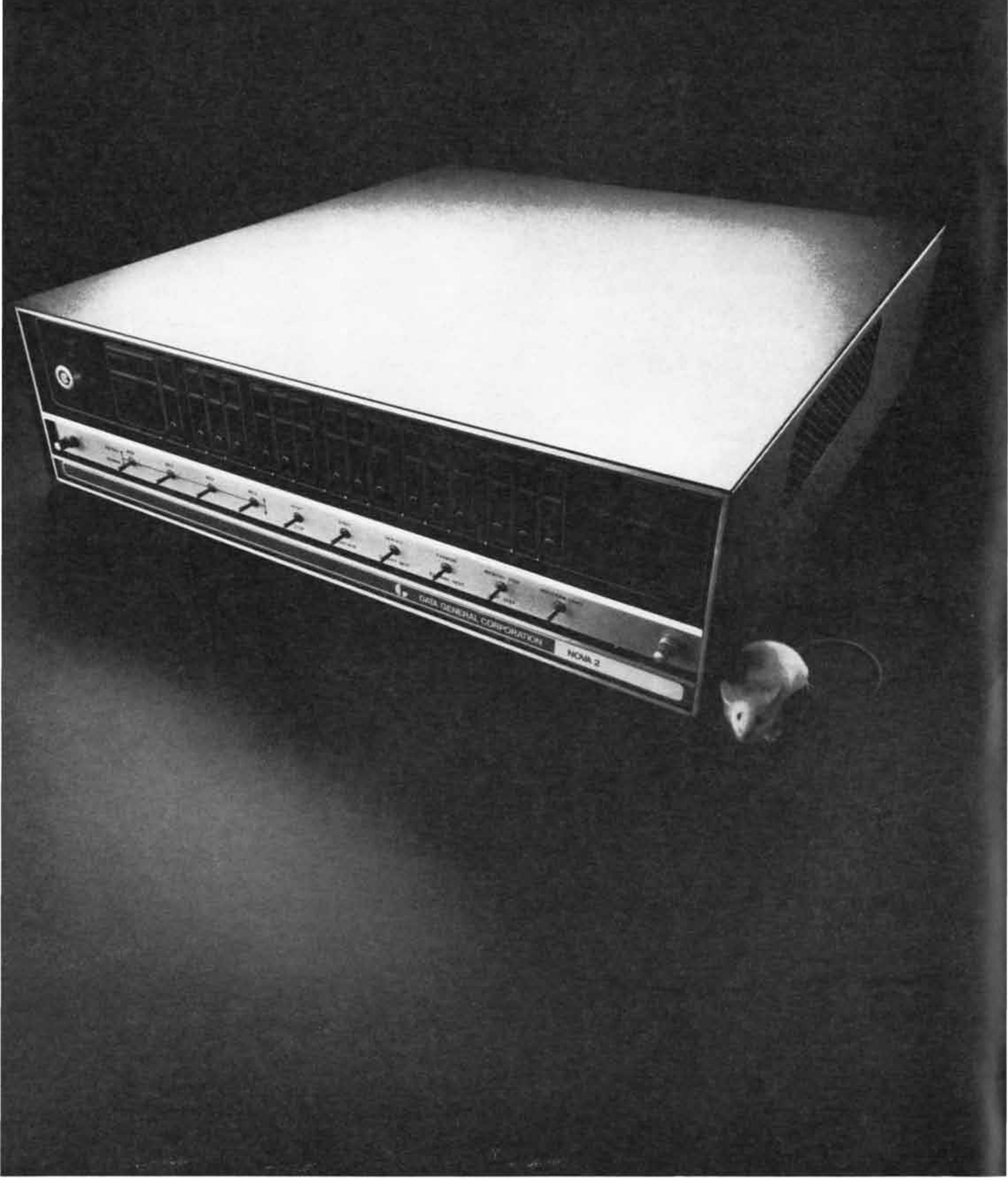
As airlines reduce the number of flights to conserve fuel, the need for the DC-10 becomes increasingly apparent. Its carrying capacity (up to 380 passengers) makes it the logical choice to meet increased density demands on routes from 200 miles to intercontinental ranges.

Compared to the biggest 4-engine jetliners flying typical transcontinental routes, a DC-10 tri-jet can save at least 4,000,000 gallons of fuel in one year, or enough for 1,000 flights between New York and Chicago in a DC-10.

The DC-10's advanced-technology engines leave no smoke in the sky. And in actual operating environment tests at airports around the world, it proved it is the quietest of the big jets.

The DC-10 is just one good example of how McDonnell Douglas is helping to meet man's challenges—in transportation, national defense, space, communications, and other fields where our aerospace technology pays dividends for everybody, year after year after year. It's all in our annual report. For a copy, write: Box 14526, St. Louis, MO 63178.

© 1974 SCIENTIFIC AMERICAN, INC



You don't have to be afraid of a computer that's smaller than you are.

© 1974 SCIENTIFIC AMERICAN, INC

That little thing next to the mouse is a computer. And it doesn't bite.

It costs a lot less than a big computer. So instead of putting it up on a pedestal, you can afford to put it to work.

You can move it to where the work is. On the production line. In the lab. Classrooms. Offices.

It's tough enough to work where it's too cold for people to work. Or too hot.

And it can tolerate double the voltage reduction of a big computer. So it's less likely to blow out in case of a brownout.

This computer is one of many made by Data General. The fastest growing maker of small computers in the world today.

Write for our brochure: "The Sensible Way To Look At Computers."

It may be just the thing you need to scare you into buying one of ours.

DataGeneral

The computer company you can understand.

Data General Corporation, Dept. C, Southboro, Massachusetts 01772, (617) 485-9100.
Datagen of Canada Ltd., Hull, Quebec (819) 770-2030/Data General Europe, Paris, France 504-23-44

© 1974 SCIENTIFIC AMERICAN, INC

THE AUTHORS

HARRY PERRY ("The Gasification of Coal") is a consultant to National Economic Research Associates, Inc. From 1940 to 1967, except for three and a half years of naval service during World War II, he was with the U.S. Bureau of Mines, serving during much of the period as director of coal research. From 1967 to 1970 he was research adviser to the Assistant Secretary of the Interior for mineral resources. After a brief consulting assignment with the Organization for Economic Cooperation and Development (OECD) he served as senior specialist for energy in the environmental policy division of the Congressional Research Service of the Library of Congress, taking up his present work in 1973. Perry obtained his bachelor's degree in chemical engineering at the University of Pennsylvania in 1940 and his master's degree at the University of Pittsburgh in 1941. His experience has been in the mining, combustion, carbonization and gasification of coal and also in work on other fuels, the abatement of air pollution and the large-scale production of helium.

RODERICK A. CAPALDI ("A Dynamic Model of Cell Membranes") is at the University of Oregon, where he serves as assistant professor of biology and as a research associate of the Institute of Molecular Biology. Born in England, he took a degree in botany and zoology at the University of London and then obtained his Ph.D. (in chemistry) from the University of York. From 1970 to 1973 he was associated with the Institute for Enzyme Research at the University of Wisconsin. He became interested in membrane structure while working on the structure and function of insulin. "It seemed that an understanding of the action of this hormone," he writes, "was severely restricted by a lack of knowledge of membrane structure. Eventually I hope to take up the problem of insulin action again." Capaldi describes himself as "an avid football fan" and notes that he plays squash in the winter and golf in the summer.

JÖRG-PETER EWERT ("The Neural Basis of Visually Guided Behavior") is professor of zoology and head of the department of neuroethology at the newly founded Gesamthochschule Kassel (Integrated University of Kassel) in West Germany. "I received my Ph.D. at the University of Göttingen in 1965," he writes, "and my Habilitation degree [qualifying him for a professorial appointment] in 1969 at the Technical University of Darmstadt. My major interests apart from the kind of work described in my article are the arts; on leaving high school my aim at first was to be a painter."

JOHN S. LEWIS ("The Chemistry of the Solar System") is associate professor of chemistry and geochemistry at the Massachusetts Institute of Technology, where he is affiliated with both the department of chemistry and the department of earth and planetary sciences. He received his bachelor's degree at Princeton University in 1962, his master's degree at Dartmouth College in 1964 and his Ph.D. from the University of California at San Diego in 1968. He lists five interests apart from his work. "First, I write: essays, science fiction and so on. Second, I am involved in research on the physiological effects of transcendental meditation and am myself a practitioner of transcendental meditation. Third, I am a dedicated Frisbee player; my amateur status has been challenged several times. Fourth, I am a beachcomber and skin diver at heart, and the whole family loves camping, usually in the most northerly parts of Canada that we can reach. Fifth, I am happy, optimistic about the future and have at least my share of utopian goals."

HARRY R. ALLCOCK ("Inorganic Polymers") is professor of chemistry at Pennsylvania State University. Born in England, he received his degrees of bachelor of science and doctor of philosophy from the University of London. He has held postdoctoral appointments at Purdue University and at the National Research Council of Canada. He also spent a number of years as a research chemist in industry. His current research interests, he writes, are in inorganic chemistry and polymer chemistry and areas where those two disciplines overlap. Allcock notes that he also has "a continuing interest in biomedical research."

NORMAN A. CHIGIER ("Vortexes in Aircraft Wakes") is reader in mechanical engineering and fuel technology at the University of Sheffield. He obtained his master's degree and his Ph.D. at the University of Cambridge. During a leave of absence from Sheffield in 1970 and 1971 he worked as a senior research associate at the Ames Research Center of the National Aeronautics and Space Administration. He has returned there on a number of occasions to conduct wind-tunnel experiments. Chigier's published writings have been mainly in the fields of fluid dynamics and combustion. He is coauthor of the book *Combustion Aerodynamics.*

PETER F. OSTWALD and PHILIP PELTZMAN ("The Cry of the Human Infant") are at the University of California School of Medicine in San Francisco; Ostwald is professor of psychiatry and Peltzman is assistant professor in residence. Ostwald writes: "I've always been very interested in sounds and how people use them to communicate. In my research I have focused on the different speech patterns, voice qualities and nonverbal signals associated with emotional behavior in health and disease. Another interest is how the basic needs for contact, food, love and pleasure are expressed in early childhood. I love music, play the violin regularly in a string quartet and enjoy working with musicians, linguists, acoustical engineers and others who are involved in the world of sound." Peltzman received his bachelor's and master's degrees at Brooklyn College of the City University of New York in 1956 and 1960 respectively and his Ph.D. from Purdue University in 1967. He went to the University of California School of Medicine in San Francisco as a research associate in 1968 after a year as a postdoctoral fellow at the Stanford University School of Medicine. His work now is in audiology and speech, medical psychology and obstetrics and gynecology.

GEORGE SHIERS ("Ferdinand Braun and the Cathode Ray Tube") is a free-lance technical writer, a consultant and the author of several textbooks on electronics. He teaches courses in scientific and technical writing and in adult-education training at the University of California extension at Santa Barbara. In addition he is writing a book on the early history of television. He is a senior member of the Institute of Electrical and Electronics Engineers and of the Society for Technical Communication. Two other articles by Shiers have appeared in SCIENTIFIC AMERICAN: "The First Electron Tube" (March, 1969) and "The Induction Coil" (May, 1971). In connection with the present article he wishes to acknowledge the assistance of Linda Gaede and Walter Knapp in translating German material.

© 1974 SCIENTIFIC AMERICAN, INC

The finest stereo receiver the world has ever known.

We recognize the awesome responsibility of making such a statement. Nevertheless, as the leader in high fidelity, we have fulfilled this responsibility in every way.

Pioneer's new SX-1010 AM-FM stereo receiver eclipses any unit that has come before it. It has an unprecedented power output of 100+100 watts RMS (8 ohms, both channels driven) at incredibly low 0.1% distortion, throughout the entire audible spectrum from 20 Hz to 20,000 Hz. Power is maintained smoothly and continuously with direct-coupled circuitry driven by dual power supplies.

To bring in stations effortlessly, clearly and with maximum channel separation, the SX-1010 incorporates an FM tuner section with overwhelming capabilities. The combination of MOS FETs, ceramic filters and phase lock loop IC circuitry produces remarkable specifications like 90dB selectivity, 1.7uV sensitivity and 1 dB capture ratio.

Versatility is the hallmark of every Pioneer component. The SX-1010 accommodates 2 turntables, 2 tape decks, 2 headsets, 3 pairs of speakers, a stereo mic and an auxiliary. It also has Dolby and 4-channel adaptors. There's even tape-to-tape duplication while listening simultaneously to another program source. This is another innovative Pioneer exclusive.

The SX-1010 is actually a master control system with its fantastic array of controls and features. It includes pushbuttons that simplify function selection and make them easy to see with illuminated readouts on the super wide tuning dial. FM and audio muting, hi/low filters, dual tuning meters, loudness contour, a dial dimmer control and a fail-safe speaker protector circuit. Never before used on a receiver are the twin stepped bass and treble tone controls that custom tailor listening to more than 3,000 variations. A tone defeat switch provides flat response instantly throughout the audio spectrum.

By now it's evident why the SX-1010 is the finest stereo receiver the world has ever known. Visit your Pioneer dealer and audition its uniqueness. $699.95, including a walnut cabinet.

U.S. Pioneer Electronics Corp., 75 Oxford Drive, Moonachie, New Jersey 07074.
West: 13300 S. Estrella, Los Angeles 90248 / Midwest: 1500 Greenleaf, Elk Grove Village, Ill. 60007 / Canada: S. H. Parker Co.

PIONEER®
when you want something better

© 1974 SCIENTIFIC AMERICAN, INC

Anatomy of a Nikon Camera (Part II)

Recently, we discussed the precision with which a Nikon is put together to make it the superb picture-taking instrument it is. But there is still another aspect of the way Nikon is put together that makes it the most versatile camera ever conceived.

We're now talking about the literally thousands of ways you can put a Nikon together yourself by way of its totally flexible, modular construction, on which the Nikon system is based.

Interchangeable Lenses

There are more than forty lenses in the Nikon system. Fourteen of them are wide angle lenses, for use at close quarters or to exaggerate perspective, including a 220° Fisheye Nikkor that actually "sees" objects behind the camera. To bring distant objects up close, you can choose any of fifteen telephotos, ranging up to 2000mm and giving up to 40X magnification compared with the "normal" 50mm lens. There are high-speed lenses for use in the dimmest available light and special lenses for extreme closeup work—even one that corrects linear distortion in architectural photos. You'll find all of these and more in the Nikon system—all sharing the exceptional quality, the surpassing sharpness and the magnificent color rendition that make Nikkor lenses the overwhelming choice of professionals.

Interchangeable Finders

Because a single lens reflex camera is built around the finder, Nikon pays particular attention to this aspect. It starts with the only 100% accurate viewing system in all of 35mm. That's right; the Nikon F and F2 are the only cameras whose finders show precisely the picture area appearing on the film—no more and no less. To implement this unique accuracy, there are five interchangeable finders for the Nikon F and six for the Nikon F2. Brilliant eyelevel prism finders. Photomic thru-the-lens meter/finder

systems, waist-level and magnifying finders — even one for viewing accurately while wearing goggles or with the camera in an underwater housing.

Interchangeable Finder Screens

The viewing screen is where the finder image is formed by the camera lens and the finder's optical system. In most cameras it's fixed; a few others accept a handful of interchangeable screens. There are no fewer than eighteen in the Nikon system, each offering practical advantages for some specific application. Even if you never use more than one, you can choose among seven universal types to suit your individual preference.

Interchangeable Backs

This feature is mainly for the professionals, who use more Nikons than any other 35mm camera. If you happen to be one, you may want to replace the Nikon back with one holding 250 (or even 800) exposure film loads for use with a motor drive.

Or, you might attach a Speed Magny back and shoot large-format Polaroid pictures (that's right, with a Nikon!).

Using just the parts discussed so far, you can come up with a staggering 21,600 different combinations for your Nikon camera. And that doesn't include closeup equipment, filters and dozens of other eminently practical accessories presently in the Nikon system, and new ones constantly being developed.

Of course, no one photographer will ever need all of it. The point is, no matter how exotic or far-fetched the situation may be, you're sure to find the equipment you need in the Nikon system.

That's versatility, Nikon style. To find out more about it and how it can work for you, visit your Nikon dealer. Ask him also about the Nikon School of Photography. Or write for detailed literature folio 19. Nikon Inc., Dept. SA, Garden City, N.Y. 11530. Subsidiary of Ehrenreich Photo-Optical Industries, Inc. EPOI
(Canada: Anglophoto Ltd., Montreal 376, P.Q.)

© 1974 SCIENTIFIC AMERICAN, INC

Established 1845 SCIENTIFIC AMERICAN March 1974 Volume 230 Number 3

The Gasification of Coal

This formerly widespread technology, which lost its markets to natural gas and petroleum, is now being reexamined. New methods promise an alternative source of fossil-fuel energy

by Harry Perry

Nearly every major city in the eastern U.S. once had its gashouse, where gas was manufactured (usually from coal) for lighting and cooking. The gashouses and the bulky cylindrical storage tanks that stood near them gradually disappeared after World War II as natural gas came to be distributed nationally by pipeline. Now, with both natural gas and petroleum increasingly in short supply, the idea of improving the nation's energy situation by the gasification of coal, which is plentiful, is attracting much interest.

Coal is found in 30 of the 50 states and represents more than 90 percent of the proved reserves of all developed fuels. At the 1972 rate of consumption the proved coal reserves would not be exhausted for some 600 years. Even if coal were the sole source of energy and the total demand for energy rose at the rate of 3.5 percent per year, the proved reserves would last for 47 years and total coal reserves for nearly 75 years.

Natural gas gained its ascendancy because its cost is low and its heating value is high. Like other gas it is clean, easy to distribute and convenient to use. In the years after World War II it therefore became the preferred residential and commercial fuel and also found many industrial applications. For these reasons the consumption of natural gas grew at a rate of 5.4 percent per year between 1947 and 1971, compared with a rate of 3.1 percent per year for the total energy consumption.

Gas made from coal could substitute for natural gas, although the old technology of coal gasification would have to be improved in order to give the product the heating value of natural gas (about 1,030 British thermal units per cubic foot). In addition the unenriched gas made from coal (with heating values of from 125 to 560 B.t.u. per cubic foot) would serve well in a number of industrial applications. It is therefore not surprising that the gasification of coal is being reexamined as a major source of clean energy from domestic resources.

Gas manufactured from coal was first produced in the late 18th century by heating coal in the absence of air. (It was heated by burning part of the coal outside a vessel containing the remainder of the coal; if the coal inside the vessel had been allowed to come in contact with air, it would have burned, forming combustion-product gases rather than gas that could itself be burned later.) The first coal-gas company, which distributed its product for lighting, was chartered in London in 1812; the first U.S. company was chartered in Baltimore in 1816.

In those days the gas was being produced by destructive distillation, that is, by heating the coal to a temperature where it decomposed chemically. The gas produced by the distillation of coal has a heating value of from 475 to 560 B.t.u. per cubic foot, depending on the type of coal and the temperature to which it is heated. A similar type of gas is produced when coal is carbonized to manufacture coke; frequently coke-oven gas supplemented coal gas when it was locally available and coal gas was in short supply.

In these processes 70 percent or more of the original coal remains as a solid residue, consisting mostly of carbon, that must be sold or otherwise utilized. With coke-oven gas this is no problem, since coke is the major product. With coal the answer lies in going a step beyond distillation to gasification.

Gasification involves not only heating the coal, as in distillation, but also the subsequent reaction of the solid residue with air, oxygen, steam or various mixtures of them. The distillation step releases a certain amount of gas that has a fairly high B.t.u. content because methane (CH_4) and other higher hydrocarbons contained in the coal are among the first components to emerge as the coal decomposes. The gasification step produces a gas that is essentially a mixture of hydrogen and carbon monoxide (which has a much lower heating value than the distillation gases do) with some of the gases distilled from the coal. The amount of the distilled gas in the final product varies with the gasification process.

Oxygen is expensive, and so the old gas companies resorted to an alternative process to make a gas that supplemented coal gas and coke-oven gas. The process produced what was known as "blue gas"

© 1974 SCIENTIFIC AMERICAN, INC

or "water gas." Anthracite coal from which the fine sizes had been removed, with the large sizes being crushed to an acceptable dimension, was fed to a vessel lined with a refractory material. After the gasifier had been brought up to a suitable temperature the coal (it could also be coke) was blasted with a stream of air and heated further by combustion with oxygen in the air. During the first part of this "blow" period the gas produced was mostly carbon dioxide ($C + O_2 \rightarrow CO_2$), but as the coal or coke in the upper level of the bed grew hotter the reverse reaction ($CO_2 + C \rightarrow CO$) began to go, meaning that carbon monoxide was being produced in increasing amounts. After enough coal or coke had been heated sufficiently the valves controlling the flow of air were closed and other valves were opened to begin the "run" period, in which the hot coal or coke was reacted with steam ($C + H_2O \rightarrow H_2 + CO$). It was a cyclical process, leading to complications in the design and operation of the equipment.

The water gas had a heating value of about 300 B.t.u. per cubic foot compared with the value of from 475 to 560 of coal gas, so that it had to be enriched. The enrichment entailed another cyclical process. The gas made at the end of a run period was hot and contained enough carbon monoxide so that it could be burned to heat refractory bricks enclosed in a "carburettor" vessel. When the bricks were hot enough, the heating gas was shut off and oil was sprayed on the bricks, where the heat caused the oil to be cracked into lighter hydrocarbon products, including methane and propane. The resulting gas, which had a high heating value, was mixed with the gas made during the run period to produce what was known as "carburetted blue gas."

The companies originally sold a coal gas with a heating value of about 550 B.t.u. per cubic foot. Burners were therefore designed to handle a gas of that quality. With burner designs fixed, manufactured gas, no matter how it was produced, was usually adjusted to that heating value. On the introduction of natural gas it became necessary in some areas to produce a gas that could be mixed with natural gas. Fortunately the heating value of the carburetted gas could be controlled to a large extent by adjusting the type and quantity of oil sprayed on the refractory bricks.

Then there was "producer gas," which was even cheaper than carburetted water gas. It had a heating value of from 110 to 160 B.t.u. per cubic foot, which meant that it could not be widely distributed because the cost would be prohibitive. (Distribution costs account for much of the consumer's bill; even with natural gas the distribution costs in the District of Columbia represent 65 percent of the total delivered cost.) Producer gas served well, however, in industries that required a clean fuel for combustion purposes or a source of heat for manufacturing processes and that could make the gas locally, so that the cost of distribution was small.

Producer gas was made in a continuous process wherein a bed of hot coal or coke was blasted with air or a mixture of steam and air. The final product necessarily contained nitrogen from the air and carbon dioxide resulting from the combustion of carbon with oxygen. The carbon dioxide could be removed, but the nitrogen diluted the heating value of the gas. This disadvantage was unimportant when the gas was burned as an industrial fuel but made it unsuitable as a synthesis gas (hydrogen and mixtures of hydrogen and carbon monoxide employed in making plastics, synthetic fuels and other synthetic products) and for distribution by utility companies.

Starting in about 1850 and continuing until World War II, the technology of making water gas and producer gas was improved steadily in the U.S. After manufactured gas lost its markets in the U.S. the technology was further improved in Europe, where coal was then the only indigenous fuel found in any significant quantity. New processes were investigated, taking advantage of technical advances in other fields such as the development of large-scale oxygen plants, new methods of handling solids in reactions with gas and improved construction materials. At about the time the technology had reached a stage where plants could be installed, however, natural gas was discovered in the North Sea and in North Africa. In addition most of the European nations decided to shift from an economy based on high-cost indigenous coal to one based on what was at the time low-cost imported petroleum. Few coal gasification plants embodying new technology were installed, and interest in further improving the technology flagged.

The two processes that were installed most often in Europe and elsewhere were the Lurgi fixed-bed, pressurized gasifier and the Koppers-Totzek fully entrained, atmospheric gasifier. Both types served mainly to make synthesis gas for the manufacture of ammonia and other synthetic products, although in some places Lurgi plants made a gas that was distributed as city gas. Neither of these processes meets all the qualifications of an ideal gasification scheme. An ideal process would be a single-stage, continuous operation employing air as the oxidizing medium; it would convert any type of coal into a combustible gas or a synthesis gas that was low in inert constituents.

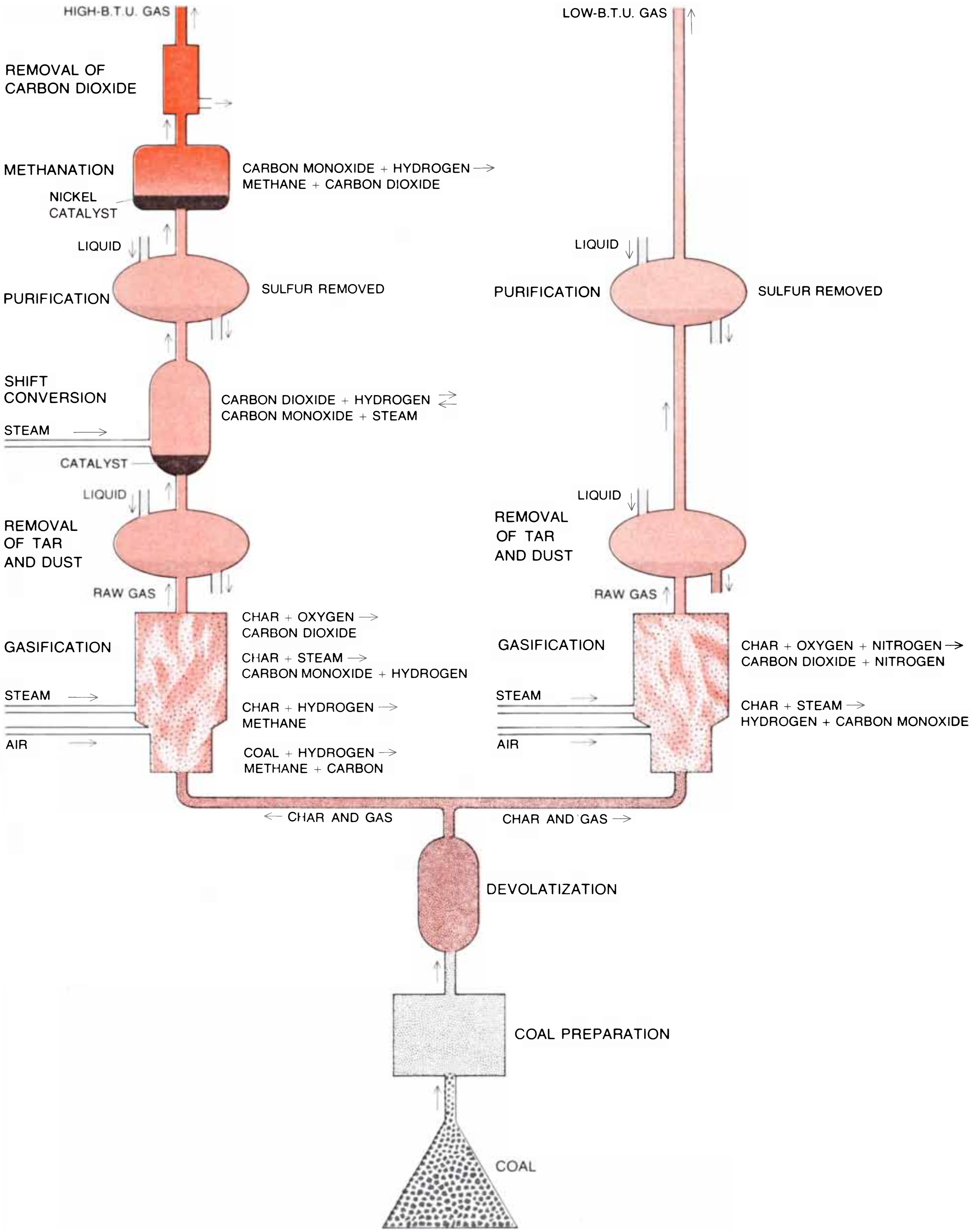
Although both the Lurgi process and the Koppers-Totzek process are single-stage and continuous, they both rely on oxygen rather than air as the oxidizing medium. The Lurgi gasifier is pressurized, which for most applications of synthesis gas is an economic advantage over processes operating at atmospheric pressure. On the other hand, the Lurgi process requires a sized coal (fines must be removed) and a noncoking or weakly coking coal (because if the bed of coal cokes during gasification, it forms a solid mass that prevents the passage of gas through the bed and brings gasification to a halt), whereas the Koppers-Totzek process will function with any coal.

In the U.S. the renewed interest in coal gasification to produce a substitute for natural gas has led to serious consideration of installing Lurgi or Koppers-Totzek plants. Neither process, however, yields a gas that could substitute for natural gas because it does not have the heating value or the composition required. The gas resulting from each process is basically a mixture of carbon monoxide and hydrogen, plus a certain amount of methane distilled from the coal during heating. Carbon monoxide and hydrogen both have a heating value of 300 B.t.u. per cubic foot—far below the heating value of natural gas. Moreover, local gas companies would not be allowed to distribute such mixtures (although they once were) because carbon monoxide is poisonous and hydrogen is difficult to contain and also has burning characteristics that would require special burners.

Both processes could, however, make a raw synthesis gas that (after purification) could be methanated to yield a gas suitable as a substitute for natural gas. Methanation involves passing the gas over a special nickel catalyst to convert it into almost pure methane, which is what natural gas is. So far, however, the methanation step has not been demonstrated on a commercial scale.

Although research in the U.S. on coal gasification is now centered on making a low-cost synthesis gas from which a substitute for natural gas can be produced, in the near future coal gasifica-

© 1974 SCIENTIFIC AMERICAN, INC



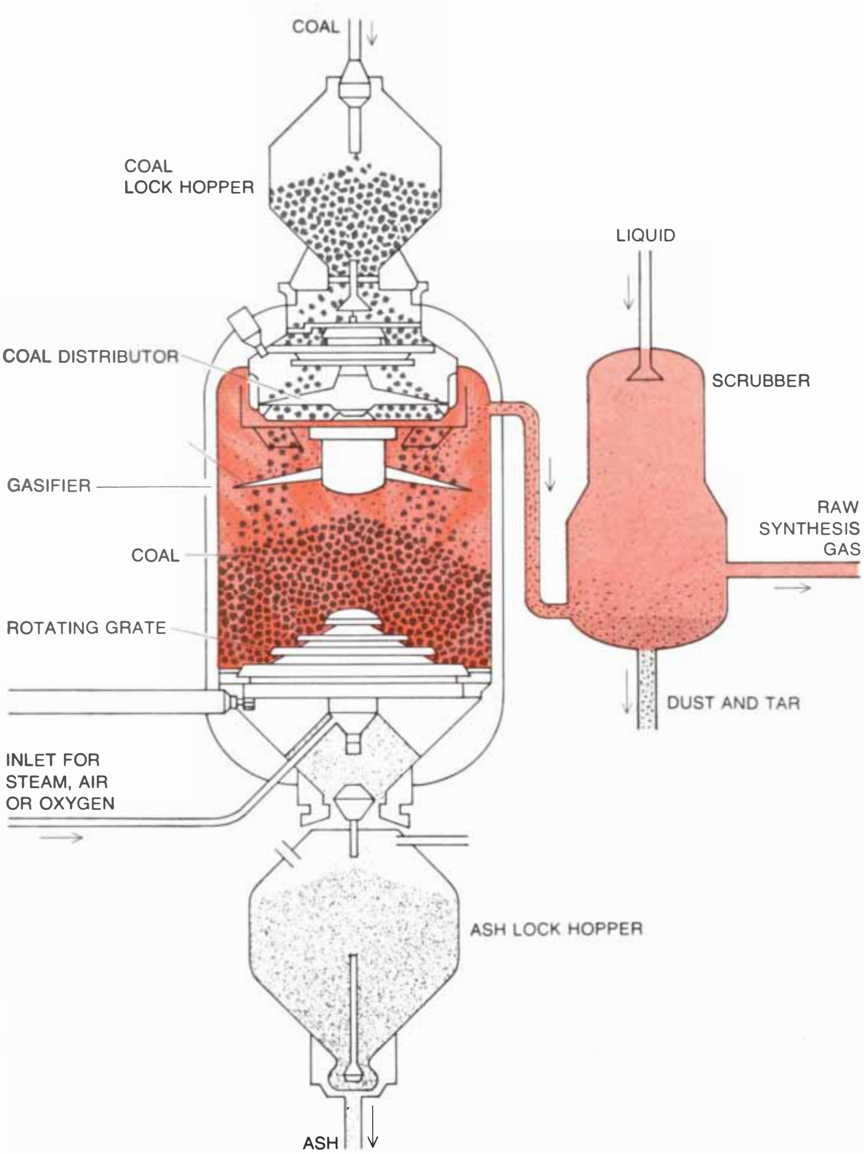
STEPS IN COAL GASIFICATION are charted for two types of gas (*color*). In each case the coal is prepared as necessary and then devolatilized, which entails heating it in the absence of air so that it decomposes chemically, yielding various gases and char. If the objective of the process is a gas with the heating value of natural gas, the steps following devolatilization are as shown at left. The simpler process charted at right yields producer gas, which has a much lower heating value but is suitable as a fuel for a boiler.

© 1974 SCIENTIFIC AMERICAN, INC

tion will also be employed to provide the raw material for the manufacture of other synthetics (alcohols, ketones, waxes and all types of petroleum products) and for the hydrogen required to produce synthetic liquid fuels from coal. (The ratio of carbon to hydrogen is much higher in coal than in oil; coal liquefaction involves producing hydrogen from coal by a gasification process and reacting it with coal so as to increase the hydrogen content of the coal and produce an oil.)

In addition new regulations on the emission of sulfur oxides and the difficulties that have been encountered with stack-scrubbing processes have led to a renewal of interest in producing a low-B.t.u., low-sulfur gas from coal to be burned as a boiler fuel. Most of the early research and development on sulfur oxide control was centered on removing sulfur oxide from the flue gases resulting from the combustion of coal. All the stack-scrubbing processes tested are reported to have been too unreliable for the sustained performance required for utility and industrial boilers, and the costs are much higher than originally anticipated. Two particularly troublesome problems are the large volume of gas that must be treated (making the cost of removal high) and the disposal in an ecologically acceptable manner of the solid waste material that is formed.

If coal is first gasified with air to form a low-B.t.u. gas, the resulting hydrogen sulfide can be removed by proved methods before the gas is burned in the boiler. The volume of raw gas to be treated would be less than half what must be handled when sulfur oxide is removed from flue gas, and no environmentally troublesome by-product would be created. Interest in low-B.t.u. gasification as a means to the control of sulfur oxides has been further stimulated by the possibility that the more advanced processes being developed for producing high-B.t.u. gas could be modified for making low-B.t.u. gas.



LURGI PROCESS is one of two methods of coal gasification that are available on a commercial scale. It is a pressurized process in which sized coal descending into the gasifier is first dried and then carbonized by reaction with oxygen and steam. In the gasifier's bottom layer the remaining carbon is burned to provide heat for the reactions proceeding above.

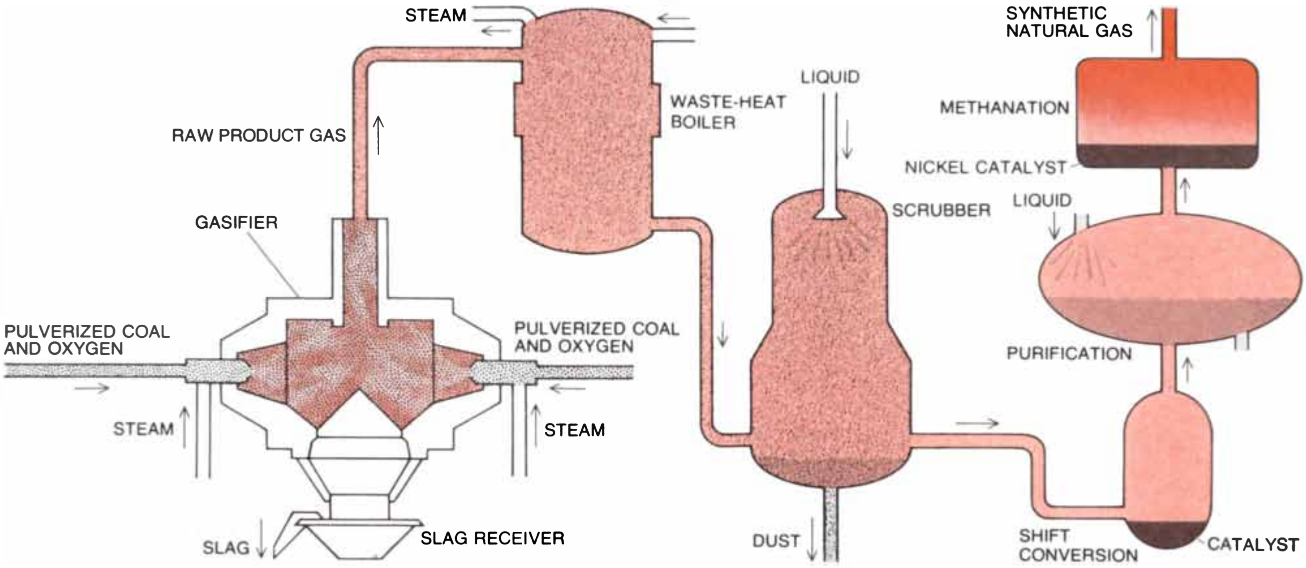
Several new methods of coal gasification are being investigated, and two have reached the demonstration stage. Since most of the research has been directed toward producing a substitute for natural gas, all the processes have attempted to retain in the product gas as much as possible of the methane that is released during the early part of the process, when the coal is simply being heated. In this way the overall capital cost and the material requirements per unit of methane are reduced. Retaining methane is also advantageous when a low-B.t.u. boiler gas is made, because the methane makes for a lower cost and a higher quality.

The basic steps required to produce a synthetic natural gas from coal can be described in general terms [*see illustration on preceding page*]. First the coal is prepared, which entails crushing it to the desired top size, removing fines (if necessary) and extraneous material and drying the coal (if necessary for the subsequent steps). The coal is then either sent to the gasifier or, if it is a coking coal and the process cannot operate with such a coal, pretreated with heat to destroy its coking properties.

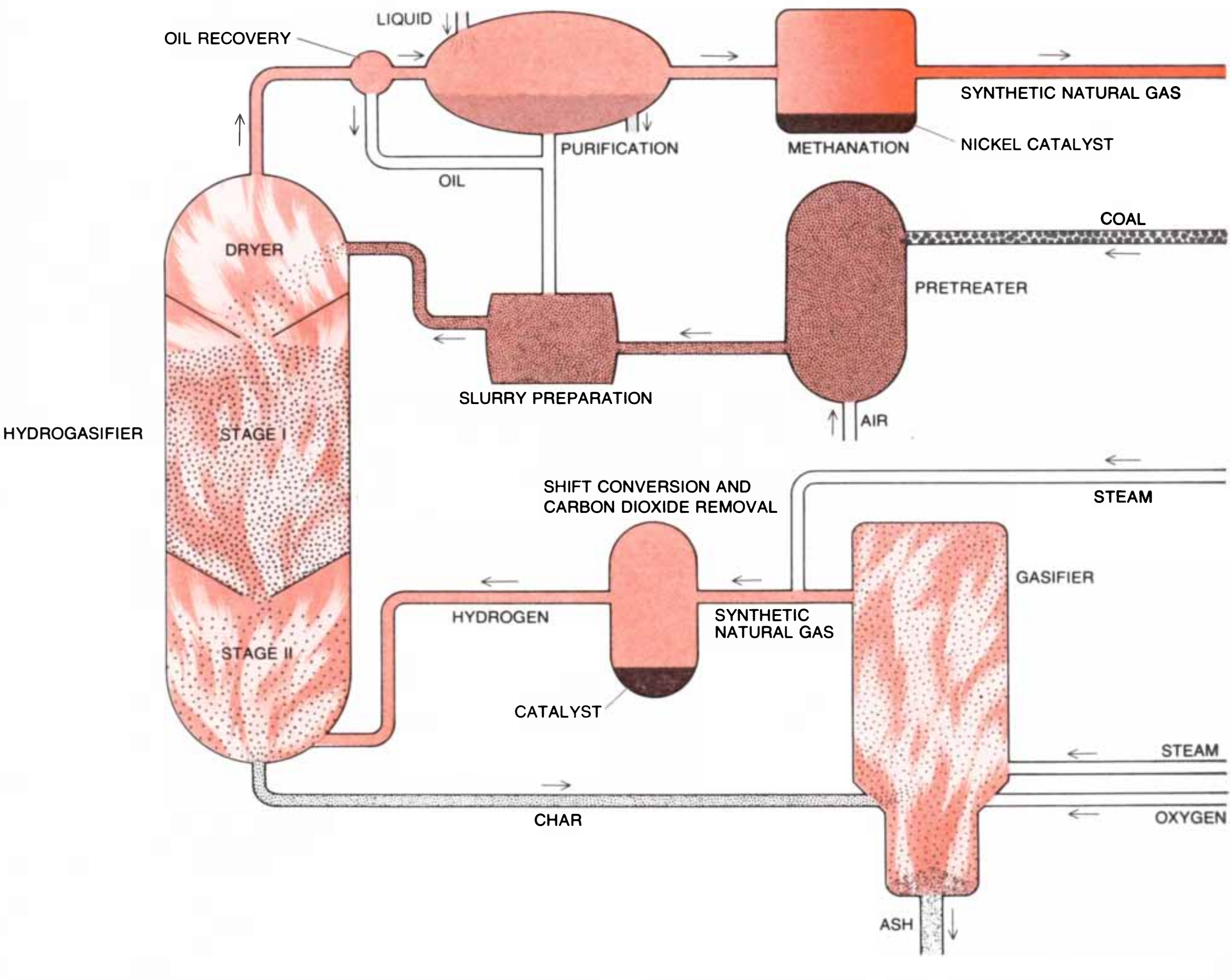
The coal is devolatilized, or distilled, either in the gasifier or in a separate vessel. As many of the distillation products as possible are retained to reduce the overall cost of gasification. Since these gases are distilled from the coal during the early part of the process and react more readily with either pure oxygen or the oxygen in the air than the char that is also produced does, the objective of retaining the distilled gases in the finished product requires special provisions to ensure that the gases do not come in contact with the oxygen and steam put into the system subsequently.

For making high-B.t.u. gas most of the processes call for steam, oxygen and either coal or char as the feed materials. The oxygen reacts with part of the carbon to provide heat and raise the temperature high enough for the balance of the carbon to react rapidly with steam to produce a mixture of carbon monoxide and oxygen. If one wanted to sub-

© 1974 SCIENTIFIC AMERICAN, INC

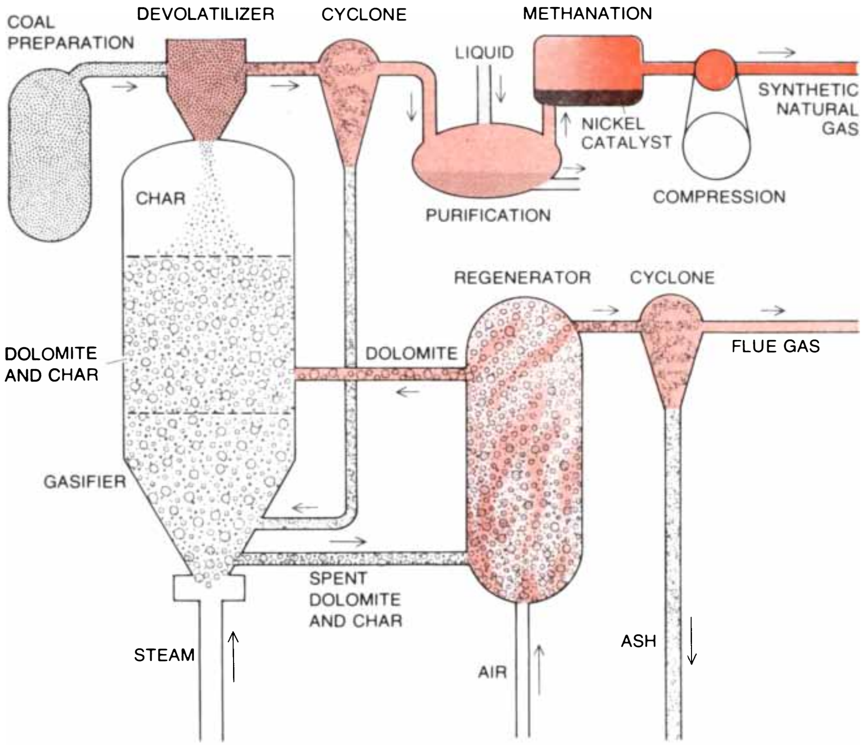


KOPPERS-TOTZEK PROCESS is the second commercial method of gasifying coal. It also employs steam and oxygen in the gasification step. The process proceeds at atmospheric pressure and will work with any type of coal. Neither of the commercial processes yields a gas with the heating value of natural gas, so that both of them require a methanation step, in which the gas produced from coal is passed over a nickel catalyst. Methanation has not been achieved on a commercial scale, but tests are now under way.

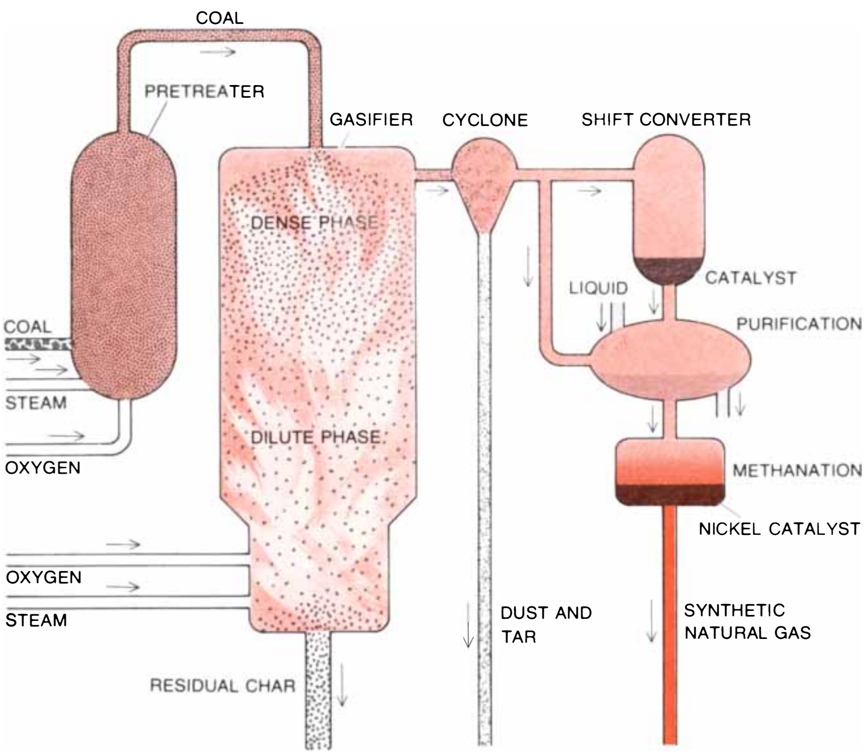


HYGAS PROCESS being developed by the Institute of Gas Technology is at the large-pilot-plant stage. Part of the coal put into the gasifier is gasified to form a mixture of carbon monoxide and hydrogen, which after shift conversion is converted to hydrogen. The hydrogen is reacted with coal or char, yielding a product that is largely methane. The gas is subjected to a final methanation reaction to attain a gas of pipeline quality. The HYGAS gasification process operates at high pressure and with a fluidized bed of coal.

© 1974 SCIENTIFIC AMERICAN, INC



CARBON DIOXIDE ACCEPTOR PROCESS, a project of Consolidation Coal Company, is also at the large-pilot-plant stage. Calcined dolomite is circulated through a fluidized bed of lignite char. The dolomite reacts with carbon dioxide produced in initial gasification of coal, liberating enough heat to sustain the carbon-steam reaction. Raw gas produced, containing methane, hydrogen and carbon monoxide, is subjected to a final methanation step.



SYNTHANE PROCESS of the U.S. Bureau of Mines is at the small-pilot-plant stage. It is designed to produce a synthetic gas that would serve as a substitute for natural gas. Coal is pretreated to destroy its caking properties, carbonized in the dense phase of the gasifier and then gasified with oxygen and steam in the dilute phase. Shift conversion to yield a 3 : 1 ratio of hydrogen to carbon monoxide is followed by methanation with a nickel catalyst.

stitute air for oxygen in order to avoid the expense of oxygen, the nitrogen in the air would dilute the product gas; therefore certain processes resort to a heat carrier as the means of providing heat. An inorganic material is heated in a separate vessel, with air as the oxidant. The material is then put into the gasifier, where it raises the temperature to the level required for the char or the coal to react with steam.

In making hydrogen, synthesis gas or high-B.t.u. gas the raw gas is next treated to remove tar and dust and is passed over a shift catalyst, which serves to speed up the reaction in which the four components of the gas (carbon dioxide, hydrogen, carbon monoxide and steam) reach equilibrium ($CO_2 + H_2 \rightleftharpoons CO + H_2O$). Any desired ratio of carbon monoxide and hydrogen can be achieved by adjusting the amounts of steam and carbon dioxide. When hydrogen is the desired final product, the carbon monoxide is all converted to carbon dioxide and scrubbed from the gas by any of several processes. In making synthesis gas, which figures in the manufacture of a wide range of chemical products, the ratio of carbon monoxide to hydrogen is adjusted to obtain the optimum composition for the chemical to be manufactured.

When high-B.t.u. gas is the objective, the appropriate ratio of hydrogen to carbon monoxide is produced. The gas is then purified to remove the sulfur compounds that may be present. The reason is that the catalysts that have been tested for the crucial step of methanation are all nickel-based and are highly sensitive to poisoning by sulfur compounds.

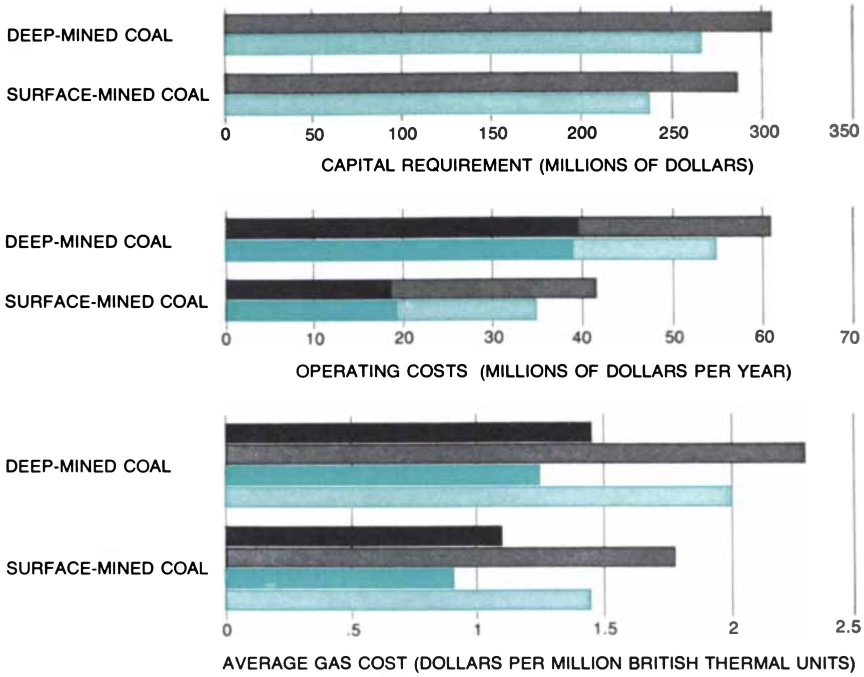
No full-scale commercial methanation plant has yet been built. A test is under way, however, at a Lurgi plant operated by the Scottish Gas Board. The test, which started last fall and is being conducted by a consortium of American firms, will use 10 million cubic feet per day of purified Lurgi gas to demonstrate the efficiency of various methanation catalysts and the amount of cleaning that must be done to keep the catalyst functioning for an acceptable period of time. A second major engineering problem is the rapid removal of heat released by the highly exothermic reaction; if the heat is not removed, it can cause deterioration of the catalyst through sintering or the deposition of carbon.

A major alternative in making high-B.t.u. gas is to gasify part of the coal to form hydrogen and then to produce methane by the reaction of hydrogen directly with coal or char ($C + 2H_2 \rightarrow CH_4$). Heat does not have to be sup-

© 1974 SCIENTIFIC AMERICAN, INC

plied, since the reaction is exothermic. The products of the reaction are then methanated to eliminate any residual carbon monoxide and to bring the gas to a heating value acceptable for distribution through pipelines. The major advantages of direct hydrogenation are that it requires fewer materials and has a higher thermal efficiency than complete gasification followed by methanation over a catalyst.

Making a producer gas is a much simpler operation than making a high-B.t.u. gas. Because it requires fewer steps and employs air rather than oxygen, the cost of making it on a commercial scale should be lower than the cost of making high-B.t.u. gas. After the coal or char has been reacted with air the hot raw gas, containing from 110 to 160 B.t.u. per cubic foot, is treated to remove tar and dust and further treated to remove hydrogen sulfide. The product is a low-B.t.u., low-sulfur gas suitable for boiler fuel. Since a substantial amount of natural gas (34 percent in 1971) is consumed by industry, any substitution of producer gas made from coal would reduce or even eliminate the problems arising from the present short supply of natural gas.



GASIFICATION COSTS are estimated for the Lurgi process (*gray*) and a typical new process (*color*). Darker parts of operating-cost bars represent coal cost, lighter parts other costs. Average gas cost is for 20 years from 1975 (*darker bars*) or from 1990 (*lighter bars*).

The two new processes that have reached the demonstration-plant stage are the HYGAS process, which is being developed by the Institute of Gas Technology, and the CO_2 Acceptor process, which is a project of the Consolidation Coal Company. Other processes, which are at or approaching the pilot-plant stage, include the BI-GAS process of Bituminous Coal Research, Inc., and the Synthane process of the U.S. Bureau of Mines. All the processes under investigation can be classified in various ways: by the method of supplying heat for the gasification reaction (internal heating or external heating); by the method of achieving contact between the reactants (fixed bed, fluidized bed or entrainment in the gasifying medium); by the flow of reactants (cocurrent or countercurrent); by the gasifying medium (hydrogen or steam plus oxygen, air or enriched oxygen), and by the condition of the residue removed (slagging, which means that the residue is liquid ash, or nonslagging, which means that it is dry ash). Nearly all the combinations of ways to gasify coal represented by these classifications have been investigated.

In the absence of full-scale commercial plants the cost of coal gasification can be only roughly estimated, and the estimates must be viewed with caution. In a presentation to the Panel on Coal Gasification of the National Academy of Sciences representatives of the Lurgi process estimated that as of about the beginning of 1972 the capital cost of a gasifier producing 250 million cubic feet of gas per day would be from $180 million to $190 million. The estimate did not include working capital, interest during construction or start-up expenses; these items would increase the cost to about $250 million, based on 1970 dollars. In the fall of 1973 the El Paso Natural Gas Company estimated the cost of a similar plant at more than $400 million. All the newer processes are estimated to have somewhat lower capital costs than the Lurgi process [*see illustration above*].

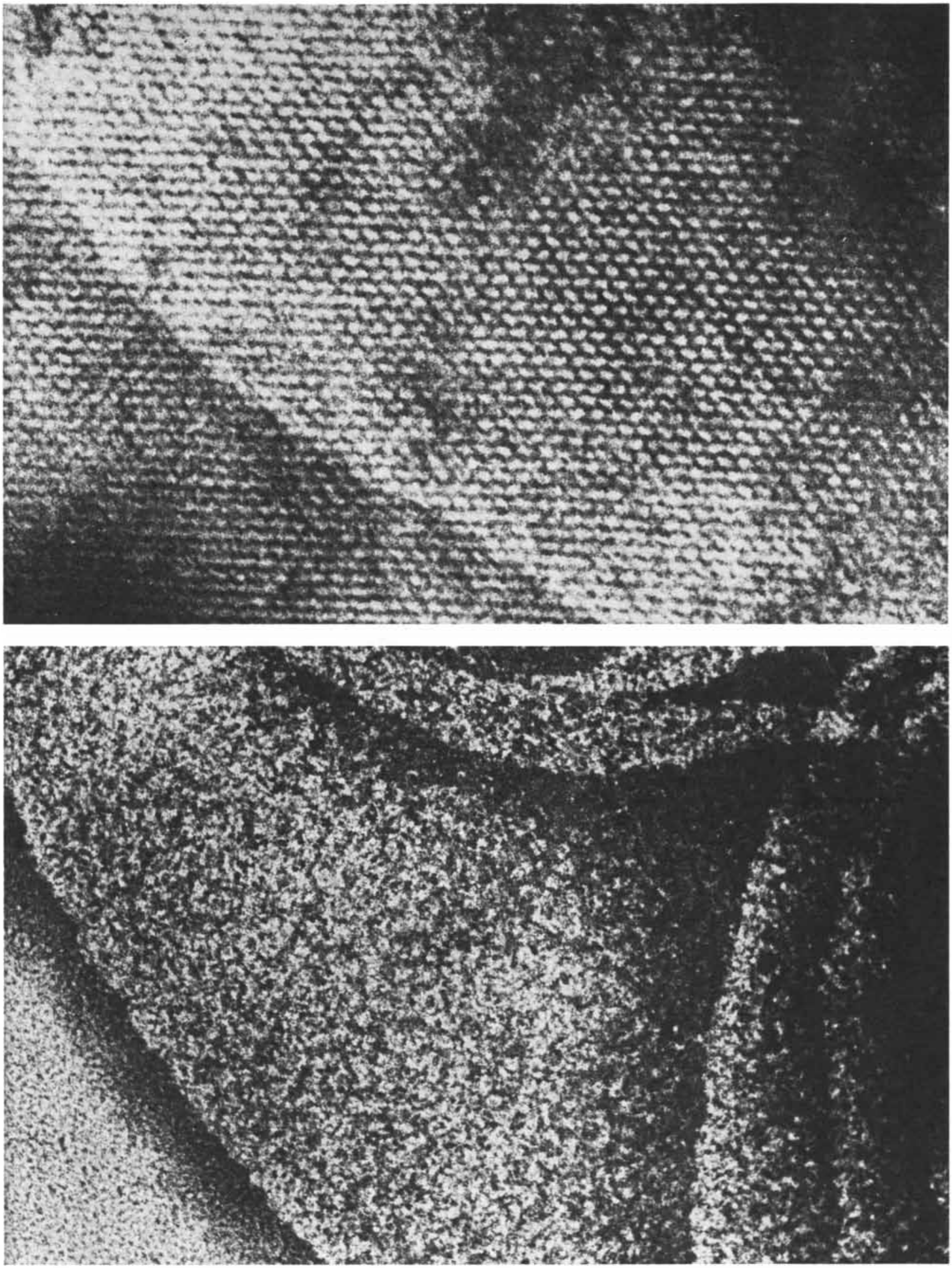
An entirely different technology that is attracting interest is gasifying coal underground. The idea is not new, and a fair amount of research was done on it in various countries until about 1960. Concern about shortages of energy has focused attention on it again.

Any process would require the creation of enough permeability in the coal so that a stream of air could flow from one point in the seam to another point without undue loss of pressure. The coal could then be ignited at either end and gasification would begin. Part of the coal would burn with the injected air, producing a mixture of carbon monoxide, carbon dioxide and heat. If good contact between gas and solid could be maintained, the hot carbon dioxide would then be reduced to carbon monoxide by a reaction with the hot coal. The product would be a producer gas.

Experiments in underground gasification of coal have created the required permeability by a number of methods that entail digging shafts or by one or another variation of the percolation method, in which two holes are drilled from the surface to the coal seam and then connected so that air can flow between them. The experiments have been able to produce a combustible gas, but they do not do so continuously and the gas does not have a constant heating value. Most of the tests have resulted in poor recovery of the available coal and have failed to achieve the level of control of the solid-gas contact that is required to produce consistently a gas of high quality. Many other problems, such as subsidence of land and contamination from ground water, have not been solved.

A number of new approaches have been proposed. They include breaking up the coal with explosives and linking boreholes by drilling with lasers. If a successful method could be developed, the energy from coal in the ground could be extracted with little or no requirement for underground manpower. Moreover, it might be possible to extract the energy from coal seams that are too thick, too thin or too poor in quality to be exploited by conventional methods.

© 1974 SCIENTIFIC AMERICAN, INC



CYTOCHROME OXIDASE, a protein present in the membrane of mitochondria, can be isolated for examination under the electron microscope, as seen here, and for X-ray-diffraction studies. In its oxidized state (*top*) the protein is organized into a crystalline lattice. The lattice structure is lost when the protein is in its reduced state (*bottom*), but the structure reappears when the reductant is removed. Measurements show that the cytochrome oxidase molecule is oval in cross section and its main axis is 80 to 85 angstroms long.

© 1974 SCIENTIFIC AMERICAN, INC

A Dynamic Model of Cell Membranes

The envelopes that surround entire cells, cell nuclei and the various cell organelles are thin assemblies of lipid and protein molecules. Their functions depend on how the membrane proteins are linked

by Roderick A. Capaldi

The fundamental unit of living tissue is the cell. In recent years it has become plain that one cellular component, the membrane, plays a crucial role in almost all cellular activity. Cytoplasmic membrane, the outer envelope surrounding the cell, acts to regulate the internal environment of the cell and to transport substances into and out of it. Internal membranes, which enclose the nucleus of the cell and such cell organelles as microsomes, mitochondria and the chloroplasts of plants, play an equally important role. For example, the mitochondrial membrane is where adenosine triphosphate (ATP) is manufactured; hence this membrane supplies the fuel for all the cell's metabolic processes. Similarly, the chloroplast membrane is the site of photosynthesis, the process by which energy from the sun is trapped in a form that can be used by living cells. How then are membranes built to accomplish so many different tasks?

A good deal of information now exists concerning the basic structure of membranes. One fact to emerge recently is that cytoplasmic and internal membranes are essentially alike; both are composed of proteins and the fatty substances called lipids. In mammalian cells small amounts of carbohydrate are also present, associated either with protein as glycoproteins, that is, carbohydrate-bearing proteins, or with lipid as glycolipids.

Lipids account for about half of the mass of most membranes. In internal membranes the lipid is almost exclusively phospholipid. Cytoplasmic membranes, in contrast, contain both glycolipid and neutral, or uncharged, lipid in addition to phospholipid. For example, as much as 30 percent of the lipid in the membrane of red blood cells consists of one type of neutral lipid: cholesterol.

Individual lipid molecules have a head and two tails [*see illustration on next page*]. At the point where the head and the tails meet, which C. Fred Fox of the University of California at Los Angeles calls the backbone, is a glycerol group. The tails that descend from the backbone are extended chains of fatty acids. The structure of these chains is quite similar to that of oil molecules and, just as oil and water tend to separate into different layers when they are mixed, so do the tails of phospholipid molecules tend to point away from water. Hence they are said to be hydrophobic. The heads of the phospholipid molecules, on the other hand, are soluble in water and are said to be hydrophilic. Molecules of this kind, with one hydrophobic end and one hydrophilic, are called amphipathic. Glycolipids and to some extent neutral lipids are also amphipathic.

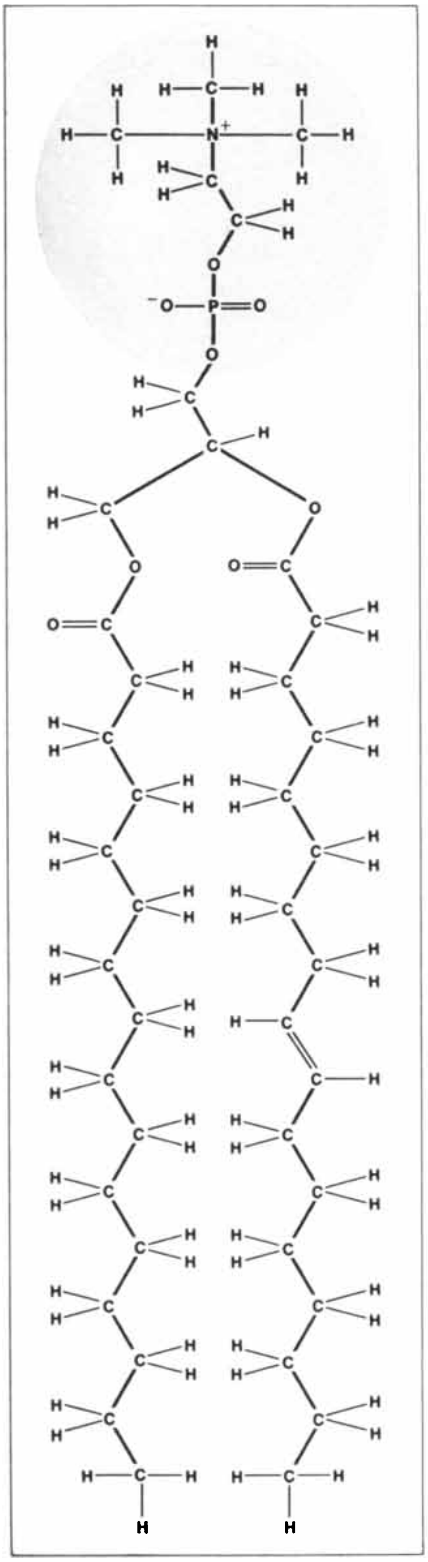
The lipids in membranes are arranged so as to accommodate their amphipathic character. They form a bilayer, two layers back to back, so that their hydrophilic heads constitute the top and bottom surfaces of the membrane and their hydrophobic tails are buried in the membrane interior [*see upper illustration on page 29*]. The lipid bilayer is a sheet about 45 angstroms thick. It is the structural framework of the membrane. It is also the anchorage for the other major component of membranes: protein.

Proteins and glycoproteins play a variety of roles in membranes. They can contribute to the structural integrity of the membrane, they can act as enzymes or they can function as pumps, moving material into and out of cells and organelles. It is the diversity of its protein activity that gives each particular membrane its distinctive character.

Just which proteins are present in which membranes can be determined in various ways. One is to assay the membrane's various enzymic activities. Another is to identify proteins by molecular weight, utilizing the technique of gel electrophoresis. Proteins are made up of the long chains of amino acids called polypeptides. Some proteins have only a single polypeptide chain; others have many polypeptide chains tightly associated with one another. In preparation for gel electrophoresis a protein is broken down into its component polypeptide chains by exposure to a detergent, sodium dodecyl sulfate. The chains are then transferred to a polyacrylamide gel with an electric potential across it. They migrate through the gel in response to the potential at a rate proportional to their molecular weight; the lower the weight of the polypeptide, the farther it moves. The gel is then stained with a protein-specific substance, for example coomassie brilliant blue, and is scanned for absorbance. When, for example, the proteins associated with the membrane of the red blood cell are identified in this fashion, the scanning trace reveals polypeptides with molecular weights ranging from 255,000 to 12,500.

The two heaviest polypeptide components, with molecular weights of 255,000 and 220,000, are collectively known as spectrin. (Vincent T. Marchesi of the Yale University School of Medicine chose the name because he first isolated the components from "ghosts," the membranes of red blood cells that have been chemically deprived of their hemoglobin.) Spectrin accounts for about a third of all the protein in the red-cell membrane. Another third of the protein lies in a diffuse absorption band with a molecular weight of about 90,000. This band contains a number of different polypeptides, including a component with a molecular weight of 87,000 pres-

© 1974 SCIENTIFIC AMERICAN, INC



AMPHIPATHIC STRUCTURE of a lipid molecule, with a hydrophilic head and twin hydrophobic tails, is exemplified by this typical phospholipid, specifically a molecule of phosphatidylcholine. Various lipid molecules comprise about half of the mass of mammalian membrane, forming the membrane's structural framework. Their fatty-acid tails may be saturated (*left*), with a hydrogen atom linked to every carbon bond, or unsaturated (*right*), with carbons free.

ent in equal copies with each spectrin molecule. Proteins of molecular weight lower than 70,000 make up the remaining third of the protein in the membrane. The red-cell membrane may be unusual in the large amount of protein of high molecular weight that it incorporates. By way of comparison, almost all the polypeptides in mitochondrial membrane are below 70,000 in weight.

Membrane proteins can be divided into two classes depending on their location with respect to the lipids of the membrane framework. One class consists of those protein molecules that are associated only with the membrane surface. These "extrinsic" proteins are located adjacent to either the outer or the inner surface of the membrane. The second class is made up of proteins that actually penetrate the membrane surface. These "intrinsic" proteins enter the lipid bilayer and sometimes extend all the way through it [*see upper illustration on opposite page*].

Whether a membrane protein should be assigned to one or the other of the two classes can be determined on the basis of its chemical properties or on the basis of other kinds of analysis, such as X-ray diffraction or electron microscopy. For example, extrinsic proteins are relatively easy to remove from membranes by chemical dissociation methods, whereas intrinsic proteins form an integral part of the membrane continuum and are much more difficult to dislodge.

Two extrinsic proteins that are visible in electron micrographs are the enzyme ATPase, found in mitochondrial membrane, and spectrin, the polypeptide in red-cell membrane. Objects termed "headpieces" are visible sticking up from the membrane surface in electron micrographs of mitochondrial membrane; they are ATPase molecules. Similarly, in electron micrographs of red-cell ghosts the "fuzz" lining the inside of the membrane is composed of spectrin polypeptides.

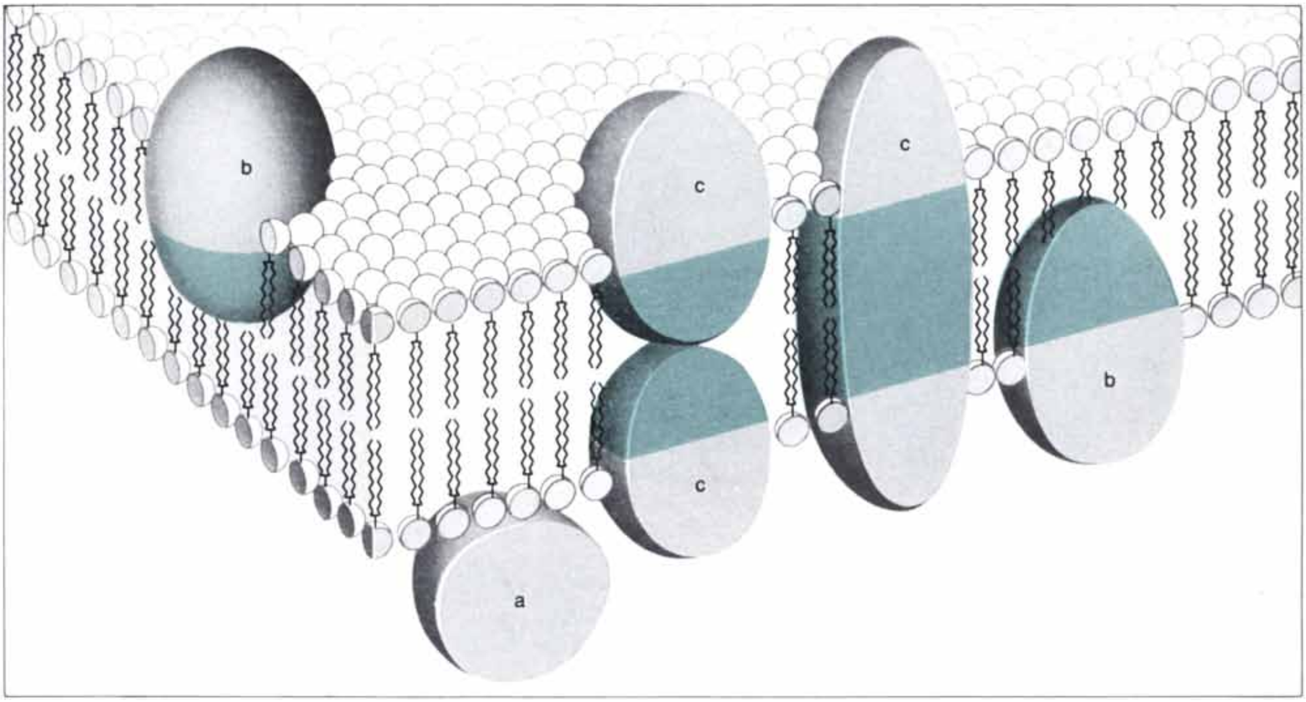
One intrinsic protein that has been studied in detail is rhodopsin, the only protein present in the membranes of the disks that occupy the outer segments of the rod cells of the retina. J. Kent Blasie and his colleagues at the University of Pennsylvania School of Medicine, working with X-ray-diffraction techniques, have found that the rhodopsin molecule is globular and some 42 angstroms in diameter. When the retinal rods are in darkness, the rhodopsin molecules of the disk membrane are submerged for about a third of their diameter in the membrane's outer surface. When the rods are illuminated, the rhodopsin molecules sink deeper into the membrane until they are about half-submerged. Even then, however, the molecules have penetrated less than halfway through the bilayer.

Working in David E. Green's laboratory at the Institute for Enzyme Research at the University of Wisconsin, my colleagues and I have examined the organization of another intrinsic protein: cytochrome oxidase, an enzyme in mitochondrial membrane that is the terminal member of the electron-transfer chain involved in the synthesis of ATP. Now, one stumbling block in the path of membrane-protein research is the fact that most membranes contain a heterogeneous mixture of proteins, including both extrinsic and intrinsic proteins. Most of the methods that can be harnessed to examine the structure of membranes, such as X-ray diffraction, are averaging methods; the resulting data give only the average properties of all the proteins in the sample, whereas we really want to know the characteristics of individual membrane proteins. This is one reason why retinal-disk membrane, with its single protein rhodopsin, is a popular subject of investigation.

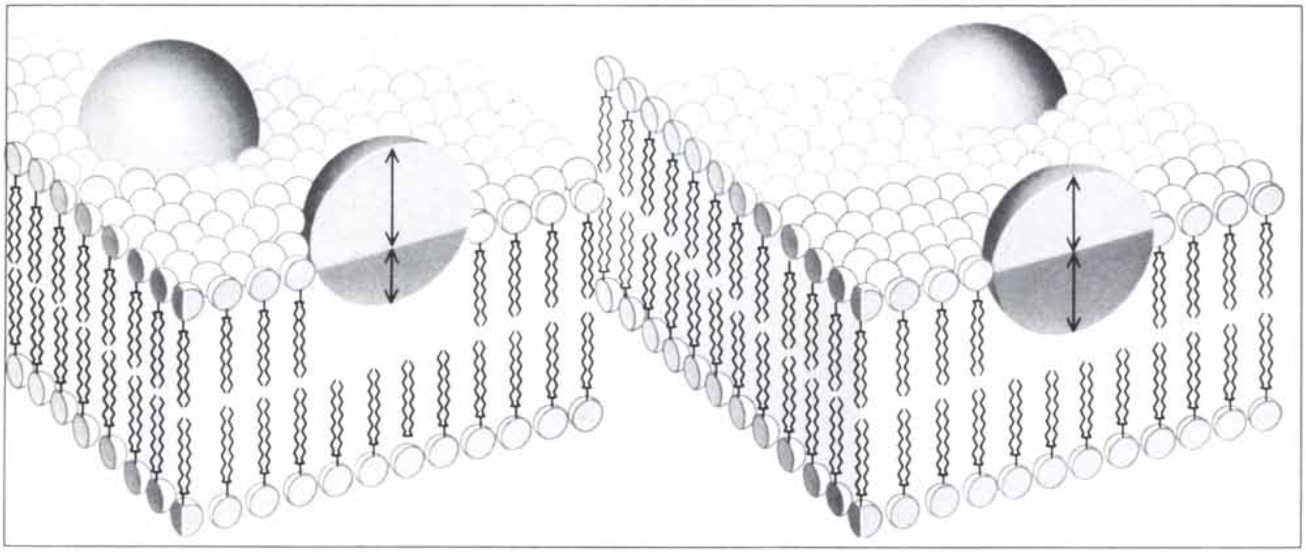
Fortunately for us it is possible to separate cytochrome oxidase from the other proteins in mitochondrial membrane. When the separated enzyme is placed in suspension with lipids, the lipids and the enzyme interact and gather in saclike vesicles that are in effect man-made membranes. The molecules of cytochrome oxidase in the artificial vesicles have the same enzymic properties that they show in normal mitochondrial membrane, and so it seems a good bet that the vesicles have the same structure as normal membrane. The advantage here is that a heterogeneous array of proteins has been reduced to a single protein.

We have used these membrane vesicles as a model system for the study of the isolated protein in its geometrical relations with the lipid bilayer. In its capacity as an electron-transfer substance cytochrome oxidase exists in one of two states, either oxidized or reduced. In the oxidized state (and in a narrow range of lipid-to-protein ratios) the enzyme is organized into a crystalline lattice that is visible in the electron microscope and can also be analyzed by X-ray diffraction [*see illustration on page 26*]. Utilizing both kinds of data, we found that individual molecules of cytochrome oxidase are some 55 angstroms long, 60 angstroms wide and 80 to 85 angstroms deep. This is enough depth to allow the molecule to penetrate the 45-angstrom bilayer completely, leaving one end jut-

© 1974 SCIENTIFIC AMERICAN, INC



STRUCTURAL FRAMEWORK typical of cell membranes is made up of a bilayer of lipids with their hydrophilic heads forming outer and inner membrane surfaces and their hydrophobic tails meeting at the center of the membrane; the bilayer is about 45 angstroms thick. Proteins, the other membrane constituents, are of two kinds. Some (*a*) lie at or near either membrane surface. The others penetrate the membrane; they may intrude only a short way (*b*) or may bridge the membrane completely (*c*), singly or in pairs.



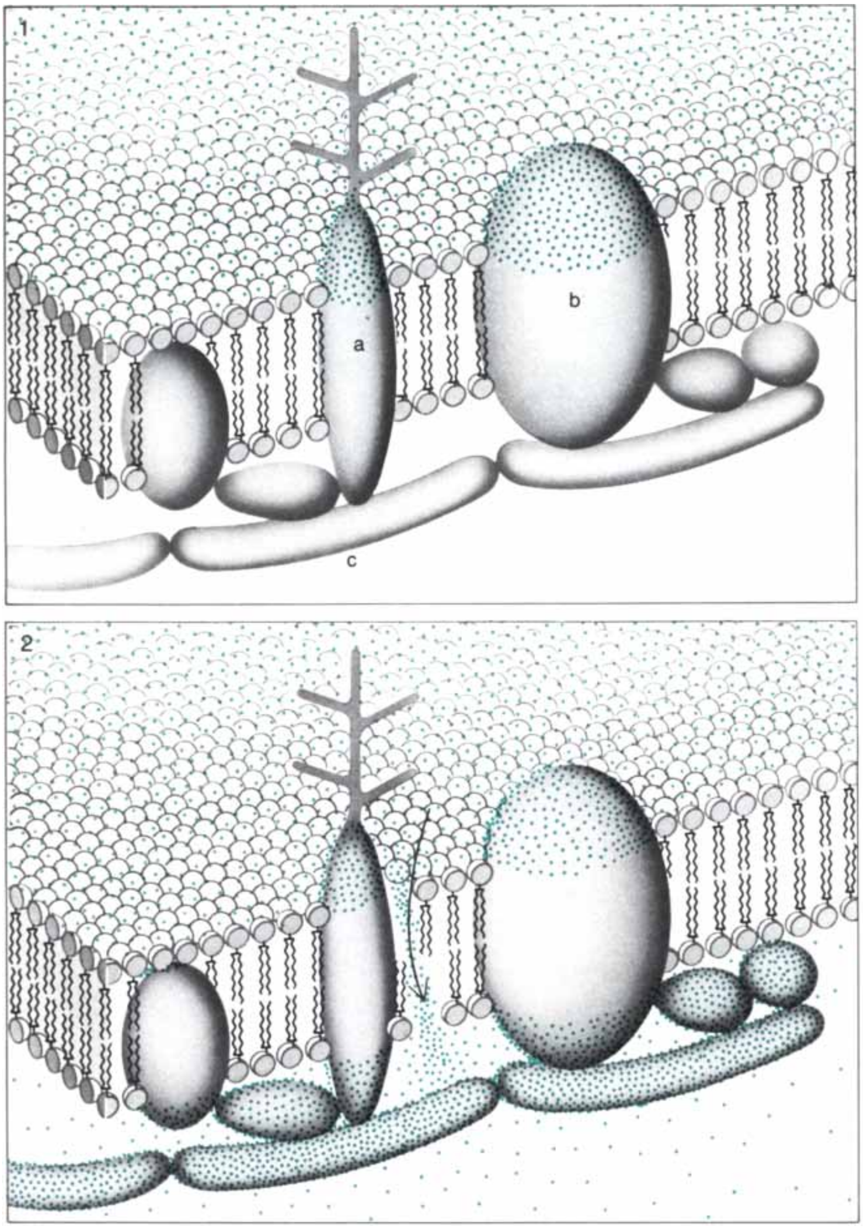
EXTENT OF PENETRATION can vary according to circumstance. Rhodopsin, the only protein in the membrane of retinal-rod disks, has been shown by J. Kent Blasie and his colleagues at the University of Pennsylvania School of Medicine to be a globular molecule 42 angstroms in diameter. In darkness a third of it is submerged in the membrane lipids (*left*); illuminated, it is half-submerged.

ting up from the exterior surface of the membrane and the other similarly exposed on the interior surface. Here is a membrane-penetrating protein that spans the bilayer by itself.

Intrinsic proteins have an environment quite unlike that of extrinsic or cytoplasmic proteins that are essentially surrounded by water. Only part of an intrinsic protein is exposed to water; the rest is essentially immersed in oil. In order to be stable in this unusual dual environment the protein must be amphipathic, as phospholipid molecules are. The parts of the molecule exposed to water must hold the majority of the seven hydrophilic amino acids: lysine, histidine, arginine, aspartic acid, glutamic acid, serine and threonine. The parts of the molecule buried in the lipid bilayer, in turn, must be mainly covered with hydrophobic amino acids.

The more deeply buried a protein is in the bilayer, the less water-surrounded surface there is to accommodate the hydrophilic amino acids in the molecule. It is therefore not surprising that many intrinsic proteins contain relatively few hydrophilic amino acids; instead they tend to be hydrophobic in character. For example, Alexander Tzagoloff of the Public Health Research Institute of the City of New York has isolated one in-

© 1974 SCIENTIFIC AMERICAN, INC



ASYMMETRIC DISTRIBUTION of the protein molecules associated with red-blood-cell membrane is demonstrated by labeling the proteins. When a chemical reagent that cannot pass through the lipid bilayer is applied to the cell surface (*1*), only two penetrating proteins with ends that extend above the surface (*a, b*) are labeled. When a lysing agent is added, making the membrane leaky (*2*), all the molecules on or near the other side of the membrane are labeled, including the spectrin (*c*) that probably holds the others in place.

trinsic protein of the inner membranes of mitochondria with a molecular weight of 10,000. Its amino acid content is 20 percent hydrophilic and 80 percent hydrophobic. This is in sharp contrast to cytoplasmic and extrinsic membrane proteins, which have on the average 47 percent hydrophilic amino acids and 53 percent hydrophobic amino acids. Another very hydrophobic intrinsic protein is the Folch-Lees protein, which can be isolated from the myelin covering of sciatic nerve; its amino acids are 29 percent hydrophilic and 71 percent hydrophobic. The intrinsic proteins rhodopsin and cytochrome oxidase are both hydrophobic; the amino acids of rhodopsin are 36 percent hydrophilic and 64 percent hydrophobic and those of cytochrome oxidase are 37 percent hydrophilic and 63 percent hydrophobic.

Two interesting intrinsic proteins have recently been characterized that, while they are not excessively hydrophobic in overall composition, have polypeptide-chain regions that are very rich in hydrophobic amino acids. One is cytochrome b_5, a protein isolated from the microsomal membrane of liver cells. Phillip Strittmatter and Lawrence Spatz of the University of Connecticut have shown that this protein, a single polypeptide, is folded at one end into a globular portion that is exposed at the surface of the membrane and is covered predominantly with hydrophilic amino acids. The polypeptide chain continues out of the globular portion into a "tail" of about 60 amino acids, almost all of them hydrophobic. The tail penetrates into the bilayer and serves to anchor the globular and enzymically active portion of the molecule to the membrane.

The second intrinsic protein with an unusual structure is the major glycoprotein found in the membrane of red blood cells. This molecule has been closely studied by Marchesi and various colleagues, first at the National Institutes of Health and more recently at Yale. Again it consists of a single polypeptide chain. One end of the chain, which holds all the carbohydrate associated with the molecule, consists predominantly of hydrophilic amino acids; this end is exposed to the water at the outer surface of the cell membrane. The other end of the chain, which also incorporates hydrophilic amino acids, extends into the watery interior of the red cell. The middle of the chain consists of some 30 amino acids. They are almost exclusively hydrophobic and lie inside the lipid bilayer of the membrane.

Because all the carbohydrate of the glycoprotein molecule is exposed at the outer surface of the red-cell membrane the membrane is asymmetric. Furthermore, labeling studies, using chemical reagents to label "available" proteins, show that the asymmetry in the red-cell membrane extends beyond carbohydrate imbalance. Reagents that cannot penetrate the lipid bilayer of the membrane will label two of the proteins in the membrane of intact red blood cells. One is the glycoprotein; the other is the protein of molecular weight 87,000; only those two proteins are exposed at the outer surface of the membrane. When the red cell is lysed and thereby made leaky to the labeling reagent, however, all the proteins in the membrane are labeled, indicating that the majority of proteins in the red-cell membrane are localized on the membrane's interior surface.

The red-blood-cell membrane is not the only one with an asymmetric organization. Labeling techniques have shown that the mitochondrial inner membrane is organized in a similar fashion. The protein molecules known as headpieces, actually ATPases, are located exclusively on the matrix, or inner, side of the mem-

© 1974 SCIENTIFIC AMERICAN, INC

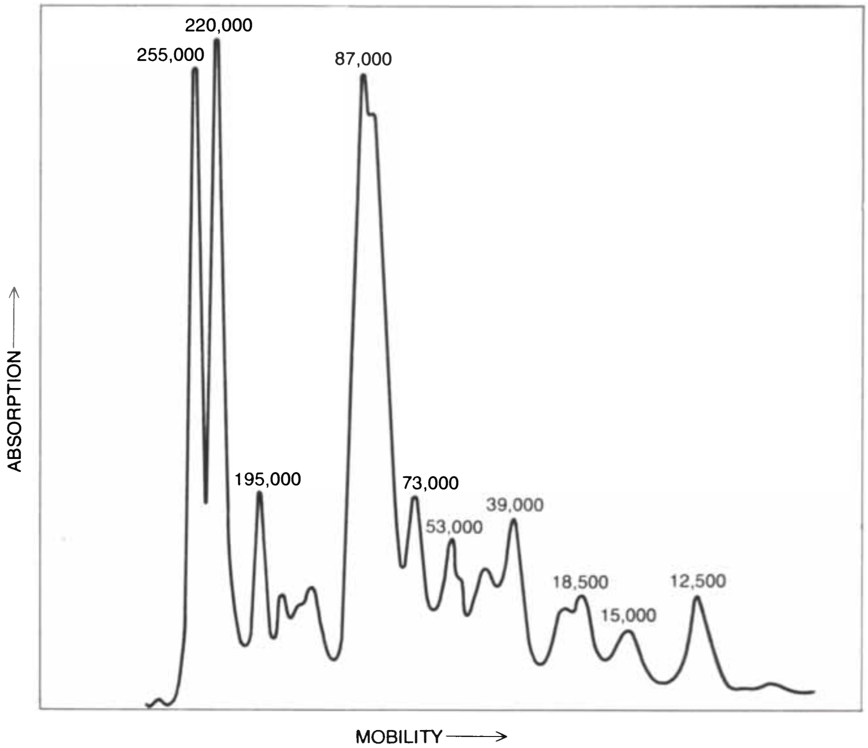
brane and the cytochrome *c* molecules are found only on the intracristal, or outer, surface. As one would expect, the molecules of cytochrome oxidase, which are exposed on both the intracristal and the matrix surfaces of the membrane, can be labeled from either side.

To review the membrane picture as it has been described so far, we see that the lipids, which account for roughly half of the mass of a membrane, are organized into a thin bilayer and that the membrane proteins are either perched on or near the two sides of the bilayer or penetrate into or completely through it. This picture is accurate but incomplete. One of the major advances in cell studies during the past few years has been the realization that membranes are by no means static. Both the lipids and the proteins have considerable freedom of movement.

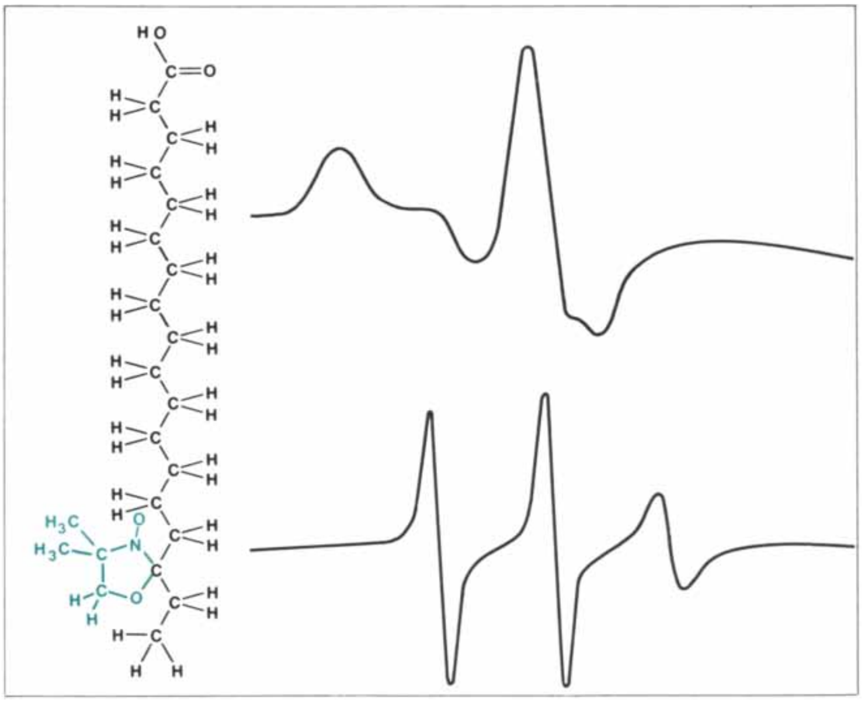
Considering the mobility of the lipids first, whether a lipid bilayer is fluid or rigid depends on two factors: first, the extent of saturation of the lipid tails (that is, the extent to which all the available carbon bonds carry hydrogen atoms), and second, the ambient temperature. Now, a considerable proportion of the lipid in mammalian-cell membrane is unsaturated, so that the melting temperature for the bilayer is below the normal mammalian body temperature. Thus the bilayer is fluid and the fatty-acid tails of the lipid molecules are free to move. The freedom of movement has been studied in some detail by Harden M. McConnell and his colleagues at Stanford University and by O. Hayes Griffith and his colleagues at the University of Oregon. Both groups have used the same analytical technique: electron spin-resonance spectroscopy. The method involves attaching a "reporter" group, usually a nitroxide group that has an unpaired electron, to one of the carbons of a test molecule's fatty-acid tail.

The test molecule is either stearic acid or a phospholipid and the reporter group is usually attached to the fifth, 12th or 16th carbon atom along the tail. Several labeled test molecules are then inserted in the membrane bilayer that is to be examined; they fit in much as ordinary lipid molecules do. The inserted group is very sensitive to motion, however, and the spectroscopic trace unmistakably records any movement of the portion of the fatty-acid tail where the reporter group is attached. By fixing reporter groups at different positions along the tail one can determine just what degree of mobility exists where in the bilayer.

That is essentially what McConnell's



IDENTITY OF PROTEINS in the membrane of red blood cells can be determined by means of gel electrophoresis. In this instance the proteins have been stained and scanned, using a densiometer, after those with the lowest molecular weights have migrated farthest through the gel in response to an electric potential. The two proteins with the highest molecular weight (*left*) form the dimer collectively called spectrin. The scan shows at least 10 more absorption bands signifying the presence of other proteins even lower in weight.



MOBILITY OF LIPIDS in the membrane bilayer can be determined by electron spin-resonance spectroscopy. A motion-sensitive "reporter" group (*color*) is attached to a carbon atom in the tail of a test molecule (*left*). When the test molecule is inserted in a bilayer, the spectrum of the reporter group will vary in accordance with the degree of bilayer mobility. If low temperature has "frozen" the bilayer so that the lipids are immobile, a characteristic spectrum results (*top right*). If the bilayer is mobile, however, the reporter group in the tail of the lipid produces a distinctly different spectrum (*bottom right*).

© 1974 SCIENTIFIC AMERICAN, INC

and Griffith's groups have done, and they have shown that there is a flexibility gradient in the bilayer. The part of the tail that is closest to the head of the lipid molecule and therefore nearest the surface of the bilayer is the least flexible. Conversely, the greatest flexibility is found at the tips of the lipids' tails and thus in the membrane zone nearest the center of the bilayer.

Not every lipid molecule in the bilayer is mobile. For example, penetrating protein molecules affect the mobility of the lipid molecules adjacent to them. In collaboration with Griffith and Patricia C. Jost, we have used spin-labeled stearic acids to probe the lipid environment immediately adjacent to molecules of cytochrome oxidase in our model membrane system. We find that the cytochrome oxidase molecules effectively immobilize sufficient lipid to coat themselves with a single layer of lipid molecules. We have given the name "boundary lipid" to this tightly associated layer. It is interesting that the amount of boundary lipid is just the amount of lipid required for full activity of this enzyme. Exactly how much of the total lipid content of a membrane will be immobilized by serving as boundary lipid depends on the number of penetrating protein molecules a particular membrane contains. In mitochondrial membrane, which contains many penetrating molecules, perhaps as much as 30 percent of the membrane lipid is so engaged.

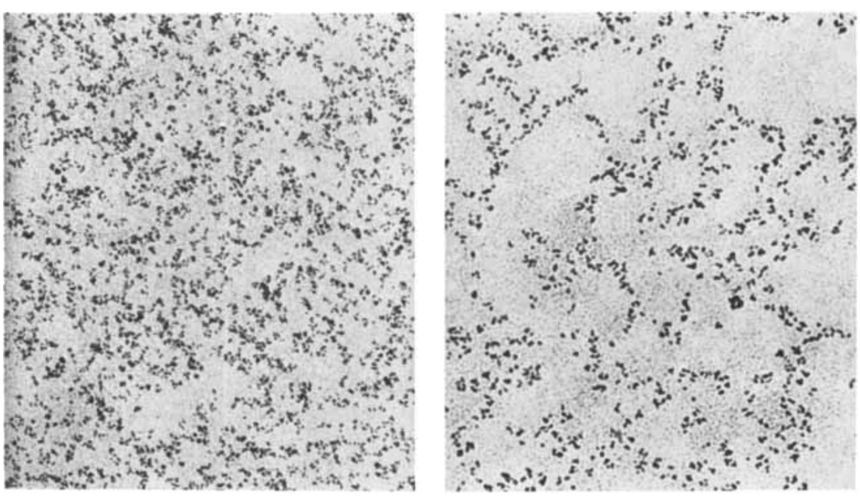
Since a major proportion of the membrane bilayer is fluid, the membrane as a whole has the consistency of a light oil. It is therefore not surprising that, when constraints do not prevent it, both lipid and protein molecules are free to move about within the sheetlike structure. Such movement is more likely to be from side to side than up and down; the stability of the asymmetric configurations in membrane testifies to that. Indeed, it would be energetically extravagant for an intrinsic protein molecule to force its highly hydrophilic end down through the hydrophobic interior of the bilayer in order to reach the other side of the membrane. The waste of energy would be less acute in the case of a lipid molecule, but even so studies of spin-labeled phospholipids indicate that the flipping of a lipid from one side of the bilayer to the other is rare.

The lateral movement of molecules through the bilayer was first demonstrated in experiments conducted by David Frye and Michael Edidin at Johns Hopkins University in 1970. They were investigating the property of the membrane in what are known as cell-fusion heterokaryons, that is, "supercells" produced by the forced fusion of a number of individual cells under the influence of Sendai virus, a well-known fusion agent.

Frye and Edidin induced the fusion of human and mouse cells in culture and then studied the distribution of certain intrinsic membrane proteins: the antigenic components of the two kinds of cells. Just where which of the proteins was present could be determined by tagging antibodies with fluorescent dyes; the differently dyed antibodies were directed either to the mouse antigen or to the human one. The investigators found that shortly after the fusion of two cells the mouse and human antigenic components were clearly segregated in what might be called the mouse and human "halves" of the supercell's membrane. After the supercell had been incubated for 40 minutes at 98.6 degrees Fahrenheit, however, the mouse and human proteins became substantially intermixed. Because the mixing had taken place in the absence of ATP synthesis the process was evidently not an energized one. Frye and Edidin could only conclude that the intermixing was a product of lateral diffusion through the bilayer. They found that when the culture was held at 34 degrees F., a temperature more than low enough to "freeze" the membrane, the two proteins did not mix.

In the years since this classic experiment was done many more antigens, and other protein molecules as well, have been found to move about within the membrane bilayer. Evidence that lipid molecules are equally mobile has come from spin-labeling studies conducted by McConnell and Philippe Devaux. They inserted patches of spin-labeled phospholipids into membranes and recorded the time required to disperse the labeled molecules by diffusion. They found that the lipids moved around at a higher rate than protein molecules; that is what one would expect, since lipid molecules are smaller than protein molecules.

Now, if lipids and proteins were all free to move around in the membrane bilayer, one would expect that the two components would be distributed at random throughout the membrane. For many membranes, however, this is not so. For example, in the membrane of cells that line the intestine the glycoproteins are concentrated at the surface end of the cell and the proteins that pump sodium are concentrated at the opposite end. Another example is the nerve cell; here a key protein, the enzyme acetylcholinesterase, is localized exclusively in the membrane at one end of the cell. Some membranes even show crystalline lattices of the kind found in the cytochrome oxidase model membrane system: the chromatophore membrane of the bacterium *Halobacterium halobium* is one example. This "purple membrane" contains only one protein in a well-defined lattice structure. "Gap junctions" between cells provide another example. Daniel A. Goodenough of the Harvard Medical School has isolated gap junctions between the cells of mouse liver. It is thought that the junctions aid the



DISTRIBUTION PATTERN of constituent proteins in red-cell membrane appears to be controlled by the location of the spectrin molecules on its inner surface. These electron micrographs by Garth Nicolson of the Salk Institute for Biological Studies show the outer surface; molecules of glycoprotein that protrude from the surface have been stained. Their distribution (*left*) is more or less generalized. The second micrograph (*right*) shows the membrane after the spectrin is aggregated; movement has made glycoprotein move too.

© 1974 SCIENTIFIC AMERICAN, INC

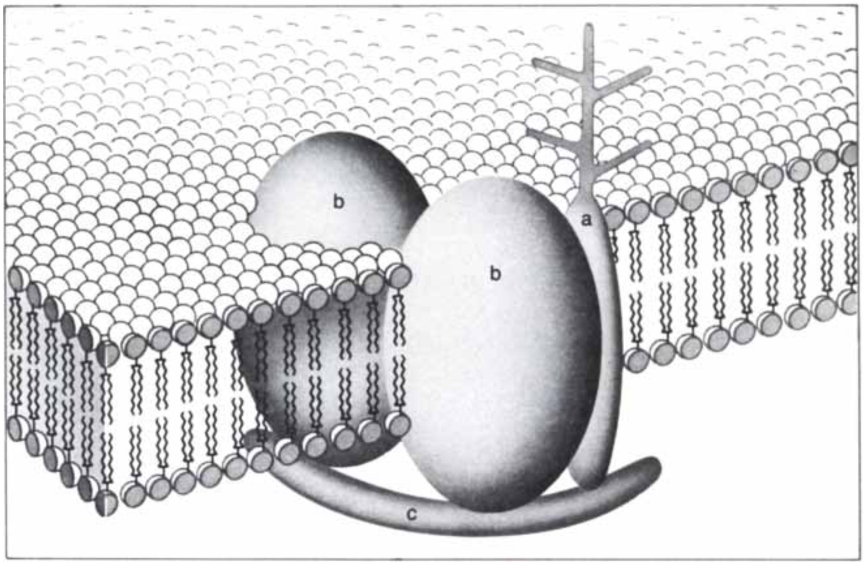
transfer of substances from one cell to another. They contain one major protein and one major lipid, presumably donated from both cells, organized into a hexagonal lattice that is clearly visible in the electron microscope.

What influences affect the mobility and hence the distribution of the proteins in membranes? The information thus far available suggests that control is exercised in many different ways. For example, it seems likely that in the mitochondrial inner membrane a tight association between the various intrinsic membrane proteins is what maintains the spatial distribution. In this membrane almost all the proteins are associated with one or another of five complexes: four that are involved in electron transfer and one that synthesizes ATP. It seems likely that proportionate quantities of each of the five complexes combine to form supramolecular aggregates that lie in an orderly array throughout the membrane. Such aggregates would promote the efficiency of electron transfer and the coupling of electron transfer to ATP synthesis.

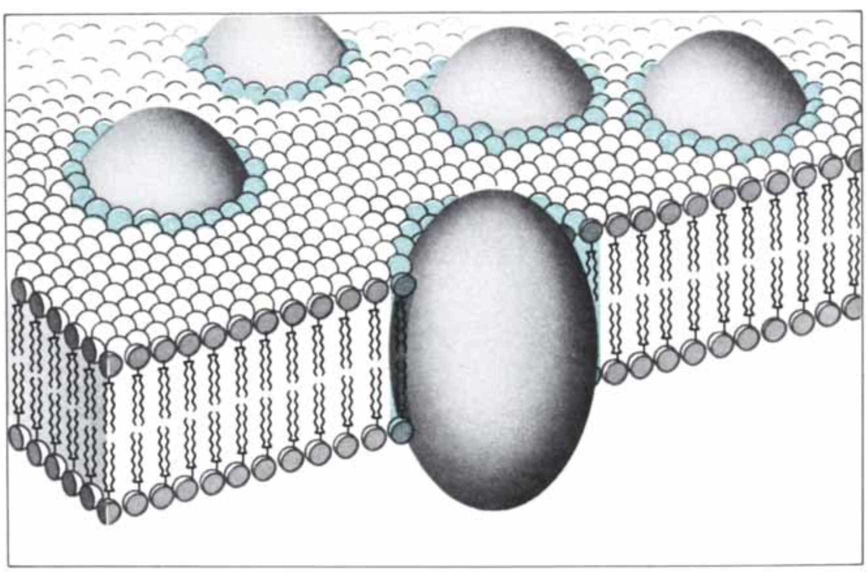
The distribution of proteins in the membrane of red blood cells appears to be maintained in a different way. Here the distribution pattern is evidently controlled by an interaction between the intrinsic proteins on the one hand and molecules of spectrin, the extrinsic protein located at the membrane's inner surface, on the other. Garth Nicolson of the Salk Institute for Biological Studies has provided a dramatic demonstration of the role played by spectrin in controlling the pattern. Nicolson has prepared an antibody to spectrin that, when it is applied to red-cell ghosts, causes the molecules of spectrin to aggregate abnormally. When this happens, the distribution of proteins intrinsic to the membrane, most noticeably the distribution of the major glycoprotein, is changed. The carbohydrate portion of this molecule contains many acidic groups, called sialic acids, that will bind an electron-dense material that is easily visible in the electron microscope. The material can therefore be used to "stain" the glycoprotein. When spectrin antibody is applied to a red-cell preparation, the position of the stained molecules is radically altered. Evidently the glycoprotein (or more probably a supramolecular aggregate that includes it) is so firmly linked to the spectrin on the inner surface of the membrane that when one moves, the other follows.

These are only two of the many systems that control the mobility and distribution of protein in membranes. A quite different system is involved with respect to the rhodopsin in retinal-rod membrane, for example, and still other systems, involving networks of microtubules and microfilaments located under the inner surface of cytoplasmic membrane, appear to control the mobility and distribution of the proteins in those membranes.

In summary, then, it appears likely that in some membranes the spatial distribution of proteins is fairly well fixed whereas in others this distribution is variable and may be controlled from within the cell. On the matter of the distribution of proteins and lipids in membranes no generalizations can be made. As was first realized by S. J. Singer of the University of California at San Diego, there is likely to be an interrelation between protein array and membrane function in cells with a variable distribution of proteins. When this is further understood, perhaps we shall be able to explain many of the key functions of membranes.



SUPRAMOLECULAR AGGREGATES in red-cell membrane include the two proteins that completely penetrate the membrane. One is a glycoprotein (*a*); the other is a protein with a molecular weight of 87,000 (*b*). The third protein in the hypothetical aggregate is spectrin (*c*). Evidently the three proteins are so linked that if one moves, the others follow.



BOUNDARY LIPID is a one-molecule-thick "coating" of immobilized lipid molecules that surrounds penetrating protein molecules (*color*). Its existence was discovered in the course of electron spin-resonance spectroscopy studies of "model" membranes that contained only one protein: cytochrome oxidase. In membranes containing many penetrating proteins, such as mitochondrial membrane, perhaps 30 percent of all lipid is immobilized.

© 1974 SCIENTIFIC AMERICAN, INC

The Neural Basis of Visually Guided Behavior

Techniques from ethology and neurophysiology are combined to show how an animal localizes a visual object, discriminates its significance and then makes the appropriate motor response

by Jörg-Peter Ewert

Animals see things and then act on the basis of what they see. What chain of events connects some key stimulus with a specific fixed pattern of responses? In recent years workers in several laboratories have sought by many different means to analyze the nerve mechanisms by which animals interpret sensory signals and select the most appropriate response. The most effective way to understand the neural basis of behavior appears to be to apply a broad spectrum of experimental techniques: to combine ethological studies of an animal's behavior with experiments involving brain anatomy and brain-cell stimulation and the recording of individual nerve-cell activity.

For the past six years, in my laboratory first at the Technical University of Darmstadt and then at the University of Kassel, we have been taking this broad approach to learn about two kinds of visually controlled behavior in the toad: orienting (prey-catching) behavior and avoidance (escape) behavior. There are several good reasons for working with the toad. Amphibians are vertebrates, so that what we learn at their relatively low level of behavioral integration contributes to our understanding of more complex vertebrate functioning. Toads in particular have a limited and easily surveyed behavioral repertory. In response to specific stimuli one can repeatedly elicit predictable reactions, such as snapping at prey, fleeing from an enemy, clasping during courtship and making particular wiping motions after tactile stimulation. (The fickle European frog, in contrast, undergoes short-term changes in motivation and is not suitable for behavioral experiments.) Finally, the toad is not easily conditioned, so that its innate behavioral functions can be measured in successive experiments for some time without being significantly affected by accumulated experience.

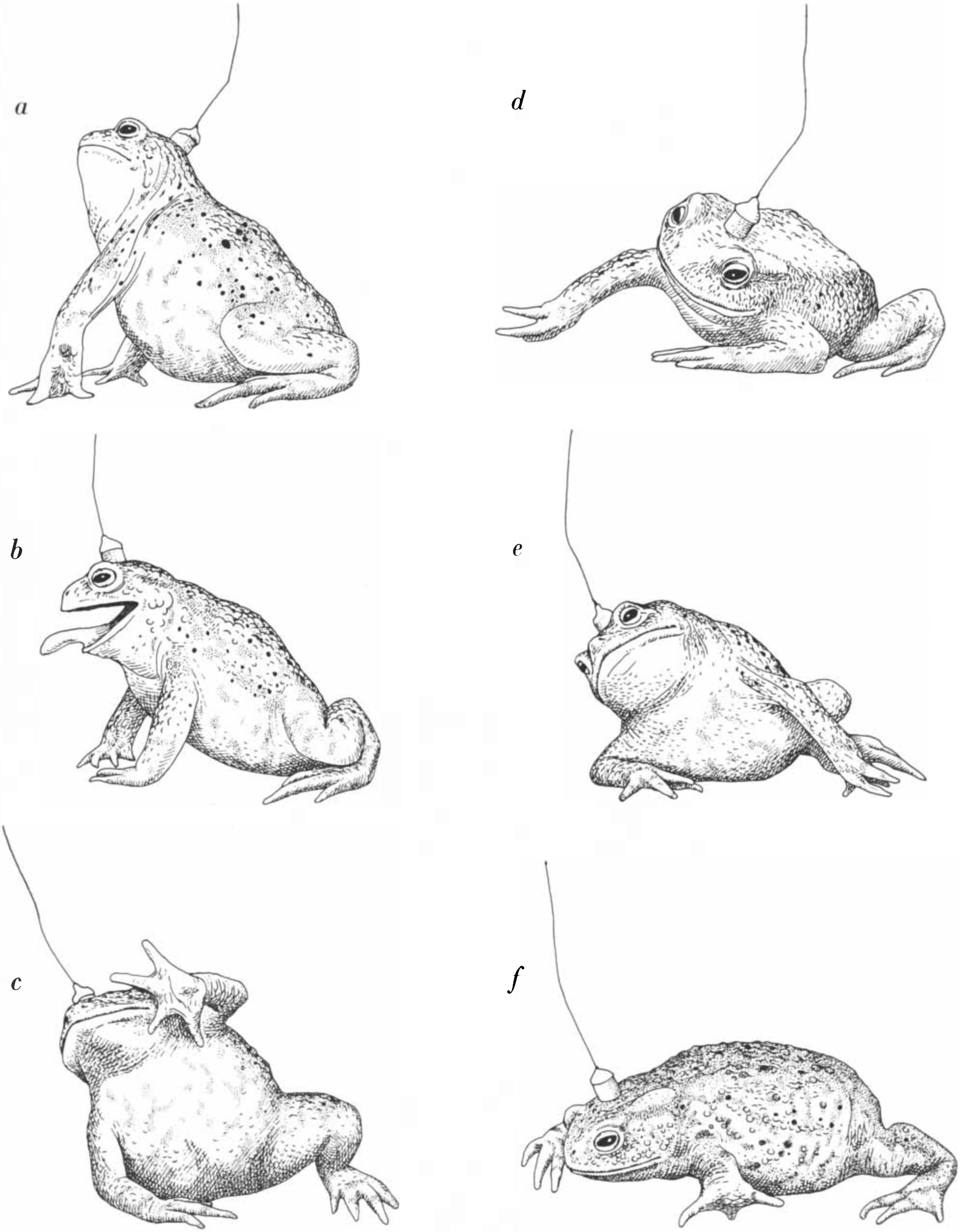
Toads respond to small objects, such as a piece of white cardboard moved over a black background, with a series of prey-catching reactions. First there is orientation toward the prey, then binocular fixation, then snapping, gulping and mouth-cleaning. Two basic processes are required to produce the overall orienting reaction: the identification of a stimulus and the location of it in space. The identification process determines the type of behavior. It is dependent on specific features of the stimulus such as its angular size, the orientation of the boundaries between light and dark, its angular velocity, its contrast with the background and so on. A detection process then localizes the stimulus and, together with the result of the identification process, determines the motor response, which can be either to turn toward the stimulus if it is identified as prey or to avoid it if it appears to be an enemy. In what follows I shall attempt to analyze the neurophysiological basis of signal identification, localization and the triggering of the associated instinctive actions.

To begin one must analyze quantitatively the key stimuli for orienting and avoidance behavior. This is done by changing various characteristics of a visual stimulus in an ordered way. The toad is placed in a cylindrical glass compartment where it observes a small square of black cardboard moving against a white background at a constant angular velocity, describing a circle around the animal at a distance of seven centimeters. The toad interprets such a stimulus as prey and tries, through successive turning movements, to keep the object fixated in the center of its visual field. The degree of orienting activity is measured by counting the number of turning responses per minute.

The angular size of the stimulus—the angle it subtends—influences the orienting activity [*see illustration on page 36*]. Of a variety of square objects toads prefer those with an edge length of four to eight degrees. (The absolute size of such stimuli is five to 10 millimeters. Experiments where the distance between animal and stimulus is varied show that it is the absolute—not the angular—size that counts; in prey-catching behavior toads display "size constancy.") The toads turn away from objects larger than 30 degrees on a side, exhibiting the avoidance response. More particular information is obtained by substituting bars of various lengths for the square stimuli. As a two-by-two-degree stimulus is elongated along the horizontal axis the orienting activity increases until a saturation level is reached; wormlike objects turn out to be particularly attractive to toads. In contrast, the response decreases as a small stimulus is extended vertically, or perpendicularly to the direction of movement.

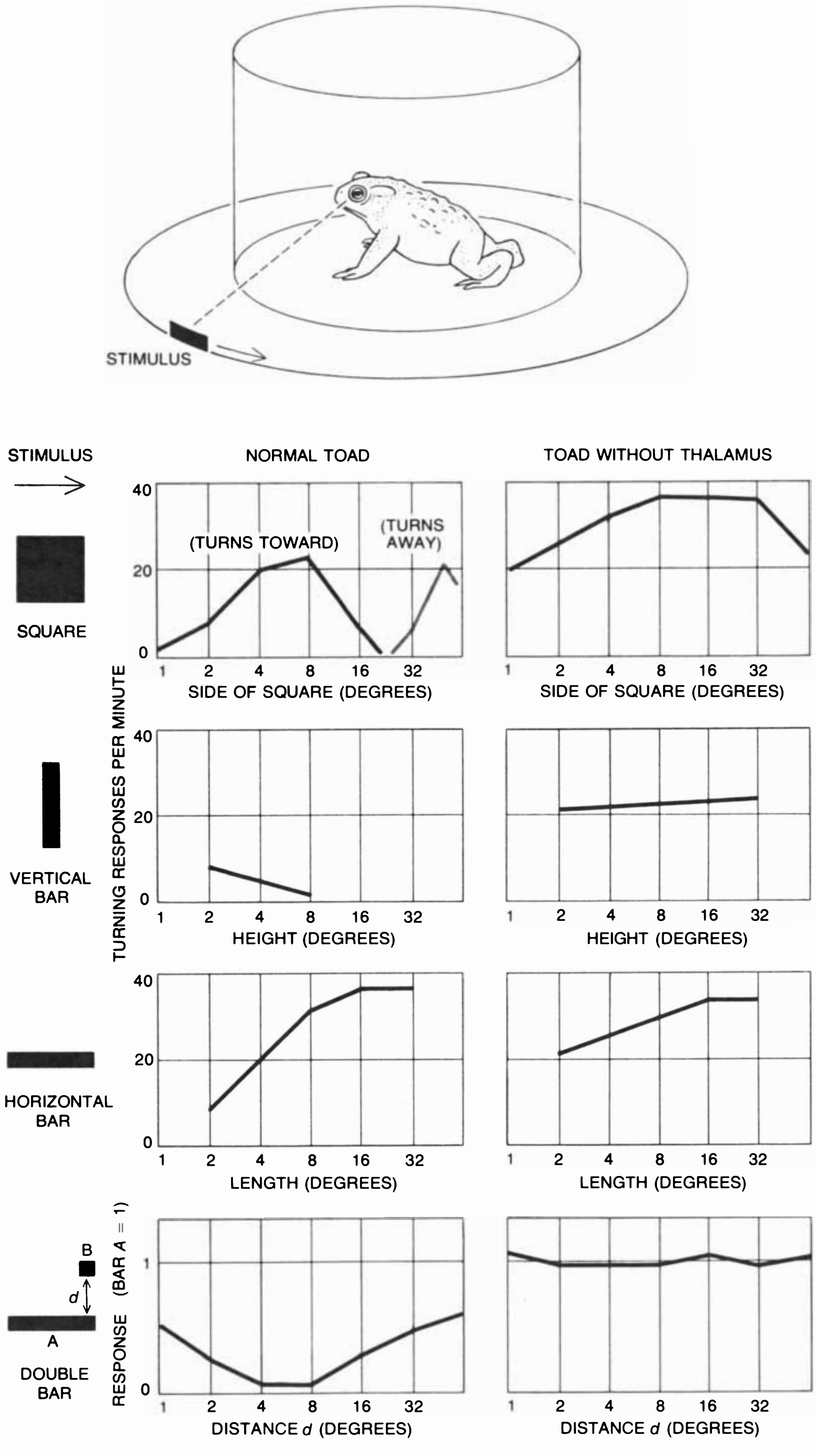
Other experiments indicate that toads discriminate prey from enemy objects through analysis of the visual stimulus in terms of point or edge configurations, also taking into consideration the direction of movement. A horizontal chain consisting of several two-by-two-degree units moving along the same path signifies prey. One such unit moving alone constitutes a prey stimulus just above the response threshold. When the horizontal chain is supplied with a separate vertical

© 1974 SCIENTIFIC AMERICAN, INC



BEHAVIORAL PATTERNS characteristic of the toad *Bufo bufo* are illustrated. The actions are commonly elicited in the animal by the sight of visual objects. These drawings, however, are based on photographs of toads whose brains were being stimulated electrically as part of the author's investigation of the neural bases of visually guided behavior. The electrode on the toad's head penetrates to the brain. An electric current applied to the optic tectum, a visual center in the brain, elicits a prey-catching sequence: orienting, or turning (*a*), snapping (*b*) and mouth-cleaning (*c*). Electrical stimulation, instead, of a site in the left or right thalamus brings a "planting-down" defensive posture (*d*, *e*) and stimulation of another part of the thalamus brings a crouching avoidance response (*f*).

© 1974 SCIENTIFIC AMERICAN, INC



BEHAVIORAL RESPONSES of the toad to objects of various shapes and sizes were quantified. Small black objects were moved across the visual field at seven centimeters' distance and the orienting response was determined for normal toads (*left*) and those whose thalamus had been removed (*right*). Prey-catching responses (turning toward the object) were elicited most effectively in normal toads by squares with sides subtending four to eight degrees; the toads turned away from larger squares. Vertical bars were ineffective as prey objects—and increasingly ineffective with increasing height. Horizontal (wormlike) bars were increasingly effective as prey objects with increasing length, up to a limit. Double bars (a horizontal bar plus a vertical extension) were less attractive, the effect varying with distance between bars; the ratio of their effect to that of a single bar is shown (*bottom*). In toads lacking the thalamus the orienting response becomes "disinhibited." The animal tends to orient toward a target without discrimination, even if the target normally signals "danger."

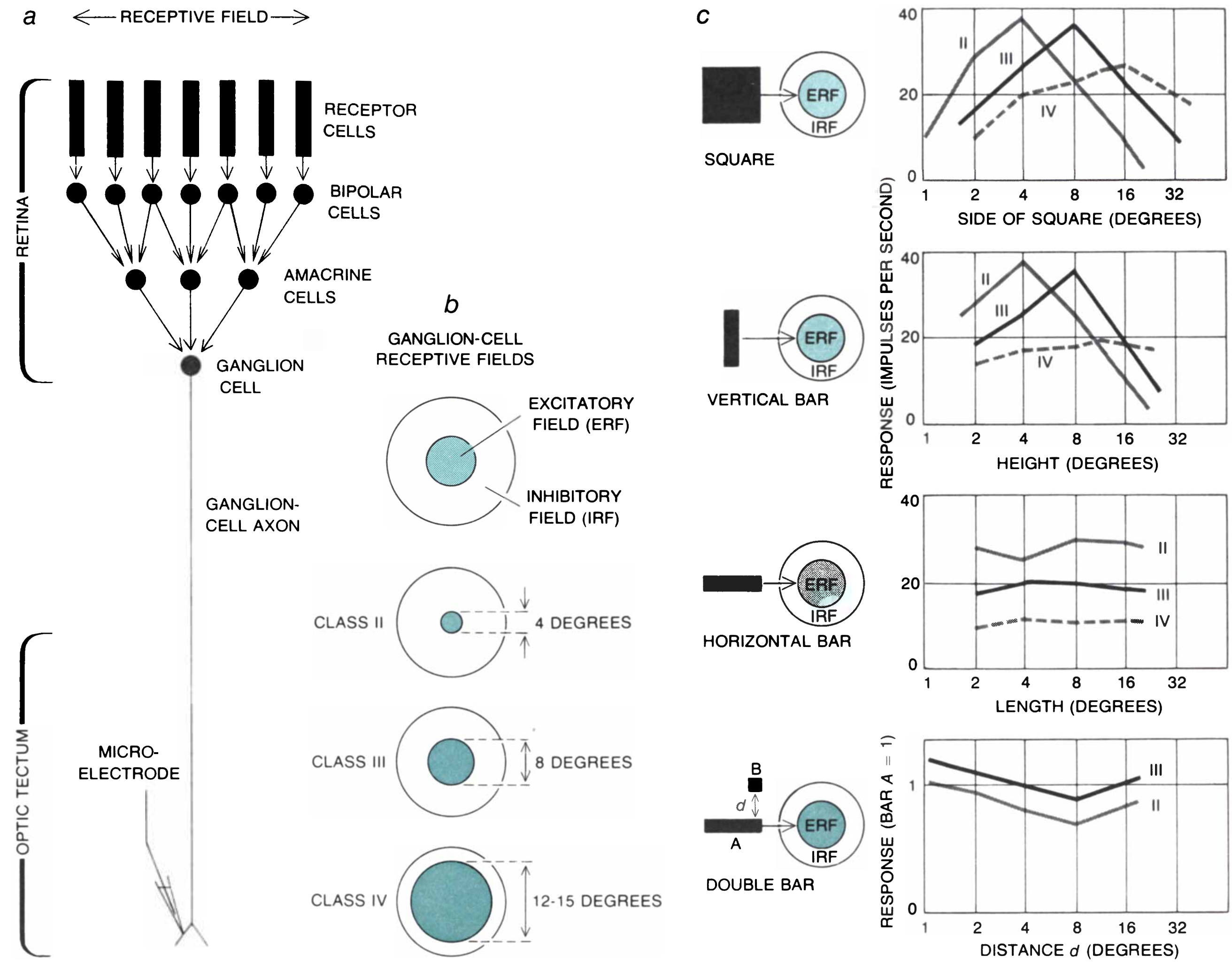
extension (making it in effect an *L*-shaped structure moving on its long side), it loses efficiency as a prey-catching stimulus. The inhibitory effect of the vertical extension depends on its distance from the horizontal element. If a second vertical extension is introduced, in effect making the stimulus a shallow *U*-shaped structure, the total configuration signifies "enemy." The ethological interpretation is that it symbolizes a "swarm," and in the toad's brain inhibitory interactions first restrain prey-catching behavior and then induce escape behavior.

For constant form and angular velocity the behavioral activity generally increases as the amount of contrast between stimulus and background increases. White objects moving against a black background are normally more attractive as prey than black objects on white; the latter, on the other hand, are more effective in eliciting avoidance behavior. When the size and contrast are held constant, behavioral activity increases with increasing angular velocity, reaching a maximum at between 20 and 30 degrees per second. Stationary objects usually elicit no prey-catching or avoidance response. The common critical feature for key stimuli representing both prey and enemy is movement, and the two kinds of stimulus are differentiated primarily on the basis of their form: extension of the object in the horizontal direction of the movement generally means prey, whereas extension perpendicular to the direction of the movement signifies "not prey" or "enemy."

What does the toad's eye tell the toad's brain? This question was first formulated for the frog and dealt with in the fascinating research of Jerome Y. Lettvin and his colleagues at the Massachusetts Institute of Technology, and was later investigated quantitatively by O.-J. Grüsser and his co-workers at the Free University of Berlin. To ask the question is to open the "black box" of the toad's brain, or at least to examine the brain functions that participate in transforming input from visual stimuli into relevant behavioral patterns. At this point I shall describe neurophysiological findings concerning whether it is in the retina of the toad's eye that the key stimuli "prey" and "enemy" are encoded.

In the toad retina there are three types of ganglion cells that send their fibers by way of the optic nerve to the structure called the optic tectum in the midbrain. One can record the action potentials, or nerve signals, from the ends of these fibers by introducing a microelectrode into the tectum. John E. Dowling, then at

© 1974 SCIENTIFIC AMERICAN, INC



NEURAL RESPONSES of the toad to the same objects were measured with recording electrodes. The electrodes recorded impulses at the terminals in the optic tectum of fibers from individual ganglion cells, the cells in the retina of the eye on which signals from the receptor cells converge via intermediate cells (*a*). Each ganglion cell has an excitatory receptive field surrounded by an inhibitory receptive field. The diameter of the excitatory fields and the strength of the inhibitory surrounds are different for each of three classes of ganglion cells (*b*). For the square object and the vertical bar (which the ganglion cell "confuses"), maximum activity is elicited when the size of the object matches the excitatory-field size of each type of ganglion cell (*c*). Horizontal length does not much affect these cells' response. Vertical extension of a horizontal bar has less effect on these cells than on behavior (*opposite page*).

the Johns Hopkins University School of Medicine, showed through electron micrography that in the frog (or toad) retina each ganglion cell is connected to a number of receptor cells by bipolar and amacrine cells. Each ganglion cell is thus fed information from a particular part of the animal's visual field. Lateral connections established by horizontal and amacrine cells play a role in determining the properties of this receptive field. In toads as well as frogs the field consists of a central circular excitatory receptive field immediately surrounded by an inhibitory receptive field. The movement of an object through the excitatory field elicits a ganglion-cell discharge, which is inhibited if another object is simultaneously moving through the inhibitory field. The three ganglion-cell types in the toad (as in the frog) differ in several characteristics, including in particular the diameter of their excitatory receptive fields: about four degrees for the so-called Class II ganglion cells, about eight degrees for Class III cells and from 12 to 15 degrees for Class IV cells. (Class I cells have been identified in frogs but not in toads.)

With microelectrodes we measure the rate of ganglion-cell discharge to see how it changes when objects (corresponding to those in the behavioral experiments described above) are moved through the receptive field of the cell. The impulse frequency increases with the length of the side of a square object until the length about equals the diameter of the excitatory field; then it decreases as the object becomes large enough to stimulate part of the surrounding inhibitory field. In accordance with the different sizes of the excitatory fields the maximum activation of each cell type is therefore elicited by objects of different sizes [*see illustration above*]. Extending a small square horizontally (making it a "worm") does not bring about any change in nerve-cell activation; this is in sharp contrast to the previously noted effect of extension on the behavioral response. The dependence of neuronal activation on the size of the stimulus is instead primarily a function of extension perpendicular to the direction of movement. Indeed, the discharge frequency is almost exactly the same in response to a narrow vertical bar as it is to a square with the same height as the bar. A retinal ganglion cell "confuses" the two stimuli—but the toad does not: the square excites behavioral activity and the bar inhibits it. When the object size is held constant, however, the dependence of the discharge rate on con-

© 1974 SCIENTIFIC AMERICAN, INC

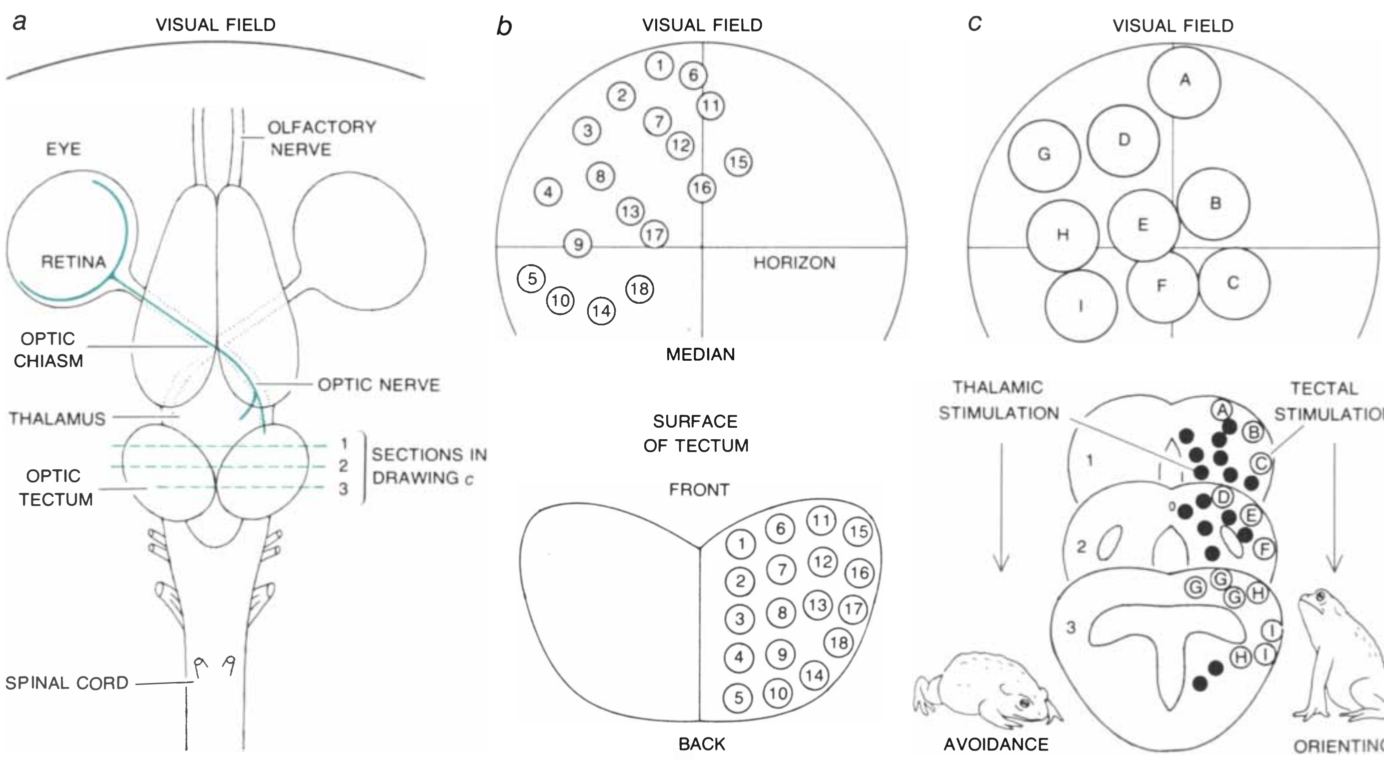
trast between stimulus and background and on angular velocity is the same as it is in the behavioral experiments.

In summary, it is clear that the first important operations on the visual input from a prey stimulus or a threatening one are performed by the toad retina. For any particular prey or enemy stimulus the behavioral response to velocity and background contrast seems to depend on information processing in the retina. The size-dependent excitatory and inhibitory processes, however, which were noted in the behavioral experiments and which play an essential role in pattern discrimination, cannot be traced to the influence of the excitatory and inhibitory fields of retinal ganglion cells. There are no retinal "worm-detectors" as distinct from "enemy-detectors." The differential analysis, and thus the behaviorally relevant interpretation of the stimulus, must be achieved in nerve-cell populations beyond the retinal level.

Since different characteristics are coded by any one type of ganglion cell the question becomes: Where is that coding interpreted? What tells the central nervous system whether an increased rate of ganglion-cell firing stems, for example, from an increase in stimulus-background contrast or from larger size? The differentiation can be made only if separate groups of cells receive different inputs from different optic-nerve fibers. In fact they do. The fibers of the optic nerve pass from each eye through the optic chiasm to the opposite side of the brain, ending in various parts of the forebrain and midbrain. Two of these destinations are of particular interest in our work. One, to which most optic-nerve fibers project, is in the surface layers of the optic tectum in the midbrain. The other is in the thalamus and the pretectal region of the diencephalon.

The optic tectum constitutes a localization system. In the tectum there is an exact topographical mapping of the retina and hence of the entire visual field. Movement of an object in a particular part of the visual field excites a corresponding region of the tectum, where the appropriate optic-nerve fibers terminate [*see illustration on these two pages*]. Recording from individual tectal neurons, or nerve cells, tells one how the individual retinal ganglion cells that excite them are reacting. In certain layers, for example, there are tectal neurons with excitatory receptive fields of about 10 to 27 degrees that are activated exclusively by moving objects. These neurons probably represent a localization system. This supposition is reinforced by experiments in which we stimulate the tectum of freely moving toads with trains of impulses delivered by means of an implanted electrode. Stimulation of a given region of the tectum always causes toads to turn toward a particular part of the visual field. Presumably the neurons we are thus activating have a direct connection with the animal's motor system, since (in contrast to the natural orienting movements made in response to a prey object) the electrically induced orienting is not disrupted by simultaneous presentation of a threatening object.

If the recording electrode is driven deeper into the tectum, it encounters neurons with larger receptive fields. Some of these cover the entire visual field on the opposite side, some the entire lower part of the field and some the entire field directly in front of the animal. Interestingly enough, all three types include the fixation point: the point of maximum visual acuity near the center of the visual field. The degree of activation of these three types of neurons could provide the toad with information about the



NEUROPHYSIOLOGICAL EXPERIMENTS yield data on the functions of different parts of the toad's visual system. Fibers of the retinal ganglion cells project primarily to the optic tectum and to the thalamus (*a*). The visual field of each eye is mapped (*b*), on a one-to-one basis (*numbers*), on the dorsal surface of the opposite side of the tectum. By the same token experimental electrical stimulation (*c*) of various parts of the tectum (*letters*) causes the toad to turn to corresponding parts of the visual field. On the other hand stimulation (*black disks*) of the thalamus, which partially underlies the tectum, causes the opposite action: avoidance, or turning away. As a recording electrode penetrates below the surface of the tectum it encounters successive populations of cells with differ-

© 1974 SCIENTIFIC AMERICAN, INC

location of a large object, since whenever the three types are excited simultaneously the object must be at the fixation point.

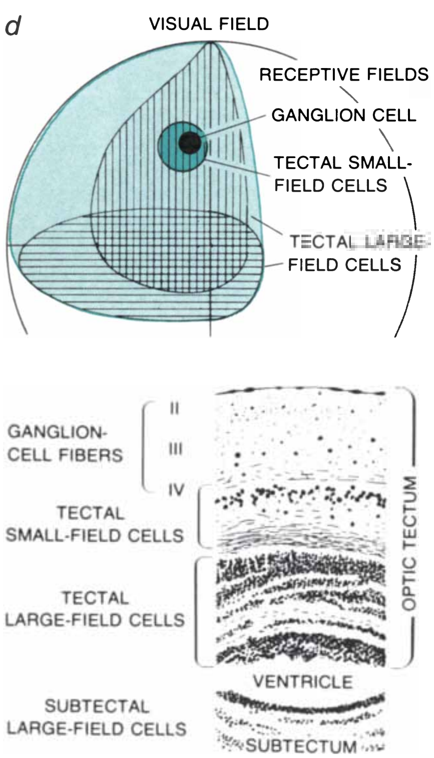
In natural situations the behavior of toads can be influenced by sensory modalities other than vision. If, for example, a beetle crosses the field of vision, the toad's orienting reaction can be either accelerated or retarded by simultaneous vibratory and tactile stimuli. Such results can be obtained in experiments if prey models are presented together with acoustic or tactile stimuli. The area for producing such changes in behavioral activity seems to be in the subtectal region, where multisensory integration is achieved. In the area below the third ventricle of the midbrain there are large-field neurons with fields similar to those of the large-field tectal cells. These subtectal neurons receive additional inputs from neurons excited by tactile and vibratory stimuli. The "mechanoreceptive" field of one of these bimodal neurons is always localized on the same side as the visual receptive field. The additional inputs from nonvisual neurons could serve to lower the threshold of a part of the visual field in which a visual stimulus is anticipated and thus in effect to raise the level of visual alertness.

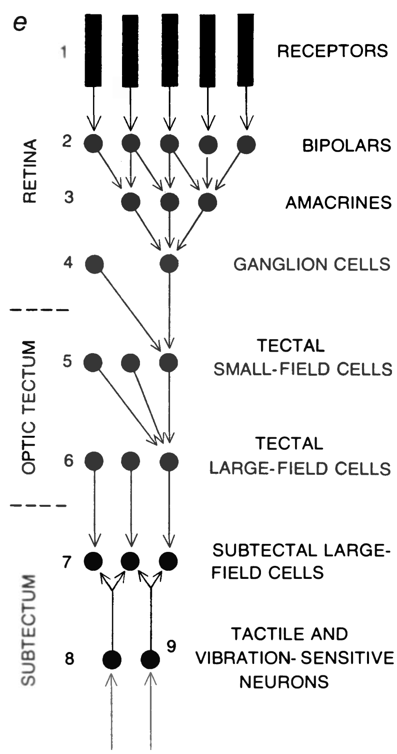
The optic tectum also comprises a neuronal system that processes behaviorally relevant aspects of moving stimuli [*see illustrations on next page*]. The cells have excitatory receptive fields about 27 degrees in diameter. Those designated Type I tectal neurons are activated mainly if the stimulus surface of an object moved through the receptive field is extended in the direction of movement; extension perpendicular to direction of movement does not have the same effect. Other cells, the Type II tectal neurons, differ from Type I neurons in that their discharge rate actually diminishes with surface extension perpendicular to direction of movement. The response of these neurons constitutes the key stimulus "prey." That is, they can presumably be considered the trigger system for the prey-catching response.

The thalamic-pretectal region, the second major destination of fibers from the retina, apparently provides what can be called a "caution" system. I have recently identified four main types of visually sensitive neurons in the toad's thalamus by means of single-cell recordings. They are activated respectively by four distinct stimulus situations: (1) movement of enemy objects extended perpendicularly to the direction of motion, excitatory receptive field of about 46 degrees; (2) movement of an object toward the toad, field about 90 degrees; (3) large stationary objects, field about 45 degrees; (4) stimulation of the balance sensors in the toad's ear by tilting. In general these thalamic neurons are activated principally in situations that tend to call for evasive movements—turning away from an enemy, sidestepping or compensating for tilting of the body. Brain-stimulation experiments support our feeling that the thalamic-pretectal region is one in which reactions can be assembled that lead to protective movements. Electrical stimulation of various sites in the region elicits the following reactions: closing of the eyelids, ducking, turning away, panicky springing away or tilting of the body.

We constructed a working hypothesis involving connections between the optic tectum and the thalamic-pretectal region: Electrical triggers in the tectum mainly elicit orienting, and triggers in the thalamic-pretectal region elicit avoidance. In a natural situation trigger impulses in particular layers of the tectum are evoked by small wormlike prey. Large objects extended perpendicularly to the direction of movement stimulate particular neurons in the thalamic-pretectal region, both directly through retinal inputs and indirectly by way of the optic tectum. These thalamic-pretectal neurons in turn inhibit the tectum and can also activate avoidance behavior [*see illustration on page 41*].

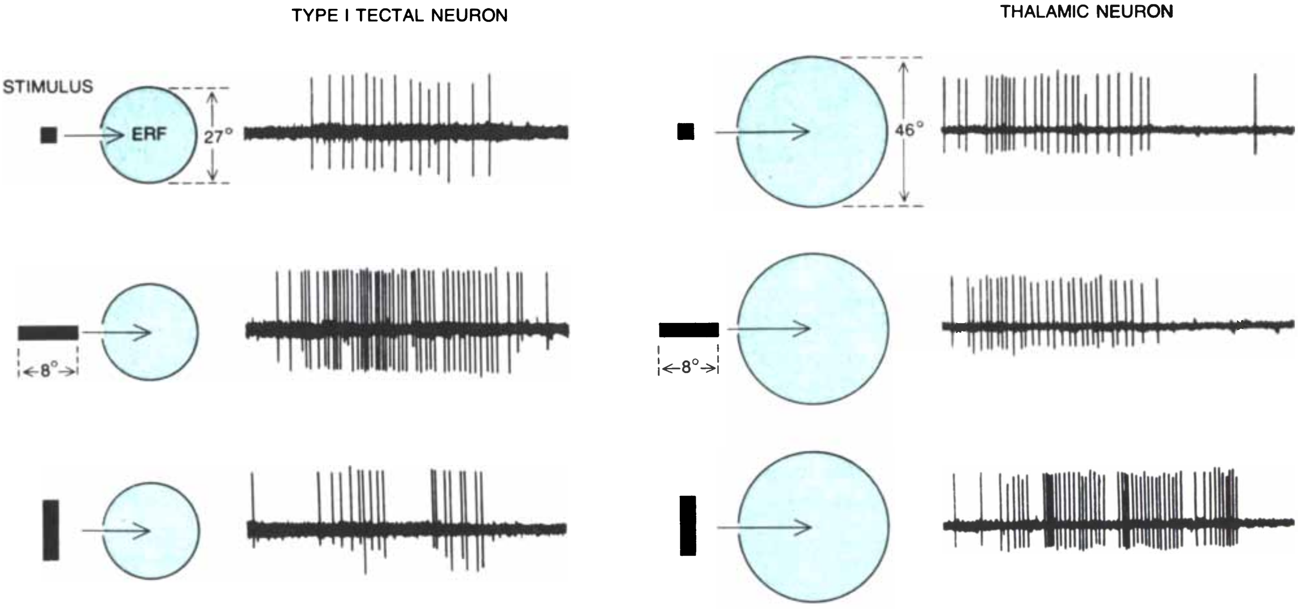
The existence of the postulated connections between the structures in the midbrain and the diencephalon has been demonstrated physiologically in two ways. One way is by direct electrical stimulation. Thalamic neurons that are sensitive to movement can also be activated by stimulation of points in the optic tectum. When the stimulating and recording electrodes are interchanged, the response of Type II neurons in the tectum to moving objects can be inhibited by the stimulation of cells in the thalamus. The other way is by surgical operation: if the optic tectum is removed, orienting movements are lost—and so are avoidance reactions, which is evidence for pathways from the tectum to the thalamus. If the thalamic-pretectal region is removed without damage to the tectum, then avoidance behavior is lost—and the orienting response is dramatically freed from inhibition even in the presence of enemy objects; this may be





ent receptive fields (*d*). There are small-field cells with fields a little larger than those of ganglion cells and, lower down, three kinds of large-field cells, each with different coverage (*color, horizontal hatching and vertical hatching*). A drawing based on a stained brain section indicates the layers at which each of these is found. The final drawing (*e*) relates the various cell populations and shows another layer of large-field cells that receive inputs from visual cells above them and also from cells that respond to tactile or vibratory stimuli.

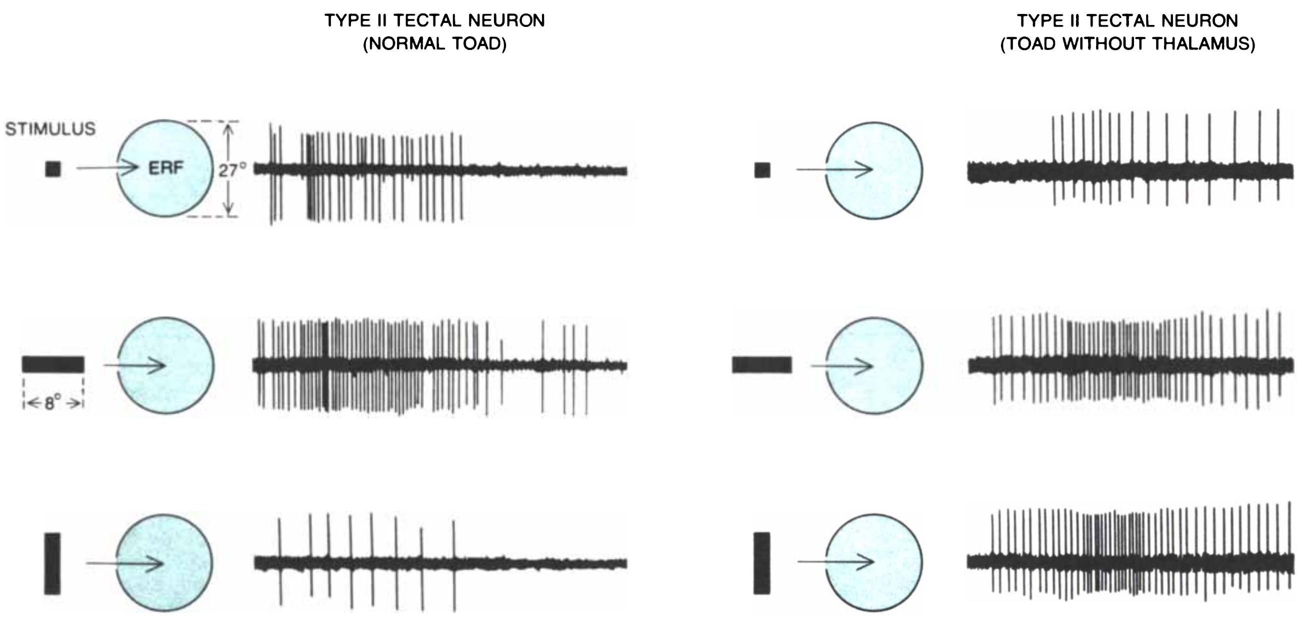
© 1974 SCIENTIFIC AMERICAN, INC



FEATURE DETECTION beyond the retinal level is accomplished by cells in the tectum and the thalamus. Recordings from individual cells indicate that tectal Type I neurons (*left*) are most activated if the object moving through the field is extended in the direction of movement. The cells in the thalamic area (*right*) respond most to an object extended perpendicularly to direction of movement.

evidence for the existence of inhibitory pathways from the thalamus to the optic tectum. In toads lacking the thalamic-pretectal region every moving stimulus elicits the orienting movements; the cautionary thalamic-pretectal system, which ordinarily allows orientation toward the stimulus only in behaviorally appropriate situations, is missing. If one lateral half of the thalamic-pretectal region is removed, the disinhibition extends to the entire visual field on the opposite side; small lesions in the thalamic-pretectal region affect only local small parts of the visual field. Quantitative experiments with toads lacking the thalamic-pretectal region make it clear that these animals cannot discriminate between stimuli that are behaviorally relevant and those that are irrelevant. The response of the Type II tectal cells to moving stimuli shows a similar "disinhibition" effect after thalamic-pretectal removal [*see illustrations on page 36 and below*].

The findings I have described so far suggest the following sequence of events: On the basis of retinal ganglion-cell input, the optic tectum tells the toad where in the visual field a stimulus is



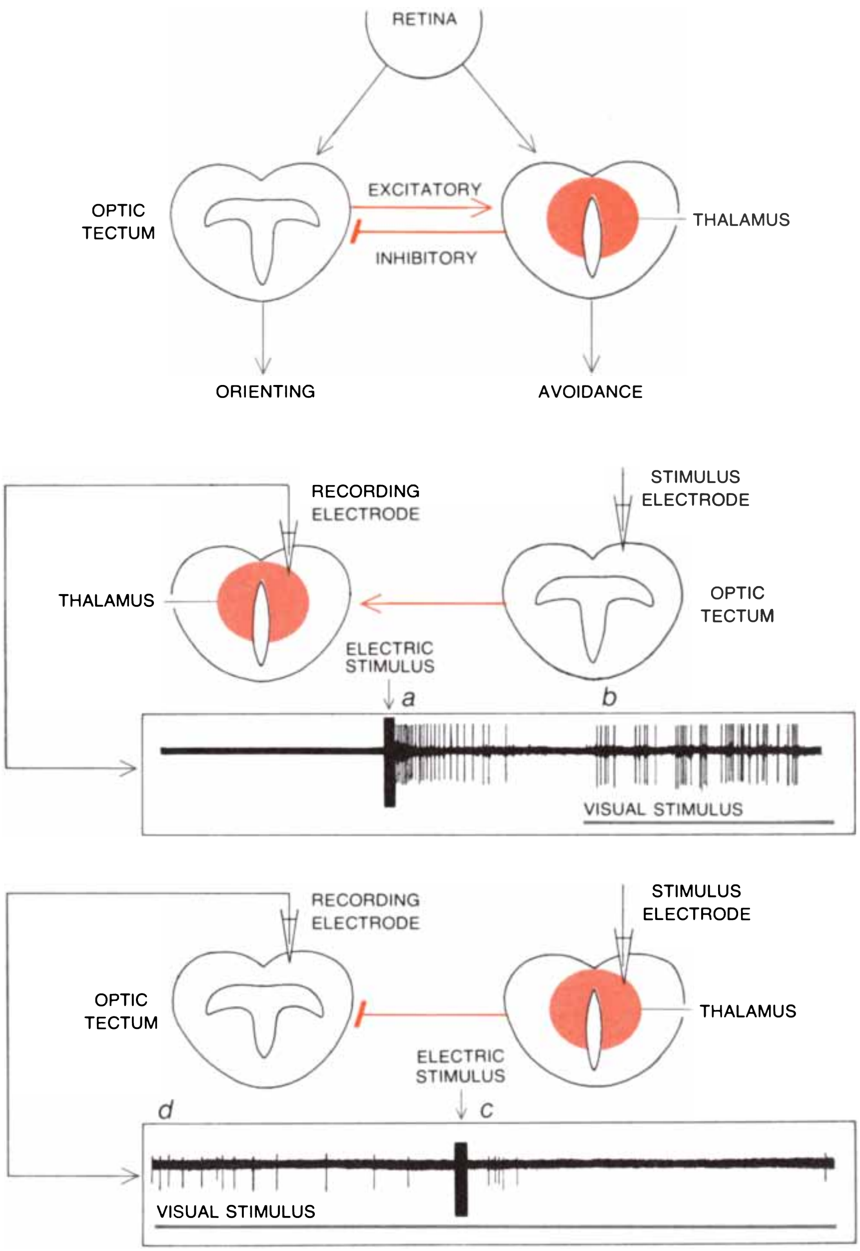
TRIGGER UNITS for the entire prey-catching response seem to be the Type II tectal neurons. In the normal toad (*left*) the cells are most activated by wormlike objects (*horizontal bar*). They are less activated (and the decrease is greater than in the case of Type I tectal neurons) by stimuli that in behavioral experiments are irrelevant for prey-catching (*vertical bar*). After removal of the thalamus, however (*right*), their response to those irrelevant stimuli is greatly increased, suggesting that the thalamic signal is inhibitory.

© 1974 SCIENTIFIC AMERICAN, INC

situated, how large it is, how strongly it contrasts with the background and how fast it is moving. The connections from the tectum to structures in the thalamic-pretectal region enable the toad to discern the significance to its behavior of the visual signals. The basic filtering process for the prey-enemy differentiation can be conceived of as passage through a series of "window discriminators" [*see illustration on next page*], each stage of which analyzes a particular aspect of the object in question. Each retinal ganglion cell acts as a vertical window that codes extension perpendicular to the direction of movement. The retinal analysis is repeated and amplified in the thalamic-pretectal region, where a neuron pool acts as another vertical window, this one with a certain minimum-response threshold. Extension in a horizontal direction is coded primarily by Type I tectal cells, which constitute a horizontal window. Type II tectal cells perform a summation, with signals arriving from the thalamic-pretectal region having an inhibitory effect and those from the Type I cells having an excitatory effect. The resultant signal acts as the trigger stimulus for the orienting movement. The triggering of avoidance behavior is probably achieved through the activation of still another pool of thalamic-pretectal neurons, the activation being proportional to an additive function of inputs from two of the window-discriminator pools.

One of the remarkable aspects of this system is a degree of plasticity, or changeability. During the summer months white prey objects moved against black backgrounds elicit orienting behavior much more effectively than do black objects against white. In fall and winter the situation is reversed, and at the same time the overall prey-catching activity of the toads decreases. Recently our recording electrodes revealed that the activation of single Class II ganglion cells in the retina exhibits this seasonal shift in white-black preference. In winter neurons with receptive fields in the lower part of the visual field are more strongly activated by black objects than by white ones; in the upper field the situation is reversed. In summer, however, the neurons whose receptive field is in the upper half of the visual field are activated primarily by black stimuli, whereas neurons receptive to the lower half of the visual field become more strongly activated by white stimuli and remain so until in the fall black stimuli again become dominant.

What is the biological significance of these observations? One can speculate that for toads, which are active at twilight, biologically important prey stimuli that appear in the lower half of the visual field are paler then their background; those in the upper part of the field, however, are for the most part relatively dark, or at least just as often dark as pale. Each of these contrast relations could be reflected in the sensitivity characteristics of the Class II retinal ganglion cells. With the approach of winter and the period of hibernation, toads stop catching prey. What makes them stop? One mechanism may be an inversion of ganglion-cell response characteristics, brought about by signals from the brain to the retina, such that the stimulus-background contrast relation is out of phase with the real world, making prey objects less visible. For the toad, in other words, identical objects appear to be dif-



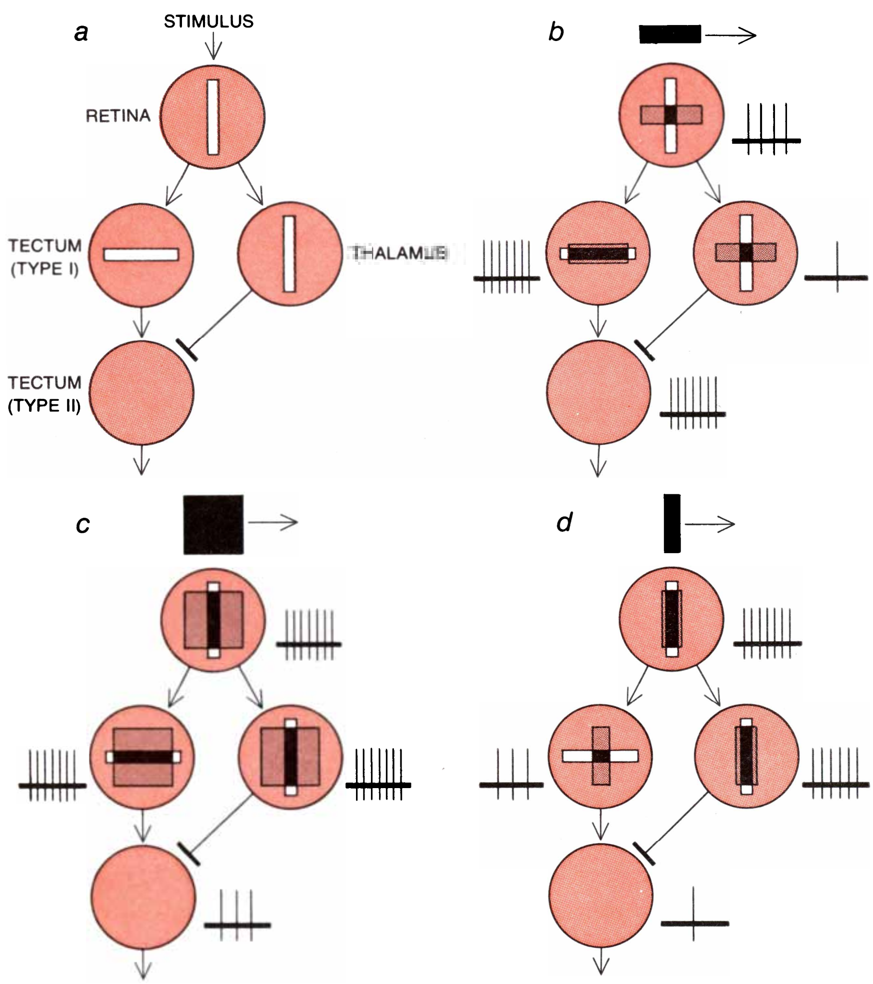
CONNECTIONS between the optic tectum and the thalamus were indicated by preceding experiments: signals from the retina excite both tectum and thalamus; subsequent impulses from Type I tectal neurons further excite cells in the thalamus, whereas signals from the thalamus inhibit activity in Type II tectal neurons. Two different kinds of motor activity are thereupon initiated by the two structures (*top*). Confirmatory evidence was obtained by electrical stimulation. Stimulation of the tectum (*middle*) elicits impulses (*a*) from cells in the thalamus that ordinarily respond to visual stimuli (*b*). Stimulation of the thalamus (*bottom*) inhibits (*c*) impulses normally elicited (*d*) in Type II neurons by moving objects.

© 1974 SCIENTIFIC AMERICAN, INC

ferent when they are seen in winter than when they are seen in summer, a reminder that an organism's picture of the environment is a product of its brain.

In contrast to the plasticity of the systems for the filtering and storage of information and for pattern recognition, brain mechanisms involved in instinctive actions are quite inflexible and do not adapt easily to changes in the stimulus situation. For each instinctive action in the behavioral repertory there is a preprogrammed "printed" neuronal circuit that coordinates the appropriate motor act—even if it becomes inappropriate! If such a circuit is triggered (either naturally or by electrode stimulation), the innate reaction proceeds automatically. Prey-catching behavior is a good example. On the basis of brain-stimulation experiments we believe the sequence of events controlling a natural orientation response is about as follows: A pattern is formed by a natural stimulus on a portion of the retina that is outside the fixation region; the retinal locus has a corresponding projection locus in the optic tectum. If the filtering process described above has identified the object as prey, then the appropriate neuronal system is activated. A value corresponding to the distance between the prey's locus on the retina and the fixation point is transferred to the toad's motor system. The result is orientation: a turning movement such that the retinal representation of the prey is brought to the fixation point. That triggers a locus in the optic tectum that corresponds to the fixation point. As soon as this triggering reaches a threshold value the rest of the prey-catching sequence is activated, quite independently of the result, or even of the short-term benefit to the animal, of such activation. For example, if an experimental prey object is removed at the instant when it is fixated by the toad, the entire normal prey-catching routine nevertheless proceeds. The toad snaps, gulps and wipes its mouth in spite of the "situational vacuum." The sequence is similar in its inevitability to what happens when the triggering region of the tectum is stimulated with an electrode.

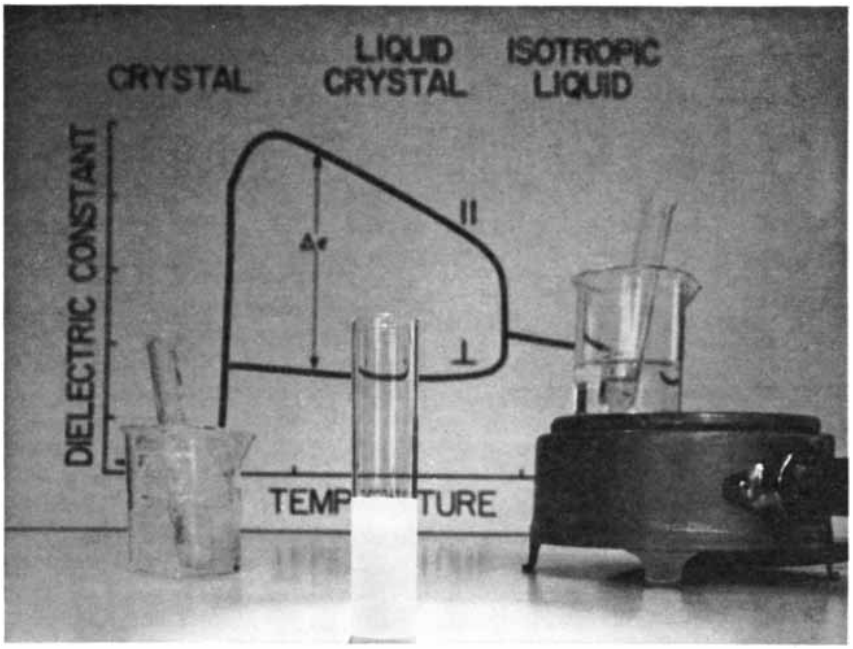
As for avoidance behavior, the results of thalamic stimulation indicate that it is controlled by a single master program. The response consists in a firm planting of the extremities on one side of the toad's body and a gathering together of the limbs on the opposite side. With the toad in this stationary, poised position the additional behavior patterns for correcting tilting of the body or making the various evasive movements can be readily incorporated.



IDENTIFICATION OF AN OBJECT AS PREY OR ENEMY is symbolized as a series of operations by "window discriminators" (*a*). A ganglion cell in the retina codes vertical extension (perpendicular to direction of movement), in effect responding to as much of a visual object as appears in a vertical window; extension beyond the window has an inhibitory effect. Cells in the thalamus do the same thing. In the tectum Type I cells code horizontal extension (in the direction of movement). Type II tectal cells sum the excitatory signal from Type I cells and the inhibitory signal from the thalamus, and the resultant signal triggers an orienting movement. At each stage the cell discharge depends on the relation between the object and the window that senses its extension either in or perpendicular to the direction of movement (*b*, *c*, *d*). (The cell-discharge patterns shown here are schematic.)

The evidence I have reviewed shows that in a lower vertebrate the neuronal processes for localization and identification of a visual signal and for releasing the associated instinctive motor responses are separated topographically but are intimately connected with one another. In the course of evolution the centers for two of these processes, visual localization and instinctive action, have apparently remained in about their original positions. They occupy the same areas of the brain, the tectum and the thalamus, in monkeys and cats as they do in toads. The organization of these parts of the brain, to which both neurophysiological and ethological methods have provided investigative access, shows remarkable constancy in all classes of vertebrates. That is not the case, however, for stimulus identification. In toads this process takes place primarily in the thalamic-pretectal region and also in the retina and the tectum. Mammals, however, underwent further evolution, corresponding to the importance of pattern recognition in the evolution of their behavior. A new substrate developed for two associated but highly specialized processes, filtering and storage of information: the visual cortex. From the investigations of Gerald E. Schneider at M.I.T. we learn that in this case ontogeny reflects phylogeny. In newborn hamsters subcortical pathways between the tectum and the thalamus are implicated in pattern discrimination. In adult animals, on the other hand, pattern discrimination takes place in the cortex.

© 1974 SCIENTIFIC AMERICAN, INC



News: For field-effect devices, EASTMAN 14080 is now ready to ship. At 18 μm spacing, a typical lot shows a threshold of only 1.5 volts rms for 60-Hz sine wave (or 2 volts dc), an *operating* temperature range of 0 to 75°C, and a dielectric anisotropy of +11.9 at 25°C. At 5 V rms, turn-on time at room temperature is about 100 ms; turn-off time, about 350 ms; contrast ratio, about 50:1.

For dynamic scattering devices, EASTMAN 14099 is equally ready, with the same 0 to 75°C operating temperature range. It aligns itself homeotropically on clean electrodes without additional magic–an important advantage. Typical threshold is 4 volts rms, 5 volts dc.

More details from George Grau at Organic Chemical Sales, Kodak, Rochester, N.Y. 14650 (716-325-2000, ext. 57288).

Glamourous, mysterious liquid crystals.

We make them.

Behind the mystery and glamour: Those who could reasonably be expected to phone Mr. Grau are probably interested in making display devices ranging from mass-marketed consumer goods to some of the more ambitious reaches of the engineering imagination.

Liquid-crystal technology has blazed up after smoldering quietly for most of the time since 1888. Organic chemists get to collaborate with electronic engineers through the medium of the marketplace, if not personally. For now, Kodak has set up its booth on the chemical side of the street.

The present generation of electronic engineers catch on fast to subjects they didn't necessarily concentrate on in school. In liquid-crystal work, one deals with the different forms and degrees of orderedness among molecules, ranging between the randomness of an ordinary isotropic liquid and the periodic architecture when it freezes to a crystal. Only certain compounds assume this mesomorphic state. Molecules of such compounds have a generally elongated shape.

The most highly ordered kind of liquid crystal, called smectic, where the molecules tier up in layers and the layers can only slide against each other, can be seen when water sits in contact with a bar of soap. In a nematic liquid crystal there are no tiers, but the molecules stay parallel. Possibilities exist for controlling their alignment. And this alignment controls what they do to light passing through. And that's what the action is mostly about in industry circles.

Field effect, one approach used to create a visible pattern in a thin layer of nematic material between patterned, transparent electrodes, amounts to electrically tuned birefringence. (We hope you didn't cut the lecture on birefringence in Physics 1.) Contrast between "on" and "off" is attained through various arrangements of light polarizers and quarter-wave retardation plates. Where and when the field is off, the molecules must line up parallel to each other and to the cell walls. This is called "homogeneous" alignment. Surface treatment to make it happen is protected by patents or trade secrecy.

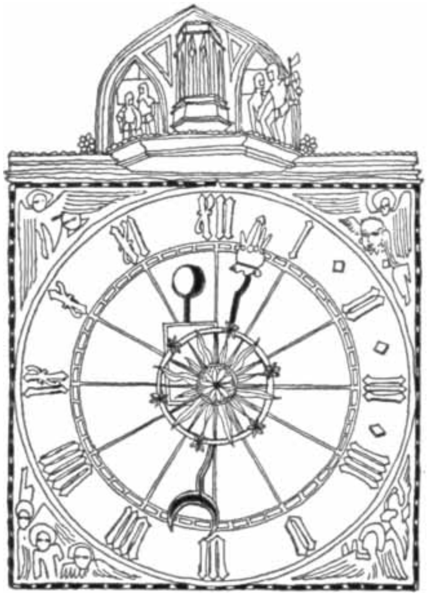
Dynamic scattering is the other approach. Here, in the "off" state, the alignment can be either homogeneous or perpendicular to cell walls. The latter is "homeotropic" in the lingo of the art. Either way, light passes straight through, and the layer looks clear. Where field is applied, then regardless of previous orientation, all the molecules strive for parallelism to the wall. But ions that have been incorporated in the mixture migrate, colliding with them to knock them every which way. The resulting optical inhomogeneity scatters light. These turbid areas can be made to look either darker or lighter than the clear areas, depending on directions of illumination and viewing. Appearance of the pattern is more sensitive to angle of view than with field effect. Power drain is more, because of the turbulence to be maintained. But you don't have to find (or license from somebody) a way to align the molecules. You just use EASTMAN 14099.

Say, if you've read this far you are probably interested enough to ask Dept. 412-L, Kodak, Rochester, N.Y. 14650 for Eastman Organic Chemical Bulletin, Vol. 45, No. 2 (1973), *where four scientists of the Kodak Research Laboratories, who write more like scientists than engineers, will take you farther into this than you'd want to go for just recreation. It has a 76-item bibliography. For our 3,281-item* Liquid Crystal Bibliography (Kodak *Publication JJ-193) on microfiche, make that Dept. 454 and send $25* (plus applicable state and local taxes).*

**Price subject to change without notice.*

© 1974 SCIENTIFIC AMERICAN, INC

SCIENCE AND THE CITIZEN



Recycling the Arms Race

Not quite two years after the formal treaty-signing ceremony in Moscow marking the conclusion of the first round of the strategic-arms-limitation talks between the U.S. and the U.S.S.R. ("SALT I"), an event heralded by the participants as a "historic first step" and a "major achievement" on the road toward ending the nuclear arms race, the two superpowers appear ready to embark on a new phase of the action-reaction cycle that has characterized the dynamics of the arms race throughout most of the postwar period. In the U.S. this impending development is manifested most clearly in the Administration's request, currently before Congress, for some $19 billion for "strategic forces and support operations" as part of the Department of Defense's record $85.8 billion budget for 1975. The projected spending for strategic arms, up some $1 billion from the current fiscal year, is described as being intended for two main purposes: to maintain and "modernize" the existing nuclear-weapons systems and to pursue an "enhanced" research-and-development program on a wide variety of "new strategic initiatives."

Under the first heading the Defense Department's 1975 budget request includes large amounts for continuing to convert its land-based and sea-based missiles to carry multiple independently targetable reentry vehicles (MIRV's). In a few years, when this conversion program is expected to be complete, the U.S. will have deployed a total of almost 10,000 nuclear warheads capable of hitting targets in the U.S.S.R. from long range. (U.S. forces in Europe are estimated to have thousands of additional nuclear weapons that could be used against targets in the U.S.S.R.) At roughly the same time the Russians, who began testing MIRV's on some of their land-based strategic missiles last August, are expected to have in their strategic arsenal a total of some 4,000 separately targetable nuclear warheads aimed at the U.S. In theory the U.S.S.R., which was granted numerical superiority in long-range missiles by the SALT I accords in part to compensate for the long U.S. lead in both MIRV technology and missile accuracy, could (by closing the "technology gap") at some point in the 1980's or later have more strategic nuclear warheads than the U.S. will have.

Citing this future possibility (and the hypothetical strategic advantage it is said to represent), Defense Department officials are now seeking a sharply increased research-and-development budget for 1975 to pay for work on an array of new strategic-weapons systems. Prominent among the requested funds are $2.6 billion for accelerating the development of the Trident missile-firing submarine (the projected successor to the Polaris-Poseidon) and $499 million for continuing the development of the B-1 long-range bomber (the projected successor to the B-52). Also included under this heading are $248 million for advanced research on MARV's, a "follow on" generation of highly accurate maneuverable reentry vehicles (see "Science and the Citizen," December, 1973); $125 million for developing a new "cruise" missile capable of being fired from either airplanes or submarines; $16 million for designing the propulsion system of a new "smaller" missile-firing submarine (to be named *Narwhal*), and $10 million to perfect a "command-data buffer system" that would enable the U.S. to switch any missile in its arsenal to a new target in 20 minutes (a procedure that now requires about 36 hours).

The last item is believed to be related to the Administration's new strategic doctrine of "nuclear flexibility," which seeks to give the President a "broader range of options" in responding to "limited" Russian nuclear attacks by placing somewhat greater emphasis on the targeting of certain military installations in the U.S.S.R. as well as population centers.

It has been suggested by some Administration spokesmen and others that the vast new array of projected weapons systems are intended at least in part as "bargaining chips" to be discarded by U.S. negotiators at the SALT II conference in order to gain concessions from the Russians. Whether the SALT II sessions, which resumed last month in Geneva, are any more successful than SALT I was in using this approach to halt the arms race, rather than simply to redirect it, remains to be seen.

Forty-two Reactors

At the end of 1973 the U.S. had 42 operable nuclear reactors capable of producing electricity on a commercial scale. Their combined capacity of 25,670 megawatts (25,670,000 kilowatts) constituted 5.6 percent of the country's electric-generating capacity. According to the Atomic Industrial Forum, Inc., which made the count in one of its periodic tabulations of reactors, "if this nuclear capability did not exist and had to be supplied instead by fossil-fired generation, it would require some 250 million barrels of oil or 57 million tons of coal annually."

The 42 operable reactors are distributed among 19 states. Illinois has seven; New York and Wisconsin have four each; Pennsylvania and South Carolina have three each; California, Connecticut, Florida, Massachusetts, Michigan, Minnesota and Virginia have two each, and Alabama, Colorado, Maine, Nebraska, New Jersey, Vermont and Washington have one each. The Atomic Industrial Forum also counts 56 reactors under construction, 101 under firm order and 14 "under letters of intent or options." A survey made in July, 1972, showed 25 operable reactors in 14 states.

The Hills of Mercury

Recent observations of Mercury by radar astronomers working with the 210-foot radio telescope at the Deep Space Station at Goldstone, Calif., have revealed that the surface of the planet has gentle hills and valleys and may also be pocked with craters like the moon and Mars. These findings may be amplified

© 1974 SCIENTIFIC AMERICAN, INC

by the high-resolution pictures that will be sent back by the *Mariner 10* spacecraft as it passes Mercury on March 29 after its swing around Venus.

A radar-astronomy group at the Jet Propulsion Laboratory of the California Institute of Technology, headed by Richard M. Goldstein, scanned the planet on 14 different dates. They used the 210-foot antenna to send out a 400-kilowatt signal at a wavelength of 12.5 centimeters; 10 minutes later the antenna received a radar echo of a fraction of a watt. A number of echoes were averaged over a period of several hours to yield a "snapshot" of a region about 600 kilometers on a side near Mercury's equator. The 14 snapshots produced a panorama of the planet from 12 degrees north of the equator to four degrees south.

Mercury seems considerably rougher than Venus, a finding that was foreshadowed by low-resolution radar observations made in 1969. The hills and valleys vary rather smoothly in elevation by about one kilometer, although one series of observations revealed a feature rising about 1,300 meters above a base 120 kilometers across. Circular features some 50 kilometers across and 700 meters deep appear to be craters. Two of the snapshots suggest that there may be several craters 500 kilometers across.

A more detailed view of these areas will possibly be provided by *Mariner 10* later in March. This spacecraft, the first to pass close to two planets, is the latest in the series that has revealed so much about the terrestrial (earthlike) bodies in the solar system [see "The Chemistry of the Solar System," by John S. Lewis, page 50].

Stock Component

Enzymes, the proteins that operate as catalysts in biochemical reactions, generally have a quite limited repertory: each mediates a particular chemical process at a particular site on a particular substrate. This specificity is a result of the distinctive structure of each enzyme's binding site or sites, where it interlocks with and becomes temporarily attached to other molecules. It now appears that certain of these structures have been incorporated into diverse enzymes wherever their function is required.

Michael G. Rossmann and his associates at Purdue University have mapped the three-dimensional structure of D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and lactate dehydrogenase (LDH), two enzymes that participate in glycolysis: the breakdown of carbohydrates to release energy. Both enzymes oxidize their substrates by removing hydrogen atoms, and both transfer these atoms to the same coenzyme: nicotinamide adenine dinucleotide (NAD). Writing in *Proceedings of the National Academy of Sciences*, Rossmann reports that in overall form the two enzymes are not closely related. In particular the sites where the substrates are bound to the enzymes are quite different, as would be expected, since the substrates themselves are dissimilar. The sites at which the coenzyme is bound, however, are substantially the same.

In order to perceive these relationships it was necessary to know not only the chemical composition of the enzymes but also how each molecule is folded to form a three-dimensional structure. Enzymes, like other proteins, consist primarily of amino acid units strung together in long sequences. The sequence can be determined by chemical methods; the three-dimensional structure is explored by X-ray diffraction. In this technique X rays are passed through a crystallized sample of the substance; from the resulting patterns it is possible to compute the position in space of each atom in the molecule.

Sites that bind NAD or related nucleotides are present in at least three other proteins: flavodoxin, liver alcohol dehydrogenase and *s*-malate dehydrogenase. In these molecules too the coenzyme binding sites appear to be variants of the same structure. More recently it has been discovered that two kinases—enzymes that control another part of the glycolytic pathway—also share this nucleotide-binding site.

The recurrence of essentially the same structure in all these enzymes suggests that they have a common ancestry; the considerable differences in their overall form, however, indicate that they must have diverged very early in the evolutionary process and have evolved independently since then.

Rossmann has suggested a possible scheme of genetic diversification that might have produced the five enzymes studied. In this evolutionary model flavodoxin separated from the other enzymes at a very early date. Liver alcohol dehydrogenase, GAPDH and LDH diverged somewhat later, and *s*-malate dehydrogenase developed later still, probably evolving from LDH.

Regardless of the evolutionary sequence, the nucleotide binding site appears to be a "stock" component that has been built into a variety of enzymes since primordial times. Rossmann writes: "The necessity of binding nucleotides to proteins in even the most primitive life forms suggests that this structure might be one of the earliest and most universal architectural units of proteins."

Seeing by Phosphene

Phosphenes are subjective sensations of light: specks, flickers or sometimes elaborate patterns that people "see" in the dark, usually after prolonged visual deprivation. Phosphenes can also be provoked, as by a blow on the head ("seeing stars"), pressure on the eyeball or electrical stimulation of the brain's visual cortex itself [see "Phosphenes," by Gerald Oster; SCIENTIFIC AMERICAN, February, 1970]. What if a meaningful pattern of phosphenes could be evoked on the visual cortex of a blind person, bypassing the nonfunctioning parts of the visual pathway? The possibility of developing a functional prosthesis based on cortical stimulation was first investigated by G. S. Brindley and W. S. Lewin of the University of Cambridge, who found that a blind person could see induced phosphenes. Their results have been confirmed and extended by investigators from the University of Utah and the University of Western Ontario, who first worked with patients undergoing brain surgery to determine what parts of the cortex were most susceptible to stimulation and how the phosphenes varied with pulse amplitude, frequency and other variables. Now that group has published the results of its experiments with two blind volunteers. In one patient actual pattern recognition was achieved, giving ground for hope that a rudimentary form of "artificial vision" may someday be possible.

W. H. Dobelle, M. G. Mladejovsky and J. P. Girvin describe their work in *Science*. Their volunteers were totally blind; one had been born with congenital cataracts and the other had been blinded in Vietnam. In each case an array of platinum-disk electrodes (64 electrodes, each a square millimeter in area, in an eight-by-eight array on three-millimeter centers, embedded in a Teflon ribbon) was placed in contact with the right occipital lobe of the brain during an operation under local anesthesia. First a map was prepared that related each electrode to the position in the visual field of the phosphene it produced. This was done by stimulating pairs of electrodes and asking the volunteer to describe their relative positions; a computer prepared a "best fit" map for a large number of electrodes. The map was displayed on a cathode ray tube. By touch-

© 1974 SCIENTIFIC AMERICAN, INC

ANTIQUE SCIENTIFIC INSTRUMENTS

Mater of Persian Astrolabe by 'Umar b. Dawlatshâh dated In 726 (A.D. 1325)

The above is one of the several hundred rare, and not so rare, instruments & early associated books to be found in our new catalog, including items of Astronomy, Dialling, Photography, Navigation, Surveying, & Microscopy.

Please send $1 ($2 for overseas airmail) for fully illustrated Catalog 108E (available early April).

HISTORICAL TECHNOLOGY
6 Mugford Street
Marblehead, Mass. 01945

Confidential borrowing by mail through:

TK Thomas Knight Financial Associates

A division of Zions Utah Bancorporation
Resources in excess of $500,000,000

Borrow up to $12,000 confidentially, by mail.

Flexible repayment plans... no penalty for prepayment

Immediate service
NATIONWIDE SERVICE BY MAIL

Write to:
W. H. Dobson
Suite 263 American Building
Salt Lake City, Utah 84101
Call collect: 801- 532-1093

ing certain positions on the display with a light pen the experimenters could then present simple patterns to the volunteers, one of whom was able repeatedly to identify such patterns as "a backward *L*."

While studying the possibility of information transfer in surgical patients and volunteers the Utah–Western Ontario group has also been studying, in animals, the biological acceptability of implanted materials and prolonged cortical stimulation, and has been developing feasible electronic technology. The group therefore has some confidence that a prototype device can be produced in a few years that will make it possible for a blind person to "see" well enough for him to detect nearby objects and to read normal printed matter. The plan is to put a miniature television camera in an artificial eye. Signals from the camera would be processed by a miniature computer, containing the individual's electrode-to-phosphene map in its memory, mounted in dummy eyeglasses. Signals from the computer would be transmitted to a receiver under the skull and thus to the electrode array.

Internal Pollution

It is well established that heavy cigarette smokers show a much higher incidence of heart attack and of death from heart disease than people who do not smoke. It has also been thought that nicotine, because of its potent effect on the cardiovascular system, was responsible for the development of atherosclerotic disease in smokers. It now seems that the real culprit is a much simpler compound: carbon monoxide.

Evidence for carbon monoxide as a cause of heart disease is reviewed in *Circulation*, a journal of the American Heart Association, by Poul Astrup of the Rigshospitalet in Copenhagen and Wilbert S. Aronow of the University of California College of Medicine at Irvine. Astrup reports that experimental animals exposed to carbon monoxide for several months show changes in the arterial walls that are indistinguishable from atherosclerosis. Degenerative changes in the myocardium, the middle muscle layer of the heart wall, are also found.

Carbon monoxide competes with oxygen for binding sites on hemoglobin, the substance in the blood that carries oxygen from the lungs to the cells of the body. In addition to removing some hemoglobin from the active transport of oxygen, the presence of carboxyhemoglobin (hemoglobin carrying carbon monoxide) in the blood causes oxygen to be more tightly bound to hemoglobin, making it more difficult for cells to obtain oxygen. Heart muscle, particularly the myocardium, is especially sensitive to oxygen deprivation.

"Nicotine," writes Astrup, "is probably of minor importance in comparison to carbon monoxide for the association between smoking and atherosclerosis, but it may have a synergistic effect.... The vasoconstrictory effects of nicotine may further impair oxygen supply to tissues of smokers with high carboxyhemoglobin concentrations."

Balancing Act

A large rock balanced on a small protuberance is an object of a certain wonder. Such rocks are not rare; for example, in Goblin Valley in southern Utah there are more than 1,000 of them. But how do the rocks stay balanced?

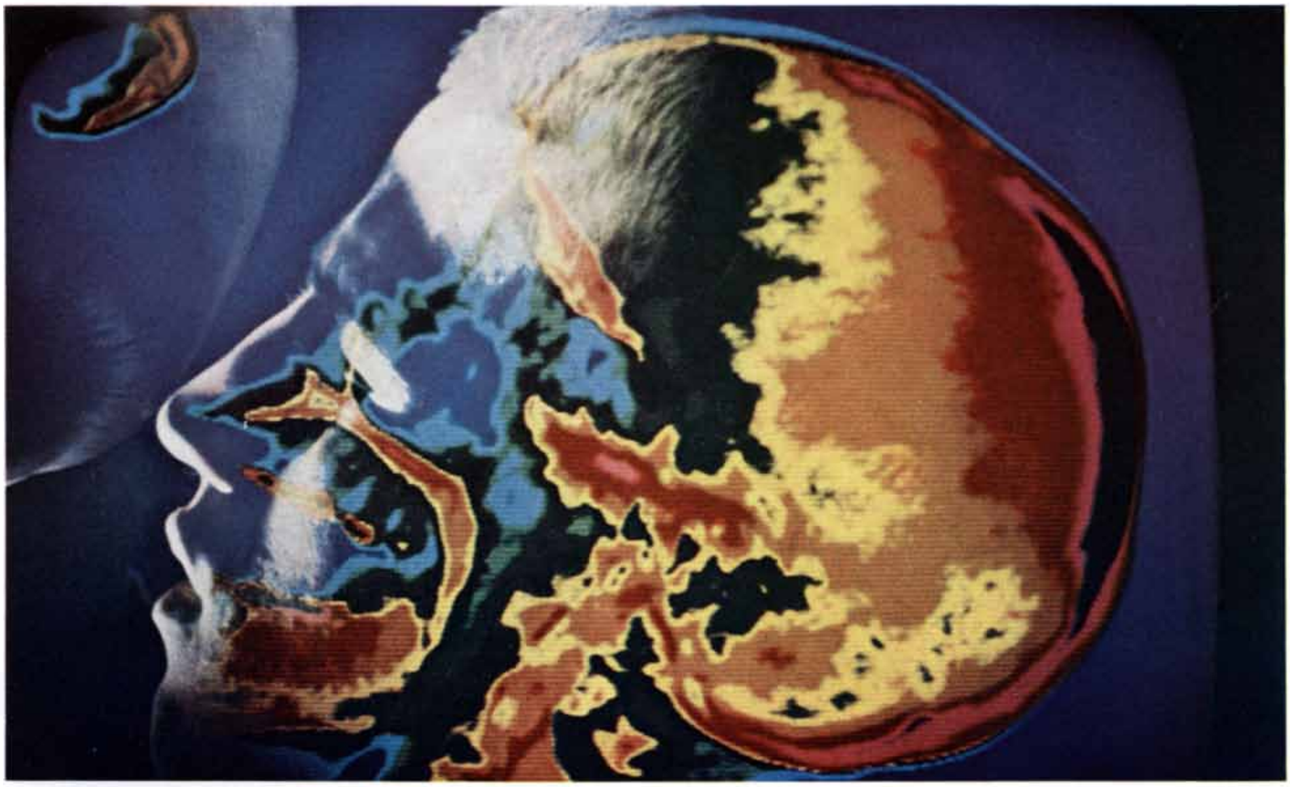
Balanced rocks originate when a bed of sediments is dissected by erosion until a column is formed. If the strata at the top of the column are harder than the strata farther down, erosion will whittle the softer rock down to a pillar narrower than the capstone.

Nothing about the erosion process, however, guarantees that the end product will be symmetrical, and so what keeps the capstone in place? Two investigators at Kansas State University, Wilson Tripp, an engineer, and Fredric C. Appl, whose specialty is rock mechanics, suggest that a dynamic process is responsible, that it starts when the capstone first begins to tilt in any direction and that the point of contact between the capstone and its supporting pillar continuously shifts, thereby remaining exactly under the capstone's center of gravity. The principle that underlies the process is simply that rock under the stress of compression is more resistant to erosion than unstressed rock.

When the capstone first begins to tilt, Tripp and Appl note, the movement will shift the stress of compression from one section of the supporting pillar to another. Thereafter the unstressed section will erode more rapidly than before and the stressed section will erode more slowly. Successive tilts in other directions will stress successive sections of the pillar, and the differential erosion that results will make the process self-leveling. As a consequence the capstone will remain poised on the pillar until the inevitable day when the area of contact becomes too small for the self-leveling to continue, and the balancing rock, ceasing its apparent defiance of the laws of statistical mechanics, crashes satisfyingly to the ground.

© 1974 SCIENTIFIC AMERICAN, INC

Why jump into a specialized field until you've studied all the possibilities?



If you're a new physician, you might not have decided on a field of specialization yet. That's easy to understand. But while you're deciding, you can be getting valuable experience and expert guidance towards that decision as a physician in the U.S. Air Force. Air Force physicians are exposed to many fields of medicine. They practice with highly professional people in modern facilities. And all Air Force physicians have seemingly unlimited opportunities for furthering their education and professional growth. Air Force graduate medical education programs are available. And because Air Force physicians participate in Group Practice (a collection of specialists in different fields complementing each other by expanding total capability) they all expand their individual ability and horizons with each new patient.

As an Air Force physician (and officer) you can apply for worldwide work assignments. The salary is excellent. And you get a 30-day paid vacation each and every year.

If you're still deciding on a medical specialty...or a future...look into the Air Force before you make a final decision. We think you'll like what you see. For more detailed information about Air Force opportunities in medicine please fill out and mail the coupon.

Air Force Opportunities
P.O. Box AF
Peoria, IL 61614

C-SA-34

Please send me more information on the Air Force Physician Program. I understand there is no obligation.

Name ______ (Please Print) Sex ☐ M ☐ F

Address ______

City ______

State ______ Zip ______ Phone ______

Soc. Sec. # ______ Date of Birth ______

Health Care at its best.
Air Force.

© 1974 SCIENTIFIC AMERICAN, INC

We'd like you

"The amount of energy locked up in a single piece of coal is incredible. We must waste no time putting it to use."

E. L. Wilson, Manager, Pilot Plant Operations, Synthetic Fuels Research Division, Exxon.

Coal is one form of energy America has in abundance. In fact, this nation has more coal than any other country in the free world. We have mined only 10% of what we have. More than 400 billion tons of commercially extractable coal are still in the ground.

In terms of the energy this coal contains, it outweighs our oil and natural gas put together. Experts estimate that at the present rate of coal consumption, our reserves could last about 250 years.

Exxon is working on ways to expand the use of coal. We are developing methods which may let America burn more of her vast reserves of high-sulfur coal without violating standards set for air and water quality. We are also developing economical ways to turn coal into gas and oil.

Gas made from coal.

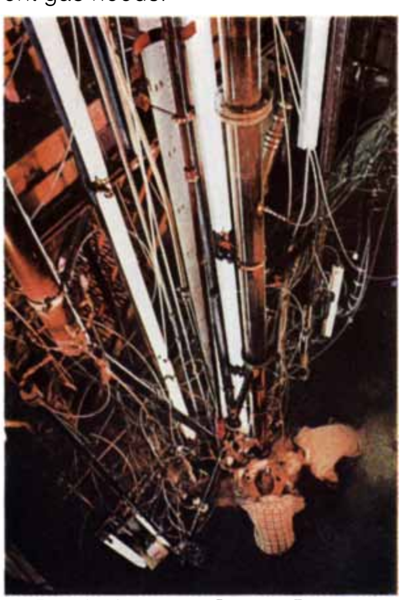
Chemically speaking, coal is similar to both natural gas and crude oil. All three are products of decayed plant or animal matter compressed into hydrocarbons over millions of years.

For years commercial plants have been gasifying coal, producing a low-energy fuel for domestic and limited industrial use. But this gas is expensive to produce, and the available processes have not been commercially proved on certain American coals.

For several years, Exxon has been developing a process to gasify the different types of American coal. Today at a pilot plant in Baytown, Texas, we are perfecting that process. High-sulfur coal from the Midwest and low-sulfur coal from the Rocky Mountain States are both being converted into raw gas.

We hope our gas will be less expensive to produce, bringing closer the day when synthetic gas can be made commercially from a wide variety of American coals.

Ten years from now, we expect plants, each producing 250 million cubic feet of synthetic gas a day, will be operating. Each of those plants could meet Cleveland's present gas needs.



At Exxon's pilot plant in Baytown, Texas, various American coals are being converted into cleaner burning synthetic gas.

Oil and gasoline made from coal.

During World War II most of Germany's planes and tanks ran on gasoline made from coal. But the problems were high cost and a relatively small yield of hydrocarbon liquid.

Now, Exxon is developing a coal liquefaction process which promises to be more efficient and less expensive than the World War II process. It also would be more flexible and reliable than other processes now being developed.

So, at another pilot plant in Texas we are converting low- and high-sulfur coal into synthetic crude oil. This oil can be refined into gasoline and other products with today's technology.

Sulfur: A major problem.

The largest potential user of coal today is industry, especially our nation's electric power plants.

Unfortunately, coal from most mines in our Appalachian and Midwestern states contains a lot of sulfur. When burned, it pollutes the air with sulfur oxides. This is a key reason why coal has not been more widely used in recent years.

Sulfur: Two possible solutions.

Exxon is now working on two processes to solve the sulfur problem.

One, being devised for the U.S. Government, would reduce the formation of sulfur oxides *while* the coal is burning.

The other process is being devel-

© 1974 SCIENTIFIC AMERICAN, INC

to know

E. L. Wilson holds a piece of high-sulfur coal from Exxon's Monterey mine in Illinois, and a bottle of synthetic crude oil made from this type of coal. This oil can be refined into gasoline and other products

oped with a major power plant builder and several electric utilities. It would remove most of the sulfur oxides from the flue gas *after* combustion but before the gas escapes from the stack.

Heat and pressure applied in the right way can convert a 4-pound piece of coal into about 1 quart of synthetic crude oil or about 32 cubic feet of synthetic gas.

When these processes—or others like them—prove commercially successful, they will enable industry to burn much more coal, freeing large quantities of oil and natural gas for other uses.

Coal's future.

The long-range success of Exxon's gasification, liquefaction and desulfurization projects depends on improving the technology and getting the costs down. As these problems are solved America will need more and more coal.

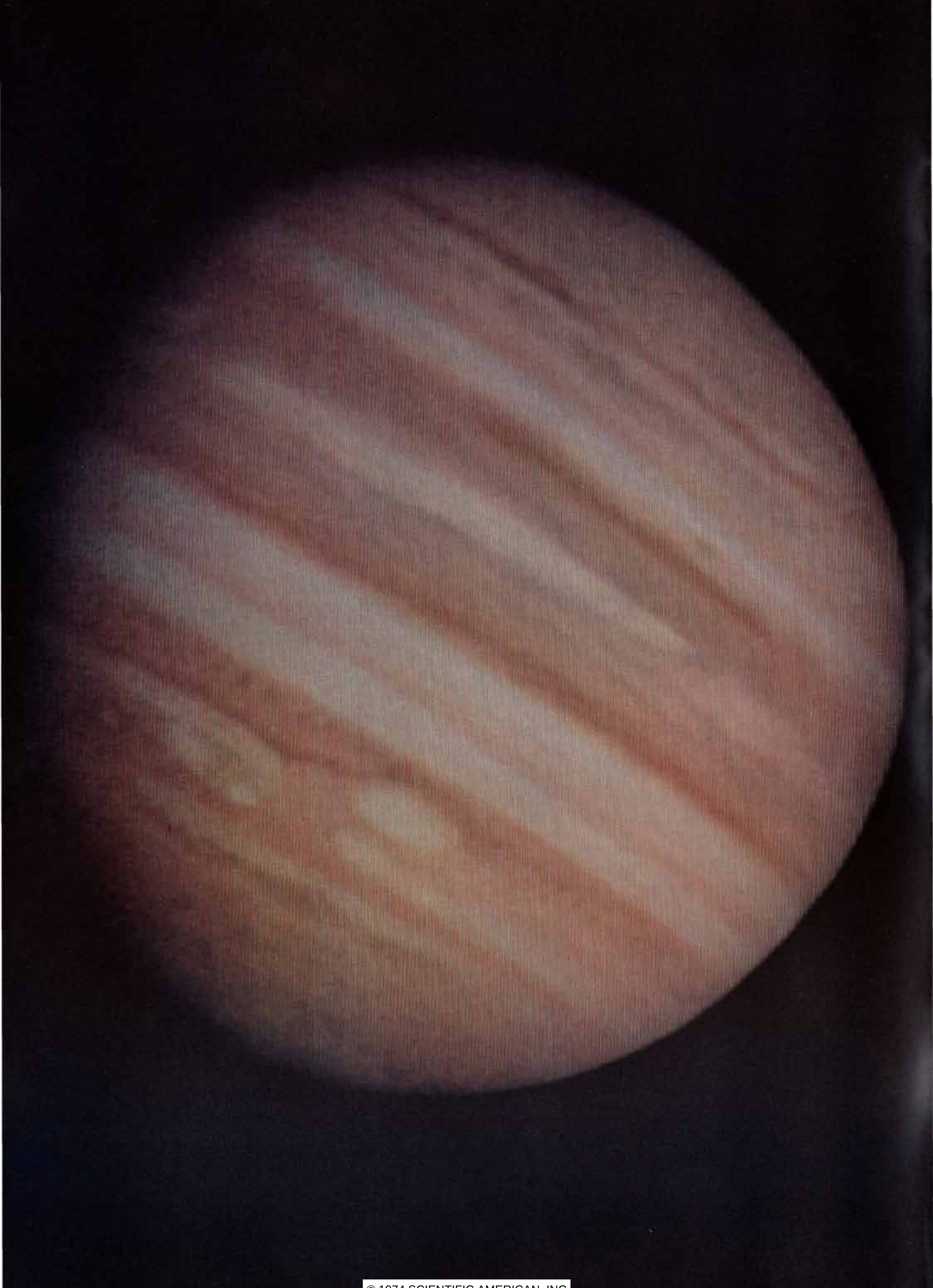
Last year, U.S. coal mines produced 603 million tons. If coal is to play a key role in filling the energy gap, production will have to more than double by 1985, and continue to increase rapidly thereafter. This means that environmental questions, primarily those of strip mining, must be resolved.

Exxon has already invested tens of millions of dollars in coal conversion and desulfurization research. In the years ahead we plan to spend many times this amount.

Coal is one of America's most important natural resources. We must waste no time putting it to use.

EXXON

© 1974 SCIENTIFIC AMERICAN, INC



© 1974 SCIENTIFIC AMERICAN, INC

THE CHEMISTRY OF THE SOLAR SYSTEM

The sun, the planets and other bodies in the system formed out of a cloud of dust and gas. What processes in the cloud could account for the present composition of these objects?

by John S. Lewis

The era of space exploration, with the landing of spacecraft on the moon and Venus and the flyby missions to Mars, Venus and Jupiter and now to Mercury and Saturn as well, has greatly enlarged our knowledge of the composition of the planets and satellites and the evolution of the solar system. Our modern understanding of the composition of the solar system nonetheless has a substantial history. It began in the 1930's with the work of Rupert Wildt on the physics and chemistry of the Jovian planets: Jupiter, Saturn, Uranus and Neptune. Wildt's investigations revealed that the atmosphere of Jupiter contains large amounts of the gases ammonia (NH_3) and methane (CH_4) and gave strong reason to suspect that all the Jovian planets consisted largely of hydrogen.

Around 1950 there was a second period of vigorous inquiry into the chemical nature of the solar system. At that time Harrison Brown proposed that the objects in the solar system could be divided into three classes according to their density and their chemical composition. The first class was the rocky objects such as the terrestrial (earthlike) planets, their satellites, the asteroids and the meteoroids. The second class was the objects consisting of both rocky and icy materials such as the nuclei of comets and the satellites of the outer planets. The third class was the sun and the Jovian planets, which consist mainly of matter in the gaseous state. At the same time that Brown suggested these classes Gerard P. Kuiper and Harold C. Urey were working on complex theories of the origin of the solar system. Urey stressed the importance of meteorites as clues to the origin of the planets, and he calculated the stability of the chemical equilibrium of numerous meteoritic minerals. Since that time the detailed study of the rocks brought back from the moon, the spectra obtained of comets and the numerous photographs made from spacecraft of the moon and the planets, together with what we know about the geology of the earth, have all contributed to the current picture of the solar system and its origin.

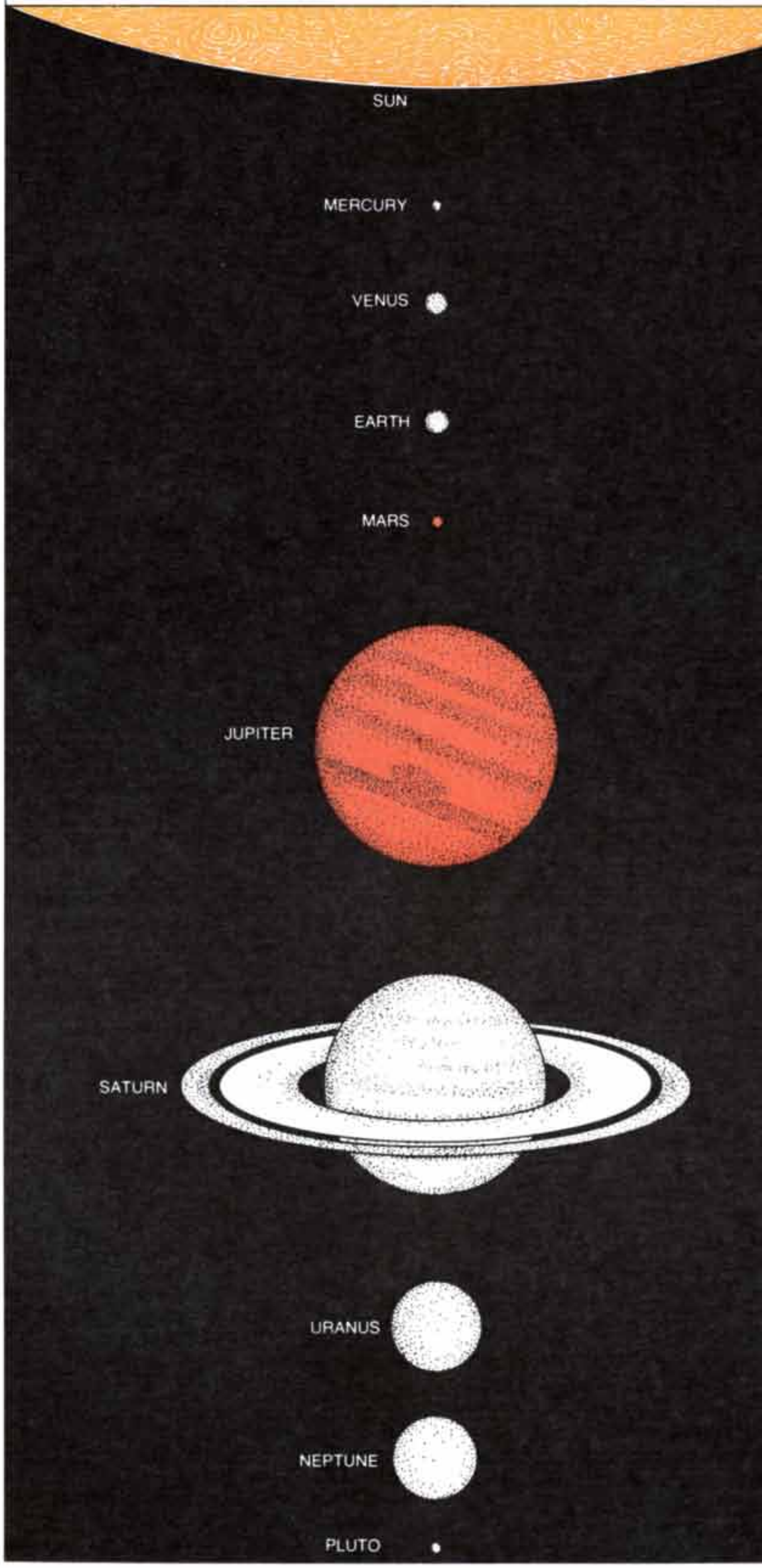
The Solar Nebula

The solar system is not a homogeneous mixture of chemical elements. Moving outward from the sun, there is a general trend for the abundance of the volatile, or easily evaporated, elements to increase with respect to the abundance of the nonvolatile elements. That general trend appears to be evidence for the hypothesis that the Jovian planets and the other bodies in the outer solar system formed at low temperatures and the terrestrial planets formed at high temperatures. According to this hypothesis, the solar system formed out of the solar nebula: a large cloud of dust and gas. The center of the nebula was hot (several thousand degrees Kelvin) and the edges were cold (a few tens of degrees K.).

The difference in temperature meant that the composition of the dust must have varied radically with distance from the center. Close to the protosun in the center all materials would have been totally evaporated. At a distance of perhaps 20 million miles from the protosun, a fifth of the way to the present orbit of the earth, a very few nonvolatile materials could have condensed into solid particles. From that point out to the present location of the asteroid belt the dust would have been composed primarily of grains of rocky material with only a limited content of volatile substances. Beyond the location of the asteroid belt the nebula would have been cold enough to allow the volatile substances such as water, ammonia and methane to freeze into

THE PLANET JUPITER, seen here from the spacecraft *Pioneer 10*, is composed of much the same elements in the same proportion as the sun. It is a representative of one of the three compositional classes of objects in the solar system. This class includes the sun and Saturn. The other two classes are the rocky objects such as the terrestrial (earthlike) planets and the objects composed of both rocky and icy material, such as the nuclei of comets. In this picture the cloud belts of Jupiter are resolved in a wealth of detail invisible from the earth. The planet appears gibbous because of the position of the spacecraft with respect to the planet and to the sun, a view that can never be seen from the earth. The terminator, the line between the sunlit hemisphere and the dark one, is at the left of the image. The picture was made at a distance of 2,020,000 kilometers from the planet at 20:58 Universal Time on December 2, 1973, about 30 hours before the spacecraft's closest approach of 130,000 kilometers. It was made with the imaging photopolarimeter experiment of the University of Arizona. The photopolarimeter consists of a telescope with an aperture of 2.5 centimeters (one inch) coupled to an optical system that splits the light from the planet into a blue channel and a red one. The device scanned the planet as the spacecraft spun on its axis 4.8 times per minute, building up the image in a raster pattern somewhat like the image on a television set. The photopolarimeter alternated between the blue channel and the red channel every half millisecond and telemetered the raw data for each channel to the earth. At the University of Arizona the image from each channel was rectified to reduce distortion and increase its sharpness; then the two images were combined to produce a color composite.

© 1974 SCIENTIFIC AMERICAN, INC



OBJECT	DISTANCE FROM SUN (MILLIONS OF MILES)	DIAMETER (THOUSANDS OF MILES)	MASS (EARTH = 1)
SUN	–	867	343,000
MERCURY	36	3.0	.1
VENUS	67	7.6	.8
EARTH	93	7.9	1.0
MARS	142	4.2	.1
JUPITER	486	89	317.8
SATURN	892	76	95.2
URANUS	1,790	30	14.5
NEPTUNE	2,810	28	17.2
PLUTO	3,780	3.7	<.2?

© 1974 SCIENTIFIC AMERICAN, INC

NUMBER OF MOONS	DENSITY (WATER = 1)	ROTATION PERIOD (DAYS)	REVOLUTION AROUND SUN (YEARS)	AVERAGE TEMPERATURE (DEGREES K.)	GRAVITY (EARTH =1)
–	1.4	27	–	5,800	28
0	5.4	55	.24	~ 600	.37
0	5.1	–243	.62	750	.89
1	5.5	1	1.00	180	1.00
2	3.9	1	1.88	140	.38
12	1.3	.4	11.86	128	2.65
10	.7	.4	29.48	105	1.14
5	1.6	.4	84.01	70	.96
2	2.3	.6	164.79	55	1.53
0	?	6.4	248.4	?	?

their ices, thus forming solid particles.

It is believed the sun formed when the center of the heated solar nebula became unstable and collapsed under the influence of its own gravity. The massive Jovian planets could have been formed by the same process. Alternatively they could have begun with a rocky and icy core so large that its gravitational attraction captured large masses of undifferentiated gaseous material. On this basis the similarity of their composition to the composition of the sun is easily understood.

The composition of the terrestrial planets and similar objects varies markedly with distance from the sun. There is a simple conceptual model, the equilibrium-condensation hypothesis, that can account for these differences. Let us assume that the chemical equilibrium between the dust and the gas in the solar nebula governed the composition of the solid materials that eventually accumulated into the planets, their satellites, the asteroids and the comets. On that basis one can calculate how the chemical composition of the condensed matter depends on its temperature at its formation. For this purpose one additional fact is needed: the relative abundances of the various chemical elements of which the primordial nebula was composed.

Our knowledge of the abundances of most elements in the galaxy is derived almost exclusively from the study of bodies in the solar system. The sun, however, is so much like most other stars that the relative amounts of the various elements in the solar system can be assumed to be typical of the large majority of stellar systems. Observations of the sun reveal that 99.996 percent of the mass of any sample of its material would consist of 15 elements. The least abundant element of the 15 (nickel) would still be commoner than all the other 80-odd elements put together.

There is a second model of how the terrestrial planets formed out of the solar nebula. The equilibrium-conden-

DATA ON THE PLANETS, incorporating information from various spacecraft missions as well as from ground-based observations, are displayed next to drawings of the planets that illustrate their relative sizes. Minus sign in front of rotation period of Venus indicates that the planet rotates in a direction opposite to the direction in which all other planets rotate. Tenth moon of Saturn, named Janus, was discovered in 1966

© 1974 SCIENTIFIC AMERICAN, INC

DEGREES KELVIN	EQUILIBRIUM-CONDENSATION MODEL	INHOMOGENEOUS-ACCRETION MODEL
1,600	1. Condensation of refractory oxides such as calcium oxide (CaO) and aluminum oxide (Al_2O_3) and also of titanium oxide and the rare-earth oxides	1. Condensation of refractory oxides such as calcium oxide (CaO) and aluminum oxide (Al_2O_3) and also of titanium oxide and the rare-earth oxides
1,300	2. Condensation of metallic iron-nickel alloy	2. Condensation of metallic iron-nickel alloy
1,200	3. Condensation of the mineral enstatite ($MgSiO_3$)	3. Condensation of the mineral enstatite ($MgSiO_3$)
1,000	4. Reaction of sodium (Na) with aluminum oxide and silicates to make feldspar and related minerals, and the deposition of potassium and the other alkali metals	4. Condensation of sodium oxide (Na_2O) and the other alkali-metal oxides at about 800 degrees K.
680	5. Reaction of hydrogen sulfide gas (H_2S) with metallic iron to make the sulfide mineral troilite (FeS)	
1,200 – 490	6. Progressive oxidation of the remaining metallic iron to ferrous oxide (FeO), which in turn reacts with enstatite to make olivine (Fe_2SiO_4 and Mg_2SiO_4)	
550	7 Combination of water vapor (H_2O) with calcium-bearing minerals to make tremolite	
425	8. Combination of water vapor with olivine to make serpentine	
175	9. Condensation of water ice	5. Condensation of water ice (H_2O)
150	10. Reaction of ammonia gas (NH_3) with water ice to make the solid hydrate $NH_3 \cdot H_2O$	6. Condensation of ammonium hydrosulfide (NH_4SH)
120	11 Partial reaction of methane gas (CH_4) with water ice to make the solid hydrate CH_4 $7H_2O$	7 Condensation of ammonia ice (NH_3)
65	12. Condensation of argon (Ar) and leftover methane gas into solid argon and methane	8. Condensation of solid argon (Ar) and methane
20	13. Condensation of neon (Ne) and hydrogen, leading to 75-percent-complete condensation of solar materials	9. Condensation of neon (Ne) and hydrogen, leading to 75-percent-complete condensation of solar materials
~1	14. Condensation of helium (He) into liquid	10. Condensation of helium (He) into liquid

MAJOR REACTIONS that would have occurred in the formation of the solar system according to two different hypotheses are shown with respect to temperature in the primordial solar nebula. The equilibrium-condensation model assumes that the nebula cooled very slowly as the planets were forming. The inhomogeneous-accretion model assumes just the opposite: that the nebula cooled quickly with respect to the rate at which the planets were forming. The two assumptions result in radically different predictions for the composition of the planets and satellites. In the author's view the equilibrium-condensation model seems to describe the planets of the solar system more accurately. Color area indicates two steps of each hypothesis that probably did not occur because the temperature never fell that low.

sation model assumes that there were no substantial changes in the temperature of the nebula at the location of planets that were accreting solid material onto their surface. Perhaps the material accreted rapidly in comparison with the rate at which the nebula was cooling. Or perhaps the composition of the solid material was determined by the temperature when it had last reacted with the gases of the nebula; after that time the gases could have dissipated and the solids could have accreted slowly onto the planets, each of which would have been homogeneous in composition. In either case the chemical equilibrium between the gas and the dust would determine the composition of both the gases and the condensates present at any time.

Comparison of the Two Models

The second model, the inhomogeneous-accretion hypothesis, is exactly the opposite of the first. Here the solar nebula cooled rapidly in comparison with the rate at which the planets accreted solid material. Since the composition of the material would have been determined by its temperature its accretion onto a planet could have given rise to onionlike layers of different condensates. It is important to add that the two hypotheses are extreme cases; the truth may lie anywhere in between them. Since we have no a priori grounds for choosing between them, however, we should understand the consequences of both extremes.

The results of the two assumptions are quite different [*see illustration at left*]. If one starts with a temperature of 2,000 degrees K. in the solar nebula, the equilibrium-condensation hypothesis would lead first to the condensation of refractory compounds containing calcium oxide (CaO), aluminum oxide (Al_2O_3), rare-earth oxides and so on. Then when the nebula had cooled to about 1,500 degrees, a metallic iron-nickel alloy similar to the one found in meteorites would condense. That step would be followed by the condensation of enstatite ($MgSiO_3$) and then by the formation of various minerals such as feldspar through the deposition of sodium, potassium and the other alkali metals. As the temperature dropped to 680 degrees metallic iron would be corroded by hydrogen sulfide gas (H_2S) to make the mineral troilite (FeS); the remaining iron would be progressively oxidized to form minerals such as olivine, which would combine with water

© 1974 SCIENTIFIC AMERICAN, INC

vapor at a still lower temperature to yield serpentine. Finally, as the temperature of the nebula dropped below 170 degrees, water vapor would condense as ice. The ice would later react with ammonia gas to make a hydrate, written $NH_3 \cdot H_2O$. At 100 degrees water ice would also combine with some of the methane gas in the nebula to yield another hydrate: $CH_4 \cdot 7H_2O$. Eventually argon and the leftover methane gas would freeze out at about 60 degrees. If the temperature continued to drop to as low as 10 degrees, neon and hydrogen would condense. Finally even helium would condense if the temperature ever fell below one degree. There is, however, no evidence that the temperature within the solar nebula was ever that low.

How does this sequence of events differ from the sequence predicted by the inhomogeneous-accretion hypothesis? One important difference is that those minerals that can be formed only by chemical reactions between gases and previously formed minerals cannot be made. This constraint rules out processes such as the corrosion of metallic iron by water vapor and by hydrogen sulfide gas, and the reaction of water ice with ammonia and methane to form the hydrates $NH_3 \cdot H_2O$ and $CH_4 \cdot 7H_2O$. Thus the sequence of chemical reactions in the process of inhomogeneous accretion is far simpler than the one in the process of equilibrium condensation.

At first the condensation of the refractory oxides, of the iron-nickel alloy and of enstatite would proceed in the same way as they would in the equilibrium-condensation process. Thereafter sodium oxide (Na_2O) would condense. Then water, ammonium hydrosulfide (NH_4SH) and ammonia would freeze into their respective ices. At the end methane and argon would freeze. Again the temperature would probably never drop low enough to allow the condensation of neon, hydrogen and helium. It is clear, however, that the dependence of composition and density on temperature for the material formed by inhomogeneous accretion would be quite different from that for material formed by equilibrium condensation.

There are ways to check the predictions of both models against reality. For example, we have quite a lot of information about the composition and the structure of one planet: the earth. Which of the two hypotheses is better able to account for what is known about that planet? Since the two models predict the content of volatile substances in

The MITS 7440

Desk-Top Scientific Calculator — Your Assistant for Expert Results.

The 7440 Scientific Calculator does the basic arithmetic functions plus ten transcendental functions. Algebraic problem entry with two deep nesting parentheses make it possible to enter a problem just the way you would write it. Either floating decimal mode or scientific notation may be used and the results will be displayed in the correct manner automatically. A separate memory register can be used for storage of an intermediate result or of a constant. Trigonometric functions can be performed using either degree or radian entry. There is also a radian mode indication.

The fourteen digit LED display provides ten digit precision for mantissa and two digit precision for exponent. There is full sign change capability for both mantissa and exponent, and full chain calculation with any sequence of functions desired. Also featured are automatic error indications for overflow, underflow, and forbidden operations. The thirty-six key keyboard has separate clear and clear entry keys to prevent having to clear the whole problem in the event of an erroneous entry. The calculation range is 10^{99} to 10^{-99}.

KEYBOARD FUNCTIONS:

+, −, x, ÷, sine, arcsine, cosine, arccosine, tangent, arctangent, e^x, lnX, $\log_{10}x$, x^y, $\sqrt{x}$, 1/x, π, x ↔ y, ( ), change sign, enter exponent, degrees ↔ radian.

SIZE: 8½" x 12" x 3½"

***PRICE: Kit . . . . $199.95 Assembled $299.95**

(MITS also makes a complete line of electronic calculators and test equipment.)

Programmer: Desk-Top Computer —

to be used with MITS 7440 desk-top scientific calculator.

- 256 programming steps with option of expandability to 512 steps.
- Program code for each function key on calculator.
- Two RUN modes of operation: continuous or step.
- Ability to look at any instruction in memory and/or its execution.
- Programs all calculator functions plus commands: A. "If neg", B. "Go to", C. "Return", D. "Remember", E. etc.

*Prices: Kit: Programmer . . . . . $199.95 Expansion: $79.95
Assembled: Programmer . . . . . $299.95 Expansion: $129.95

(Also models available to interface with MITS 816, 1440, and 908DM desk-top calculators.)

Warranty: Full two-year warranty on assembled units, ninety days on kits.
*Prices subject to change without notice.

MITS INC.
"Creative Electronics"

6328 Linn, N.E., Albuquerque, New Mexico 87108
505/265-7553 Telex Number 660401

☐ Enclosed is a Check for $ ____
or ☐ BankAmericard # ____
or ☐ Master Charge # ____
Credit Card Expiration Date ____ ☐ Kit
Include $5 00 for Postage and Handling ☐ Assembled
☐ Model 7440 ☐ Programmer ☐ Expansion
☐ Please Send Information on Entire MITS Line.
NAME ____
ADDRESS ____
CITY ____
STATE & ZIP ____
MITS/ 6328 Linn, N.E., Albuquerque, New Mexico 87108 505/265-7553 Telex # 660401
SA-3-74

© 1974 SCIENTIFIC AMERICAN, INC

the material that formed the planets, which hypothesis is in better accord with our knowledge of planetary atmospheres? Can we discriminate between the two hypotheses on the basis of the substantial amount of information we have about carbonaceous chondrites, the oldest and most primitive meteorites?

Evidence from Meteorites and Planets

Let us look first at the evidence from the carbonaceous chondrites. That evidence strongly favors the process of equilibrium condensation. Unlike the crust of the earth, chondrites are composed of unheated, unmelted, undifferentiated primordial material. The mineralogy and relative abundances of the elements of this preplanetary stuff are available for direct inspection. A mineral that is nearly universal in meteorites is troilite, or FeS. Troilite is an important feature of the sequence of events in equilibrium condensation, but it is absent from the sequence in inhomogeneous accretion. The next most widely distributed minerals are pyroxene and olivine, most often containing between 5 and 20 percent ferrous oxide (FeO). Ferrous oxide is an essential part of the equilibrium-condensation sequence.

Carbonaceous chondrites are rich in volatile materials, including a large quantity of water bound to other substances such as the mineral serpentine. It is an interesting fact that the spectra of many asteroids are indistinguishable from laboratory spectra of carbonaceous chondrites. That would be expected on the basis of the equilibrium-condensation model. Finally, carbonaceous chondrites are highly homogeneous. There is absolutely no tendency for the material in these meteorites to be strongly sorted into many different pure minerals; instead they tend to be remarkably well mixed.

The available data on the atmospheres of the planets are also much more easily understood on the basis of the equilibrium-condensation process than on that of the inhomogeneous-accretion one. The content of volatile elements in the planets formed by inhomogeneous accretion would be zero until the time when water had condensed out of the solar nebula; thereafter the condensed material would be more than 60 percent water! On the other hand, the equilibrium-condensation model predicts the right amount of water for the earth while leaving Venus extremely arid; it also predicts a water content for Mars that is six times higher than that for the earth.

The amount of carbon in the planets is predicted well by both models. The source of the oxygen to make carbon dioxide on the terrestrial planets would apparently have to be ferrous oxide. The content of ferrous oxide in Venus, the earth and Mars predicted by the equilibrium-condensation model is sufficient in all three cases to make the observed amounts of carbon dioxide. Ferrous oxide, however, is totally absent in the sequence predicted by the inhomogeneous-accretion model.

The basic chemical and physical properties of the earth all favor the equilibrium-condensation model. The chemical composition of the earth as a whole cannot be distinguished from that of the iron-rich carbonaceous chondrites. Unfortunately for students of such problems, however, some 99 percent of the mass of the earth consists not of crustal material but of mantle material and core material, neither of which can be analyzed directly. One must be content with inferring the composition of these materials from remote observations of such properties as density and the velocity of seismic waves.

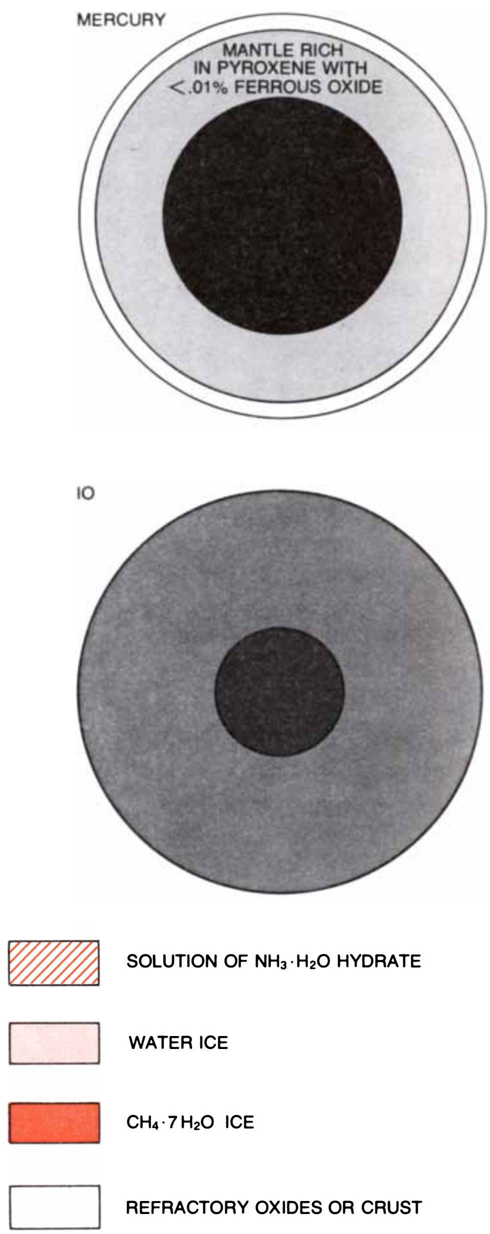
It is generally agreed that the earth's mantle largely consists of compounds containing magnesium oxide (MgO), silicon dioxide (SiO_2) and ferrous oxide. Currently it seems probable that the entire mantle is about 10 percent ferrous oxide. The mantle is divided into two distinct layers: the upper layer is mainly silicates of iron and magnesium and the lower one, where the pressure is higher, is largely a mixture of oxides of iron and magnesium.

The core is also divided into two layers. The inner core, which accounts for only 1 percent of the earth's mass, has a density that corresponds to the density of a solid iron-nickel alloy under immense pressure. The outer core is liquid and has a density appreciably less than that of liquid iron alone at the same pressure, indicating that there must be a lighter element mixed in. The lighter component has generally been thought to be either silicon or sulfur. It is a remarkable coincidence that the equilibrium-condensation model alone provides the correct abundance of sulfur to explain the observed density of the outer core. Moreover, the equilibrium-condensation model explains why the earth has a slightly higher density than Venus by predicting that the earth has sulfur in its core and Venus does not.

Whether or not Venus actually is deficient in sulfur compared with the earth is not yet known, but it is certainly true that Venus' atmosphere lacks detectable sulfur compounds. If Venus emits volcanic gases as the earth does, it should have in its atmosphere a readily detectable amount of carbonyl sulfide (COS). Carbonyl sulfide has not been found, however, even by very sensitive spectroscopic methods.

Internal Structure of the Planets

On the basis of all the evidence from meteorites and planets let us accept the equilibrium-condensation model as yielding a good approximation of the primordial diversity in the composition of the solar system. Now let us briefly



© 1974 SCIENTIFIC AMERICAN, INC

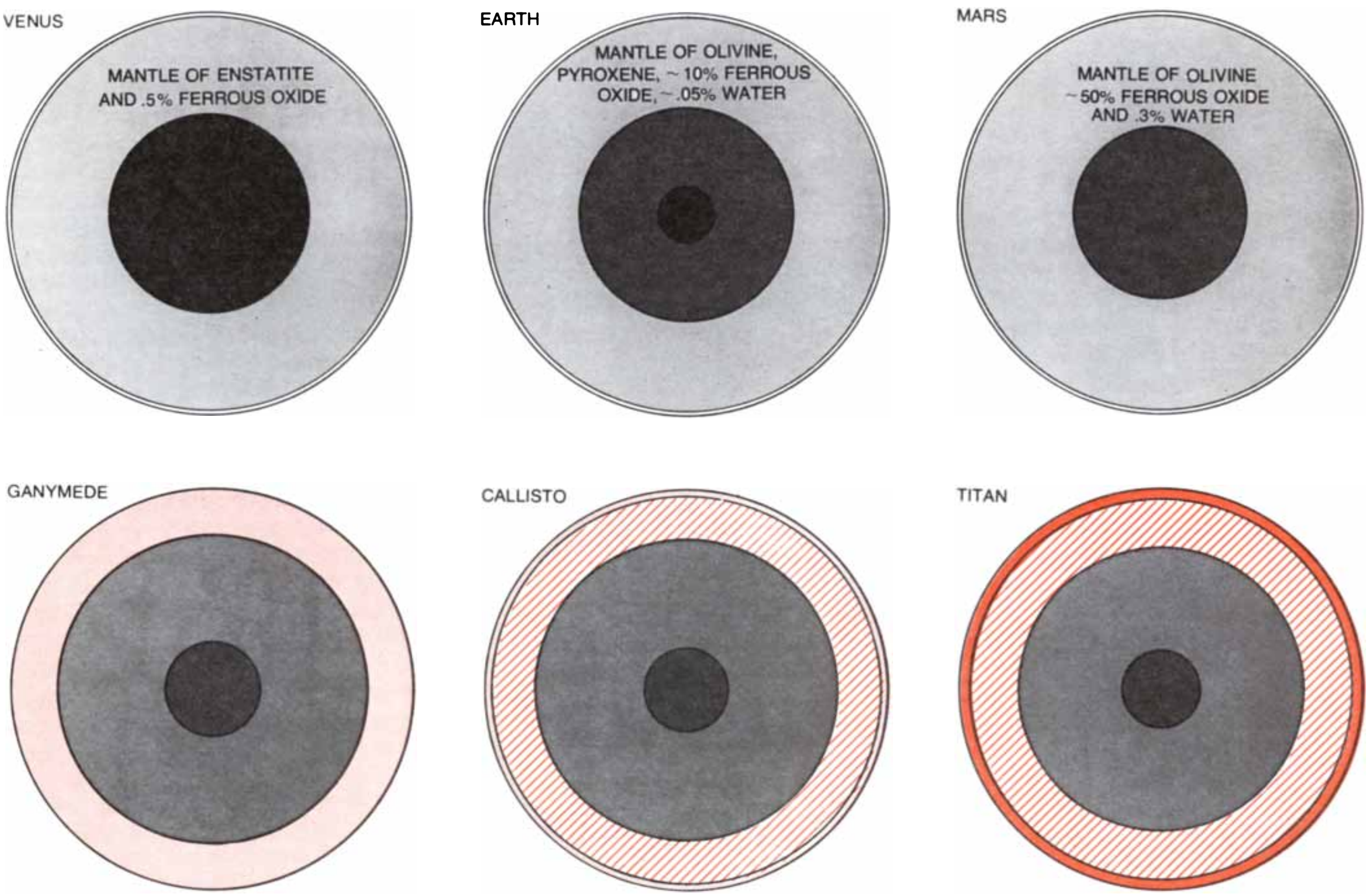
survey the nature of the physical and chemical processes going on inside the planets and satellites in order to gauge what types of internal structure could have been formed in their parent bodies.

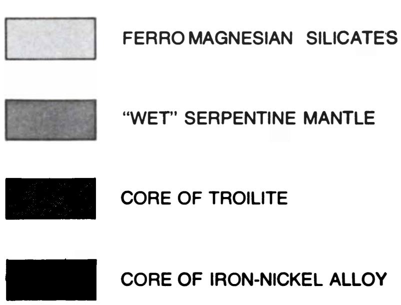
The easiest question to ask is how the present internal structure of differentiated objects such as the earth depends on the temperature at which they formed. (Small bodies such as meteoroids and asteroids that were never warm enough to differentiate would of course remain as homogeneous mixtures of their constituent minerals.) In order to answer this question we must retrace the steps of the equilibrium-condensation process as the temperature within the solar nebula dropped from 2,000 degrees K.

The first minerals formed, the refractory oxide minerals, would have given rise to a class of protoplanets with a high concentration of calcium, aluminum, titanium and the rare-earth elements. The protoplanets would also have been rich in uranium and thorium, and hence they would have been heated rapidly by the radioactive decay of those elements in their interior. Thus in spite of the fact that the refractory oxide minerals have high melting points, the temperature inside the protoplanets would have been high enough for them to melt readily. The present-day interior of a large body of this type would be mostly homogeneous. Sulfides, iron and iron oxides would be totally absent. It has recently been suggested by Don L. Anderson of the California Institute of Technology that the moon is a member of this compositional class. The moon would have had to have been extensively contaminated with sulfides and iron oxides, however, in order to account for its observed composition.

The second compositional class would have consisted of protoplanets formed just after the metallic iron-nickel alloy condensed out of the solar nebula. After these bodies had differentiated they would have had a thin crust of oxide minerals resting directly on a massive metallic core. They are not represented in the solar system today.

The protoplanets of the third class would have been formed after the temperature in the nebula had dropped far enough for enstatite to condense. The chemical makeup of a planet that came into being at this point would have been





CROSS SECTION OF OBJECTS IN THE SOLAR SYSTEM as predicted by the equilibrium-condensation model agrees with what is known about the earth and other planets. Io, Ganymede and Callisto are three of the four largest moons of Jupiter. Titan is the largest moon of Saturn. "Wet" mantle of serpentine contains much water bound into the rock. The boundary between the inner mantle and the outer mantle of the earth is not shown because it is a change in the mineral structure and not a difference in chemical composition.

© 1974 SCIENTIFIC AMERICAN, INC

dominated by silicates. A crust of oxide minerals would have formed a thin layer on top of a mantle rich in pyroxene containing a negligible amount (less than .01 percent) of ferrous oxide. The planet Mercury is a member of this class.

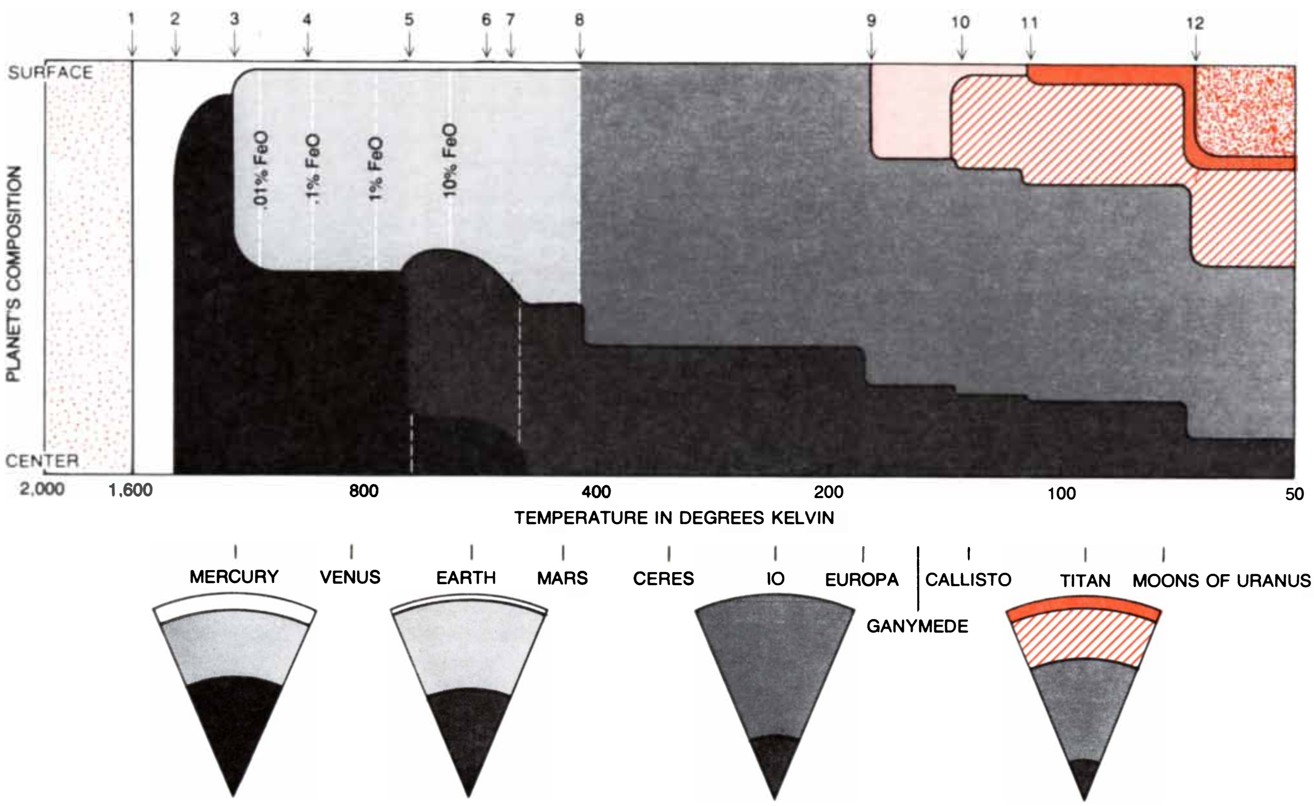
The protoplanets of the fourth class would have differed from those in the third in that their crust would have retained alkali metals, resulting in a composition chemically similar to the crust of the earth except that the content of volatile elements would have been much lower. Such a body would have had a mantle containing perhaps .5 percent ferrous oxide. The bodies in this class would have corresponded to Venus.

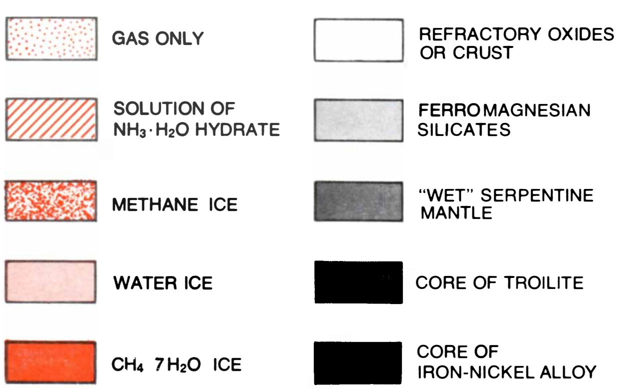
The protoplanets of the fifth class would have retained sulfur as a constituent of troilite. On melting, the troilite would have sunk to form a dense outer core. Most of the mass of the core would have consisted of this sulfide melt rather than solid metal. The mantle, which would have been composed of olivine and pyroxene, would have contained at least 2 percent ferrous oxide. This class is not represented in the present solar system except perhaps in meteorites.

The sixth major class would have been made up of protoplanets that had formed when the temperature in the solar nebula had dropped enough (to about 600 degrees K.) for water to be retained, bound in the crystal structure of minerals such as tremolite. Up to .3 percent of the mass of these bodies could have been water. The earth, with a core rich in sulfur, its mantle containing some 10 percent ferrous oxide and its water content about .05 percent, belongs to this class.

Mars and Beyond

The seventh major class would have consisted of protoplanets that had formed after metallic iron had been completely oxidized. Such a planet would have contained some .3 percent water. Its olivine mantle would have been very rich in ferrous oxide, perhaps consisting of as much as 50 percent fer-





EQUILIBRIUM-CONDENSATION MODEL of the formation of the solar system is compared with the inhomogeneous-accretion model (*see illustration on the opposite page*). The left edge of the diagram, at 2,000 degrees Kelvin, represents the interior of the solar nebula; the right edge, at 50 degrees K., represents a distance of about 30 times the distance of the earth from the center of the nebula. Numbers across the top indicate the temperatures and the locations of the 12 major reactions of the equilibrium-condensation hypothesis. At bottom are shown the temperatures at which some of the objects in the solar system would have been formed. Vertical axis represents model's prediction of objects' internal composition from its center (*bottom*) to its surface (*top*) regardless of the specific distance in miles. Pie-shaped sections at bottom show the cross section of four objects in the solar system (Mercury, the earth, Io and Titan) that formed at different temperatures. FeO is ferrous oxide. Small area of iron-nickel alloy core represents solid inner core of the earth; small area of troilite is melted outer core.

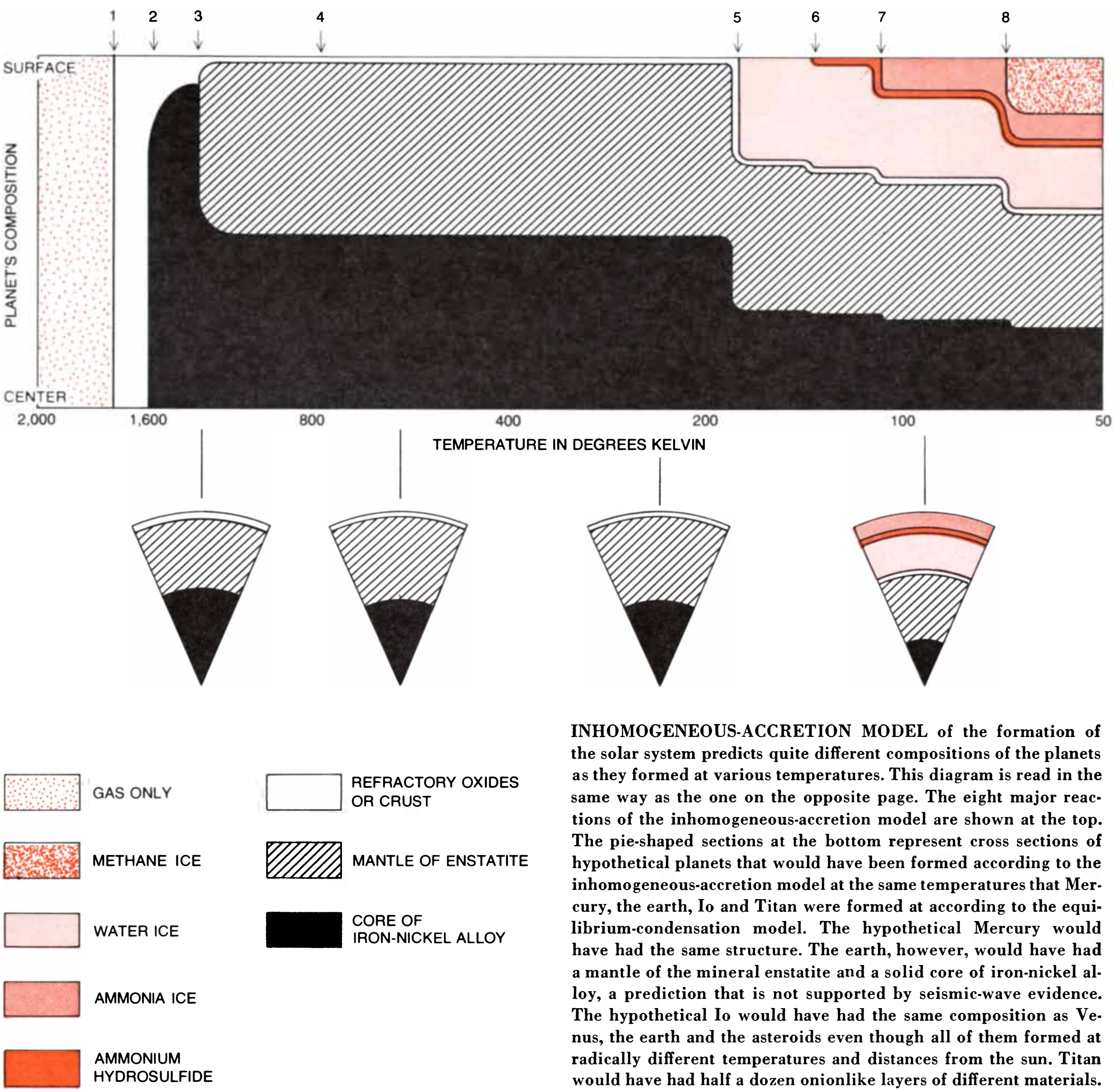
© 1974 SCIENTIFIC AMERICAN, INC

rous oxide. The core would have been either solid or liquid troilite, devoid of metallic iron. The resulting object would closely resemble Mars.

The protoplanets of the eighth class, which would have formed at a temperature low enough for the olivine in the mantle to have combined with the hydroxyl radical in the solar nebula to yield a "wet" mantle of serpentine, would have had a tremendously high water content: about 14 percent by mass. The difference in the density between the minerals in the crust and those in the mantle would have essentially disappeared, but it is possible that a core of ferrous sulfide could have existed. A homogeneous parent body of this composition could have differentiated enough to form a core even at such a relatively low temperature as 400 degrees because the melting point of silicates with a large water content is fairly low. The asteroids probably belong to this eighth class. Calculations based on the known melting points of silicates should reveal whether or not such small objects could be warmed enough by the decay of radioactive elements in their interior for the minerals to be differentiated in their feeble gravitational field. If the asteroids are in fact like the carbonaceous chondrites, then perhaps the ordinary drier, less oxidized chondrites originated between the orbits of the earth and Mars instead of in the asteroid belt.

The protoplanets of the ninth class, formed below the condensation temperature of water, would have contained water ice. In such a body the water content, both as ice and as water bound in minerals, would have been about 60 percent by weight. Thus its density would have been quite low: only about 1.7 grams per cubic centimeter and less than a third the average density of the earth. Jupiter's largest satellite, Ganymede, may fall into this class.

The protoplanets of the 10th class would have retained ammonia as the solid hydrate $NH_3 \cdot H_2O$. A mixture of this hydrate with ice begins to melt at the low temperature of 173 degrees K., which is only a few degrees above the



INHOMOGENEOUS-ACCRETION MODEL of the formation of the solar system predicts quite different compositions of the planets as they formed at various temperatures. This diagram is read in the same way as the one on the opposite page. The eight major reactions of the inhomogeneous-accretion model are shown at the top. The pie-shaped sections at the bottom represent cross sections of hypothetical planets that would have been formed according to the inhomogeneous-accretion model at the same temperatures that Mercury, the earth, Io and Titan were formed at according to the equilibrium-condensation model. The hypothetical Mercury would have had the same structure. The earth, however, would have had a mantle of the mineral enstatite and a solid core of iron-nickel alloy, a prediction that is not supported by seismic-wave evidence. The hypothetical Io would have had the same composition as Venus, the earth and the asteroids even though all of them formed at radically different temperatures and distances from the sun. Titan would have had half a dozen onionlike layers of different materials.

© 1974 SCIENTIFIC AMERICAN, INC

A Sierra Club book with no pictures?

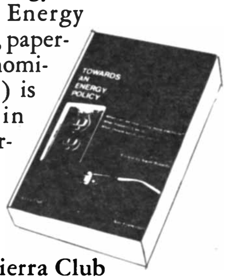
TOWARD AN ENERGY POLICY

Yes. Just a hard, sophisticated, technical assessment of the realities of America's electric power situation. This is a soft-bound collection of 17 original papers written by scientists, lawyers and academicians of the stature of DAVID ROSE, MARC ROBERTS, GEORGE AVERY, JAMES McKENZIE and HARRY PERRY. They are responses to specific questions on key issues of public energy policy posed by the Sierra Club, with the help of former FPC Commissioner Charles Ross, and former Vermont Governor Phillip Hoff.

The papers present what is known about the supply and technology of electric energy, they examine possibilities of reducing consumption; suggest policies to encourage that end; and consider the economic and social consequences of energy conservation.

We do not recommend this volume for the newly interested layman. He will find it, to paraphrase that succinct book review, "too far between the covers." We do believe, though, that this scholarly approach to questions of public energy policy will be of real value to the professional who must come to grips with the realities of our present energy situation.

"Towards an Energy Policy" (634pp., paperbound and economically reproduced) is not available in bookstores. To order send your name and address to:



Energy Book, Sierra Club
Box 7959 Rincon Annex
San Francisco, Calif. 94120

Enclose your check, made out to Sierra Club Special Publications. $7.50 per copy. ($6.00 to Sierra Club members) Add tax: California, 5%; BART Counties, 5.5%; New York City, 7%; New York State, 4%; New Jersey, 5%.

Name________________

Address________________

____________ Zip______

daytime surface temperature of Jupiter's second-largest satellite, Callisto. If a protoplanet in this class had ever been heated enough by radioactive decay to differentiate, it would have had a deep mantle of a liquid water-ammonia solution surmounted by a thin crust of water ice. During the differentiation a core of sulfides, oxides and hydrous (water-containing) silicates would have been formed as a sediment. The overall result would have been a stratified ball of mud and slush.

The protoplanets of the 11th class would have been about 4 percent methane by weight. The heat generated by radioactive decay would have melted the solid hydrate $CH_4 \cdot 7H_2O$, liberating methane gas to form an atmosphere. Differentiation would have given rise to a solid crust of the hydrate. A prime candidate for membership in this class is Saturn's largest satellite, Titan, which has a massive atmosphere rich in methane.

The protoplanets of the 12th and last class would have held solid methane as the temperature in the solar nebula dropped to almost 50 degrees K. Since carbon is nearly as abundant as oxygen in the solar system, and since the density of solid or liquid methane is very low (about .6 gram per cubic centimeter), the density of such a protoplanet would have been only one gram per c.c. Several of the smaller satellites of Saturn have been reported to have densities of near one gram per c.c., but the margin of error in such measurements is at least a factor of two. It is entirely possible that some of the satellites of Uranus or Neptune belong to this class, but their masses and radii will have to be determined by spacecraft flyby missions before we can confirm that speculation with any certainty.

I have mentioned the satellites of Jupiter, Saturn, Uranus and Neptune but not the planets themselves. There are reasons for the absence of the Jovian planets from the compositional classes. As we have seen, these classes are based on the temperature at which the protoplanetary material condensed out of the solar nebula according to the equilibrium-condensation model. Now, since the composition of Jupiter and Saturn is essentially the same as the composition of the sun, they could have formed just about anywhere in the solar nebula with no strong dependence on the temperature in that region. All that would be required is an initial protoplanetary core whose gravitational field was strong enough to accrete the undifferentiated solar material. Uranus, Neptune and Pluto are left off the list simply because not enough is known about their composition.

Important Results

One important result of the sequence of reactions I have been discussing is that the densities of the terrestrial planets, like their chemistry, can all be explained as a direct consequence of the variation of the temperature at which they were formed according to their distance from the sun. It has often been argued that, by analogy with a hypothesis about the formation of meteorites, the different densities of the terrestrial planets are due to a physical process that fractionated the material in such a way that metals were preferentially accreted in one case (Mercury) and silicates were preferentially accreted in the others. Such preferential accretion, as it is seen in meteorites, might have been the result of differences in the magnetic or electrical properties of the material, or in how well the accreted matter adhered to the surface of the protoplanet. Since the sequence of reactions of the equilibrium-condensation model leads automatically to the high density of Mercury, we are led to deny that any fractionation process affected the planets. Planets are so massive that they must have accreted their material by gravitation, a totally nonselective process. Temperature alone determined the composition of the accreted material. On the other hand, the parent bodies of meteorites, which could have had a mass as small as 10^{-12} times the mass of a planet, would have had only a feeble gravitational attraction. They probably accreted their material almost entirely through selective interparticle forces such as magnetism. Thus the paradox vanishes.

In summary, the equilibrium-condensation model shows very simply how the density and the volatile content of the solid material in the solar system are related to the temperature at which the material was formed. The model leads to a large variety of quantitative predictions that can be tested in the laboratory, by astronomical observations and by planetary probes. Within a single chemical sequence we can now interrelate such seemingly diverse matters as the bulk composition of the solar nebula, the internal structure of the planets and the content of volatile compounds in the planets.

Perhaps the most striking result of this approach, however, is the underlying vision of an intimate interrelationship among all the objects in the solar

© 1974 SCIENTIFIC AMERICAN, INC

I am you.

My name is Noritsune Chiba, and these are my kids. They're giving me a workout... it seems like they'll never run out of energy.

I'm trying to make sure they don't. I'm an Hitachi engineer. And energy is my business.

For the last four years I've coordinated Hitachi's contributions to a hydroelectric plant at Ludington, Michigan. A really enormous project.

At night, when off-peak power is available from other power generating plants, we use it to pump nearly 2,000 tons of water per second up hill from Lake Michigan into an artificial lake. Then during the day, when power *is* in demand, we let it flow down again—through turbines, in order to generate electricity.

Building the colossal pump/turbines and generator/motors for this project has required all kinds of new technology. For instance we had to develop a new thick steel welding technique for the turbine runner. And we had to design a thrust bearing capable of supporting over 2,000 tons.

The best thing about the project is the result. The people of the American Midwest will now have an additional 1,872,000 kilowatts during the hours of peak demand. What's more, they'll have their additional power without additional pollution.

Ludington, Michigan is half-way round the world from where I live. But even so, I feel like I've done something for my family. I feel that the technology I've helped develop will help everybody.

I feel that when I work for Hitachi, I'm working for all of you.

Because my work will benefit me.

And I am you.

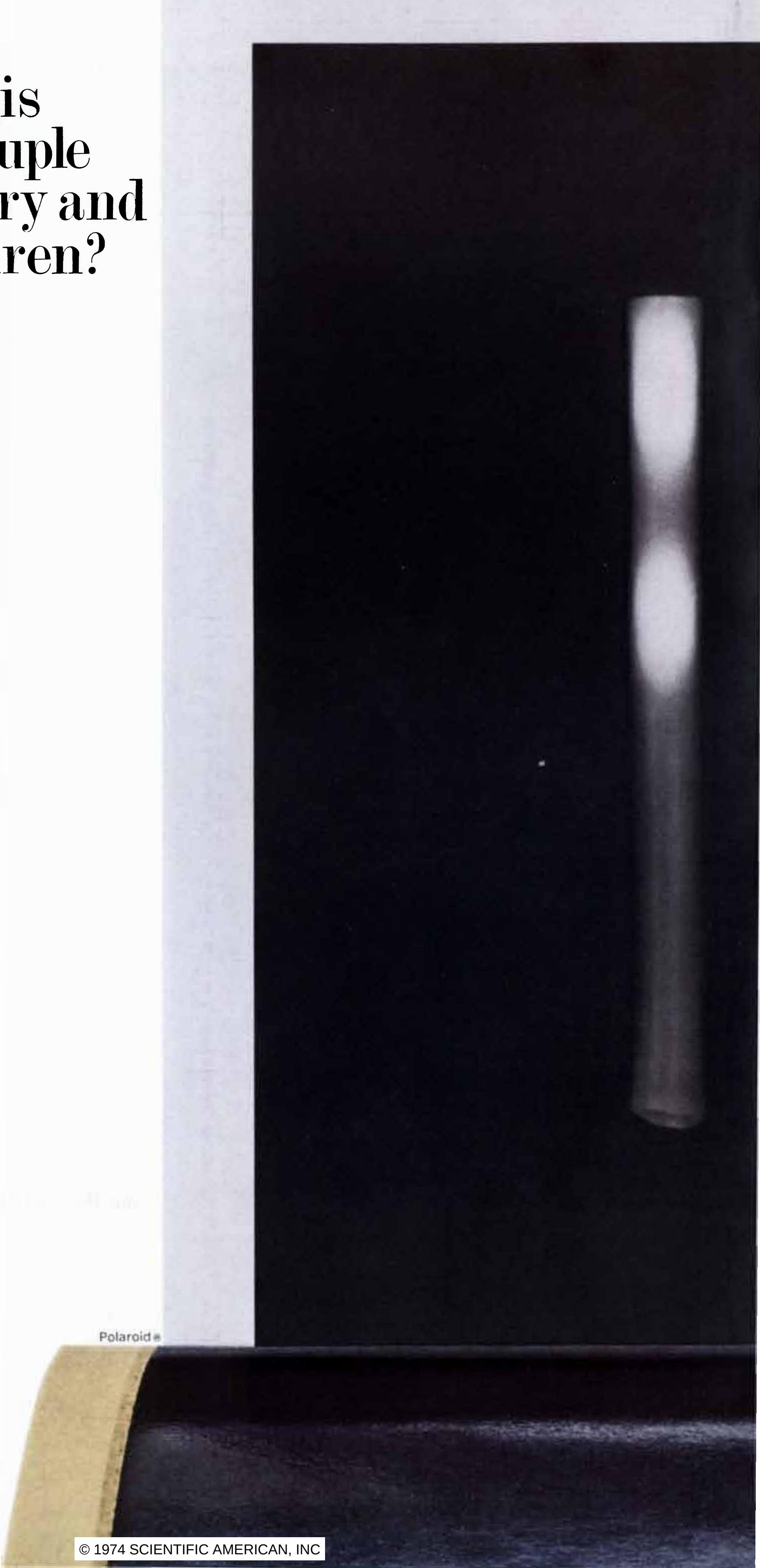
HITACHI
Resourceful people

Hitachi America, Ltd.
Head Office: 437 Madison Avenue, New York, N.Y. 10022
Tel: (212) 758-5420
Branch Offices: San Francisco, Chicago, Atlanta.

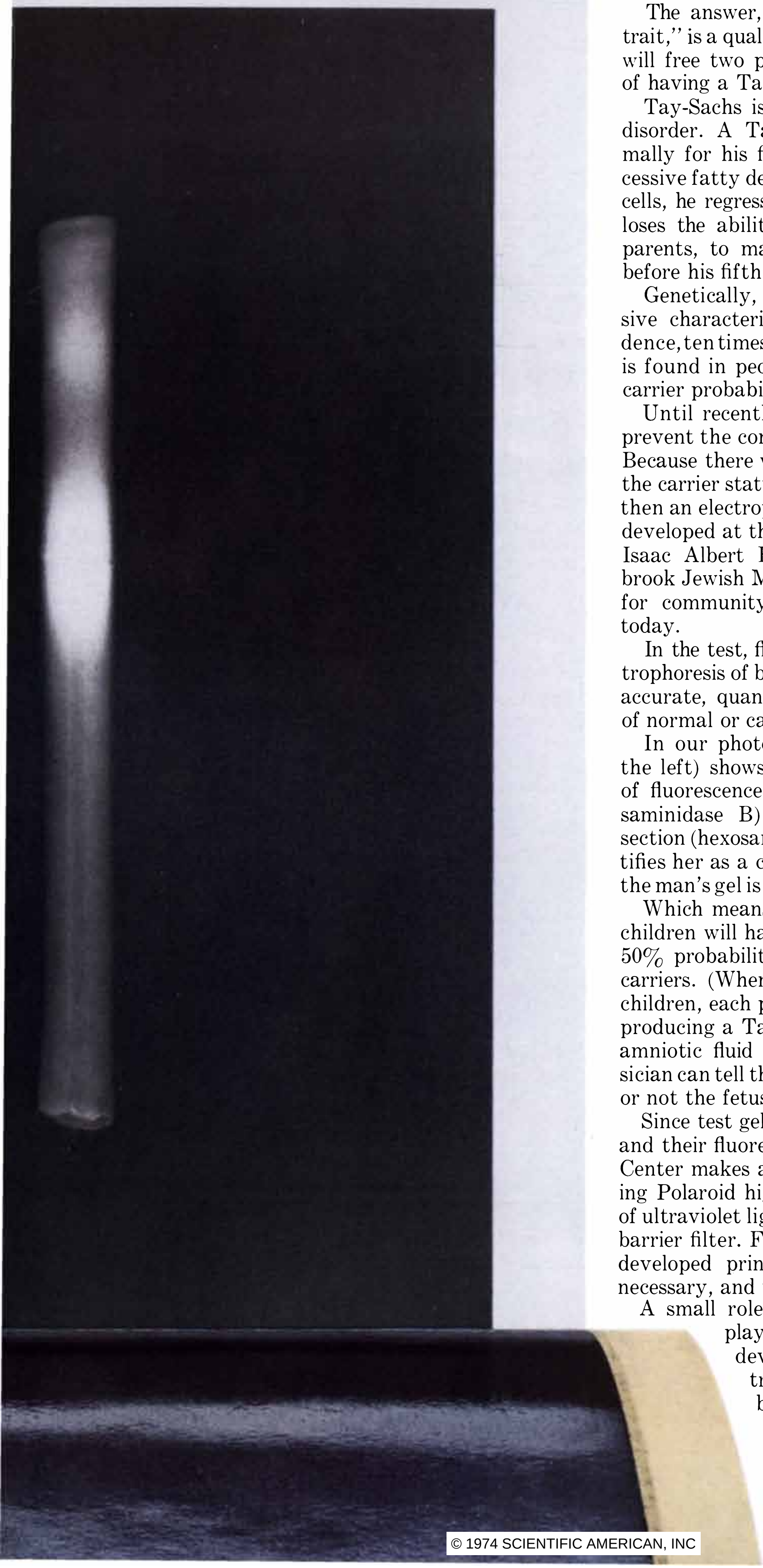
● Power Generation & Transmission Equipment ● Industrial Machinery ● Transportation Equipment ● Communications Equipment ● Electronics & Home Appliances

© 1974 SCIENTIFIC AMERICAN, INC

Does this Jewish couple dare to marry and have children?



© 1974 SCIENTIFIC AMERICAN, INC



The answer, shown in our Polaroid "portrait," is a qualified "yes." It's an answer that will free two people from the haunting fear of having a Tay-Sachs child.

Tay-Sachs is a fatal, inherited, metabolic disorder. A Tay-Sachs child develops normally for his first six months. Then, as excessive fatty deposits accumulate in his brain cells, he regresses. Slowly and inexorably, he loses the ability to sit up, to recognize his parents, to manage a smile. Then, usually before his fifth birthday, he dies.

Genetically, Tay-Sachs is a simple recessive characteristic. Its highest carrier incidence, ten times that of the general population, is found in people of Jewish ancestry where carrier probability is 1 in 30.

Until recently, there was no sure way to prevent the conception of a Tay-Sachs baby. Because there was no sure way to determine the carrier status of prospective parents. And then an electrophoretic testing technique was developed at the Birth Defects Center of the Isaac Albert Research Institute at Kingsbrook Jewish Medical Center. It's being used for community screening in many centers today.

In the test, fluorescent acrylamide gel electrophoresis of blood samples provides a rapid, accurate, quantitative, visual determination of normal or carrier status.

In our photograph, the woman's gel (on the left) shows a relatively higher intensity of fluorescence in the upper section (hexosaminidase B) as compared to the lower section (hexosaminidase A). This pattern identifies her as a carrier. The reverse pattern in the man's gel is typical of a normal individual.

Which means there's *no* chance that their children will have the disease. But there *is* a 50% probability that their children will be carriers. (When two carriers decide to have children, each pregnancy holds a 25% risk of producing a Tay-Sachs child. By testing the amniotic fluid early in pregnancy, the physician can tell the prospective parents whether or not the fetus has the disease.)

Since test gels deteriorate within 12 hours, and their fluorescence dissipates rapidly, the Center makes a permanent record by exposing Polaroid high contrast film to 5 seconds of ultraviolet light through a Wratten No. 2A barrier filter. Fifteen seconds later, the fully developed print can be studied, reshot if necessary, and filed for future reference.

A small role for one of our products to play? Maybe. But when anybody develops a procedure that helps troubled people, we're proud to be involved.

Just for the record.

Polaroid Corporation

© 1974 SCIENTIFIC AMERICAN, INC

Toyota introduces ESP.

A warning system that helps keep minor problems from becoming major expenses.

ESP is a built-in feature located above the driver.

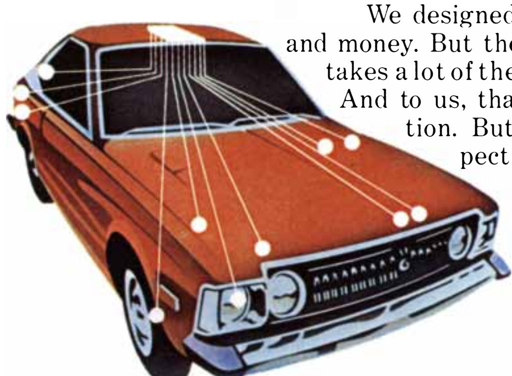
Take a close look at the Electro Sensor Panel pictured here. We call it ESP. And you've never seen anything like it before. That's because no other car has it. It's that new.

Toyota created ESP to monitor 11 key service areas. If just one of them needs attention, ESP pops a light on to let you know. Before you find out the hard way.

ESP works constantly when the car is running. And it gives you advance warning. That means you can correct your problem before it becomes critical.

ESP is standard equipment on the Corona Hardtops. It's a built-in feature located above the driver. To check the system, turn the key, and ESP lights up momentarily to show you that it's working.

We designed ESP to save you both time and money. But there's more. We feel that ESP takes a lot of the worry out of caring for a car. And to us, that's important. A real innovation. But then, that's what you'd expect from Toyota.



ESP monitors 11 key service areas while you're driving.

TOYOTA

Small car specialists for 40 years.

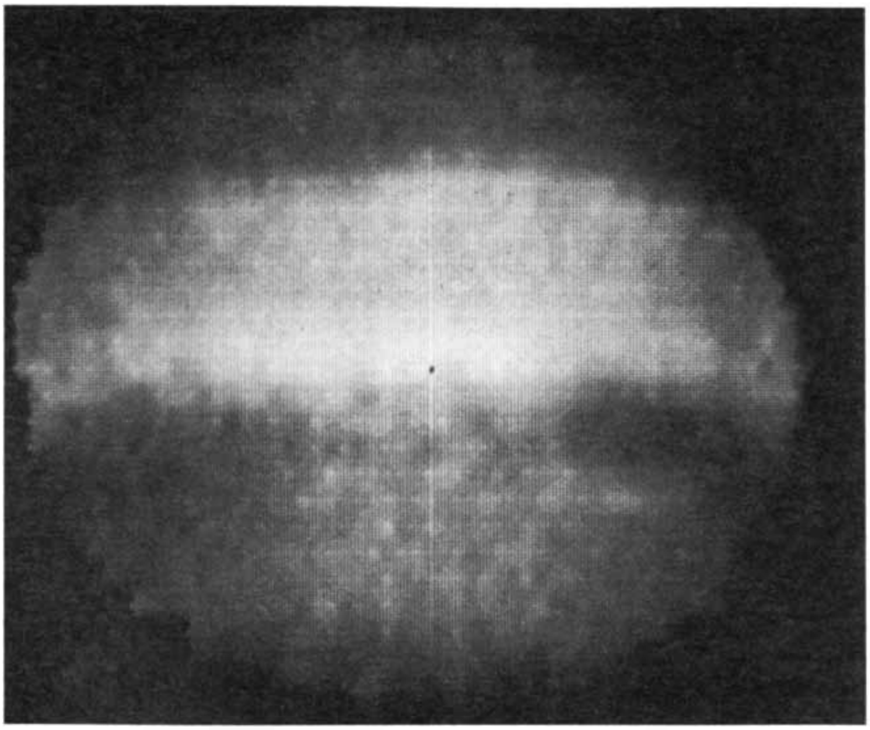
ESP may be in limited supply in your area.

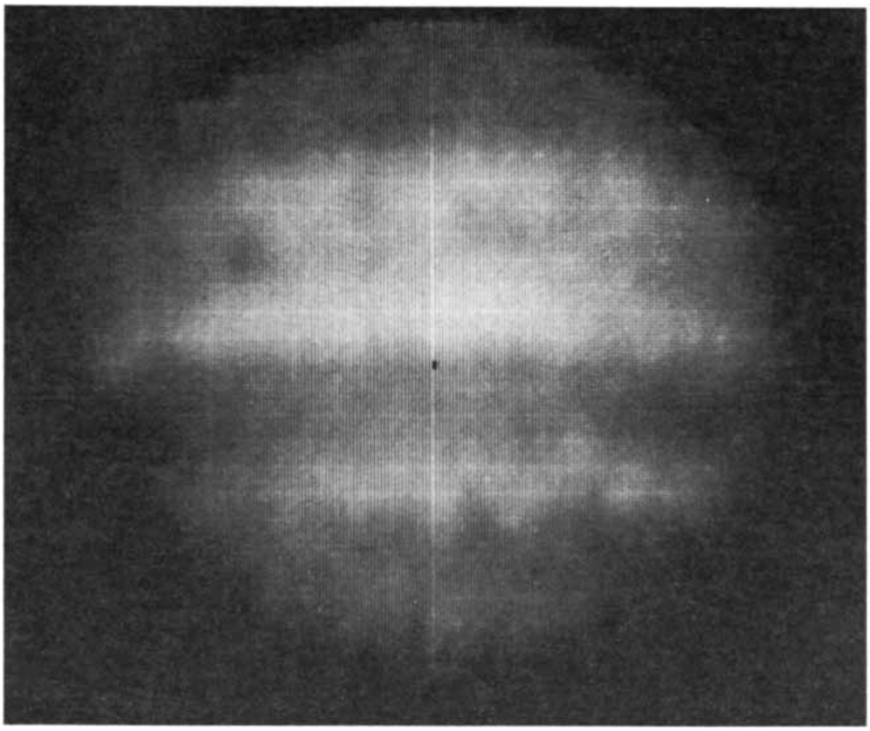
© 1974 SCIENTIFIC AMERICAN, INC

system. We see that there is strong evidence that the planets formed from a medium of uniform elemental composition. We see that the same qualitative processes of condensation and accretion are at work in the same general way throughout the entire system. The presence or absence of sulfur gases in the atmosphere of Venus, the density of the core of Mars, the electrical conductivity of the mantle of Callisto, the atmospheric pressure at the surface of Titan, the radius and mass of the satellites of Uranus—all tangibly influence our comprehension of the origin, composition and structure of our own planet. In this era in the history of geology we have seen how important it is to study the origins of all the chemical elements in order to understand the origin of any one of them, and we have been obliged to study global tectonics in order to understand the origin of any one mountain. Now we must be prepared to learn from comparative studies of the planets exactly what processes affect planets in general and how planets evolve. From this kind of investigation we can learn for the first time how the earth originated, evolved and attained its present structure.

Another long-range benefit of such comparative studies will be to provide insight into how frequently planets might be formed around other stars, and what the composition of those planets might be. Since the composition of most stars is indistinguishable from that of the sun, and since present theories of the origin of the solar system suggest that the formation of planets is a normal by-product of the formation of stars, it would not be surprising to find that planets similar to those in the solar system are common throughout our galaxy and other spiral galaxies. It is interesting to speculate whether or not planets belonging to some compositional classes that are not seen in the solar system actually exist in other planetary systems. One such example would be a planet made up of iron and refractory oxides that was far denser than Mercury; another would be a very earthlike planet that was devoid of water; a third would be a true planet composed of material rich in volatile compounds such as the material ascribed to the asteroids.

The next few years will provide numerous opportunities for directly testing these predictions. A vigorous program for exploring the planets is the only way to realize the full benefit of the opportunities. It is the logical next step in our efforts to understand the planet on which we live.





HEAT MAPS OF JUPITER at two different wavelengths were made from *Pioneer 10* with the infrared radiometer experiment of the California Institute of Technology, directed by Guido Munch. The radiometer is a reflecting telescope with an aperture of three inches mounted rigidly on the spacecraft. The spin of *Pioneer 10* and its motion in its trajectory past Jupiter allowed the radiometer to scan the planet and build up its images. The top picture was made at a wavelength of 20 microns and measured the heat being emitted at a distance of 40 kilometers below the top of Jupiter's atmosphere; the bottom picture was made at a wavelength of 40 microns and measured the heat being emitted at a distance of 50 kilometers below the top of the atmosphere. Both were made on December 3, 1973, during the spacecraft's closest approach to the planet. The warmest (lightest) areas are eight degrees K. warmer than the coolest (darkest) areas. The warmest areas correspond to the dark belts of Jupiter; coolest areas correspond to the bright belts. Planet is warmest near equator. It is estimated that the planet has an average temperature close to 128 degrees K. It radiates more than twice the amount of thermal energy it receives from the sun, confirming observations from the earth indicating that planet generates some heat of its own.

© 1974 SCIENTIFIC AMERICAN, INC

INORGANIC POLYMERS

Most commercial polymers are long-chain molecules with carbon atoms in their backbone. New families of polymers with backbones of atoms other than carbon are yielding materials with unusual characteristics

by Harry R. Allcock

The products of the polymer chemist are all around us: clothing made from synthetic fibers, cups of polystyrene, bottles of polyethylene, bearings of nylon, Teflon-coated cooking ware, plastic bags, polymer-based paints, epoxy glue, synthetic rubber, polyurethane foams—the list is almost endless. The past 20 years have seen a technological revolution in which traditional materials, such as cotton, wood, glass and metals, have been steadily replaced by synthetic polymers. It is an article of faith among polymer chemists that the trend will continue.

Nearly all the synthetic plastics and elastomers in use today are organic polymers. They consist of very large molecules, usually linear or branched chains with carbon atoms in their backbone. For example, polyethylene molecules consist of long chains of carbon atoms with two hydrogen atoms attached to each carbon. Polystyrene has a similar structure, except that a phenyl ring (C_6H_5) replaces one hydrogen atom on every other carbon. Linear chains are also found in polypropylene, Nylon 66, poly(methyl methacrylate) (Lucite or Plexiglas) and poly(ethyleneterephthalate) (Dacron and Mylar). Synthetic polymers such as these can be used to make strong fibers, intricate solid shapes, tough films and rubberlike elastomers.

The fact remains that many organic polymers have serious deficiencies. Few can be heated for prolonged periods above 150 degrees Celsius without either melting or decomposing. When decomposition occurs, it usually results from the reaction of the carbon atoms with atmospheric oxygen. Many organic polymers dissolve or swell in hot organic liquids or oils and for this reason cannot be used as seals or components in automobile or aircraft engines.

Furthermore, few organic polymers remain flexible or rubbery over a wide enough temperature range for them to serve at both low and high temperatures. That is one reason why automobile tires that contain nylon cords tend to "bump" on cold mornings and why plastic garden hose cannot be unwound in winter. In the Arctic or in high-flying aircraft this hardness and brittleness of polymers can create serious hazards. There are other applications for which no suitable organic polymers have yet been found: in spacecraft, where prolonged resistance to ultraviolet radiation is essential; in the textile industry, where flame-retarding fabrics are needed, and in medicine for the fabrication of artificial organs that will not lead to blood clotting or other undesirable effects. Many polymer chemists now believe the solution to problems such as these will require the development of radically different polymer systems. That is why attention is increasingly being focused on inorganic polymers.

Most long-chain synthetic polymers show a characteristic sequence of changes as they are heated. All linear polymers are glasses at low temperatures. As they are heated a certain temperature is reached at which they change from a glass to a rubber. This change is known as the glass-transition temperature. At still higher temperatures the polymer may soften slowly to a gum and eventually liquefy. If the material is a crystalline polymer, that is, if sections of adjacent chains are packed together in a regular array, it will melt to a viscous liquid at a sharply defined temperature. Extensive cross-linking of the chains yields a matrix polymer, a rigid material with a high melting point [*see illustration on page 68*].

The observable differences among polymers are chiefly attributable to different glass-transition temperatures, varying degrees of crystallinity and specific melting characteristics. To a large degree these differences are due to the presence of different substituent groups attached to the main chain or to the presence of elements other than carbon in the chain. Thus by varying the side-group structure or by introducing oxygen or nitrogen atoms into a carbon chain an almost bewildering variety of polymers have been made available, with an enormous range of different properties.

Fundamentally, inorganic chains should show the same thermal transitions as organic polymers. For example, they should form glasses at low temperatures, become rubbery above a glass-transition temperature and finally become viscous and melt. In addition extensive cross-linking of the chains should give rise to rigid matrixes that have a high melting point. At the same time inorganic polymers should exhibit considerably more resistance to oxidative degradation than their organic counterparts.

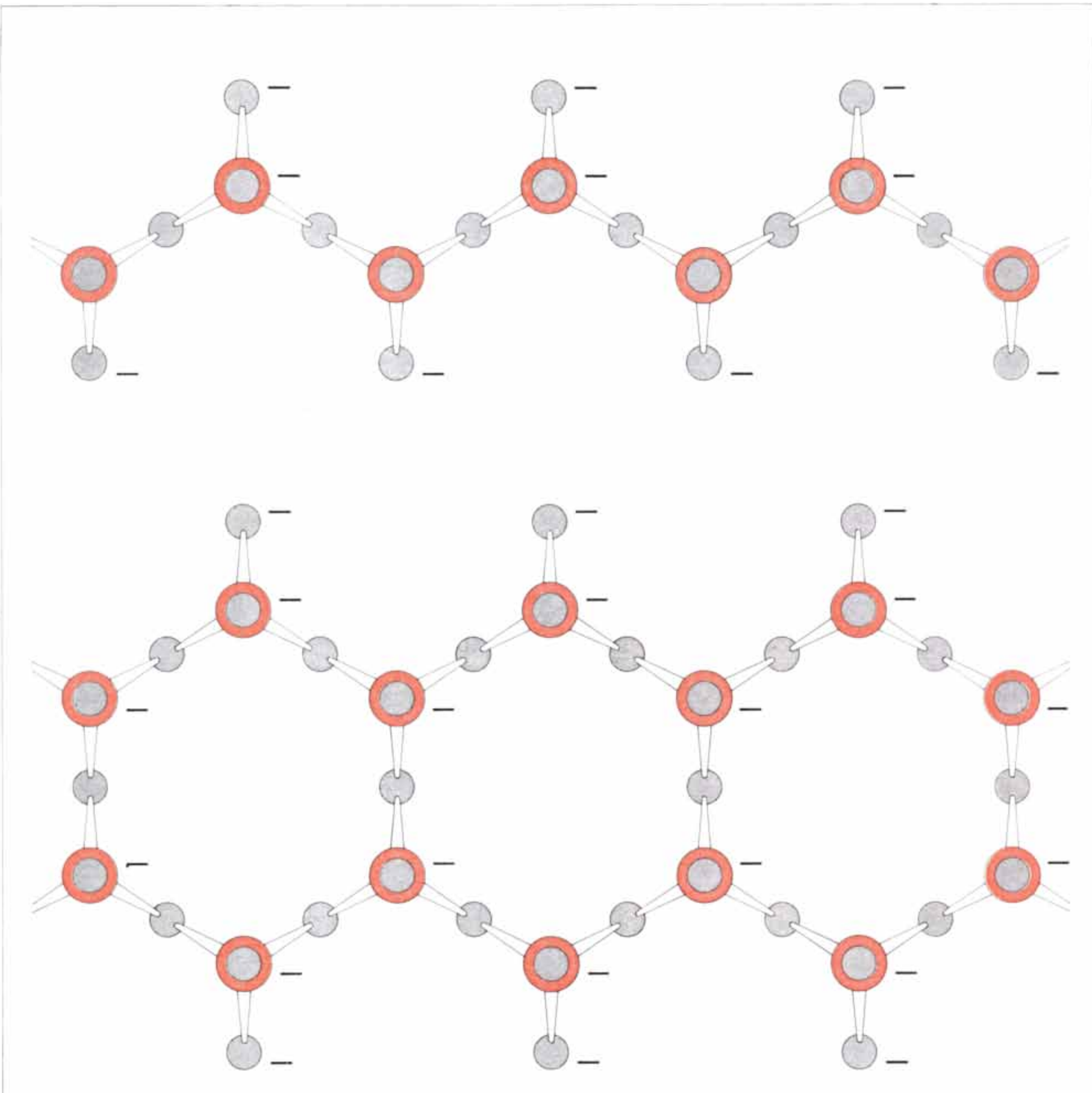
Some very important inorganic polymer systems have been in use for a long time. Glass is an inorganic polymer made up of rings and chains of repeating silicate units. It is well known that glass fibers can be used for the fabrication of textiles. Similar chains of silicate units are found in pyroxene minerals. Ladder polymers, or double chains, are found in amphibole minerals, such as one form of asbestos [*see illustration on opposite page*]. In both of these structures the negative charges of the oxygen atoms lying off the central chain are neutralized by positively charged metal ions, such as the ions of sodium,

© 1974 SCIENTIFIC AMERICAN, INC

magnesium and calcium. These charged side-group structures tend to bind adjacent chains together to form a cross-linked matrix. Thus instead of displaying rubbery, elastomeric properties or flexible, film-forming properties, the polymers have a high melting point. When covalent cross-links are present, as they are in silica glasses, a three-dimensional system of covalent silicon-oxygen bonds generates a very-high-melting-point matrix.

Phosphate polymers, with a backbone of alternating phosphorus and oxygen atoms, are also known, and these formally resemble linear silicates. For example, crystalline high polymers of alkali phosphates have been known for many years as "Maddrell's salt" and "Kurrol's salt." Again the presence of the charged side groups lowers the torsional mobility of the chains, with consequent elimination of flexibility at low temperature. An additional problem encountered with some of these polyphosphates is that treatment with water causes breakage of the chains and brings about conversion to products of low molecular weight.

Most rocks, brick, concrete and ceramics are three-dimensional inorganic polymers. Their uses are limited for many applications by the fact that they are difficult to fabricate into useful objects except at high temperatures. They are not, in general, flexible, elastomeric



INORGANIC CHAINS AND RINGS, which occur in certain classes of minerals, provide useful models for the polymer chemist. Pyroxene minerals (*top*) are single-strand polymers in which oxygen atoms and silicon atoms (*color*) alternate along the backbone of the chain. Each silicon is also bonded to two oxygen atoms that project outside the backbone. These oxygen atoms bear a negative charge that is neutralized by positively charged ions of such metals as sodium (*not shown*). The arrangement of charges binds chains together in a highly cross-linked matrix. Amphibole minerals (*bottom*) have a double-strand structure cross-linked in similar fashion.

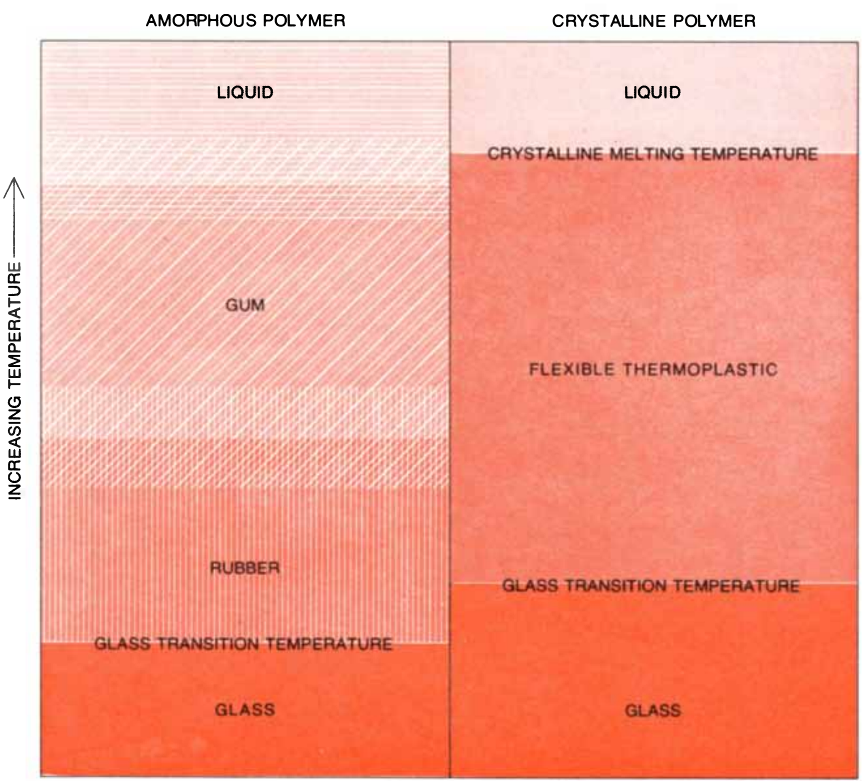
© 1974 SCIENTIFIC AMERICAN, INC

or resistant to impact. Although they have excellent stability in air at high temperatures, they do not fulfill many of the requirements of modern technology.

For a number of years chemists have realized that an answer to this problem lies in the middle ground, between the organic polymers on one hand and the mineralogical inorganic materials on the other. Basically the need is for polymers that have a linear structure, nonionic substituent groups to favor flexibility and inorganic elements in the backbone to provide stability against heat and oxidation. The presence of inorganic elements in the polymer chain should also confer flexibility and elasticity over a wide temperature range. In theory hundreds of different inorganic chain structures can be envisioned. In practice certain restrictions must be considered.

The primary restriction is that the elements that make up the backbone should be linked together through covalent bonds. This restriction is employed to reduce the chances that the chains will be susceptible to attack by water or that ionic cross-links will exist. For that reason the primary search for useful inorganic polymers has concentrated on the covalent "main group" elements, particularly boron, aluminum, silicon, germanium, tin, nitrogen, phosphorus, arsenic, antimony, oxygen, sulfur, selenium and tellurium. One of the earliest covalent inorganic polymers to be made, and still one of the most fascinating, is polymeric sulfur. The stable form of sulfur at room temperature is rhombic sulfur, which contains cyclic molecules with eight sulfur atoms in a ring [*see top illustration on opposite page*].

Rhombic sulfur is a brittle, crystalline material that melts at 113 degrees C. to form a yellow-red liquid. Liquid sulfur has the curious property that above 159 degrees C. its viscosity increases as the temperature is raised, contrary to the behavior of nearly all other liquids. The viscosity increase results from the opening of the eight-member rings and their conversion to long chains. Since long chains can become entangled more effectively than rings, the viscosity rises. As the temperature is raised above about 175 degrees C., however, the viscosity begins to decrease, an indication that depolymerization is now occurring to convert the chains back to rings.



AMORPHOUS AND CRYSTALLINE POLYMERS exhibit distinctly different properties in the middle temperature range when they are too warm to be glassy and too cool to be liquid. When heated above the glass-transition temperature, amorphous polymers first become rubbery, then gumlike and finally liquid, with no clear demarcation between phases. Crystalline polymers are flexible and thermoplastic over a fairly wide temperature range.

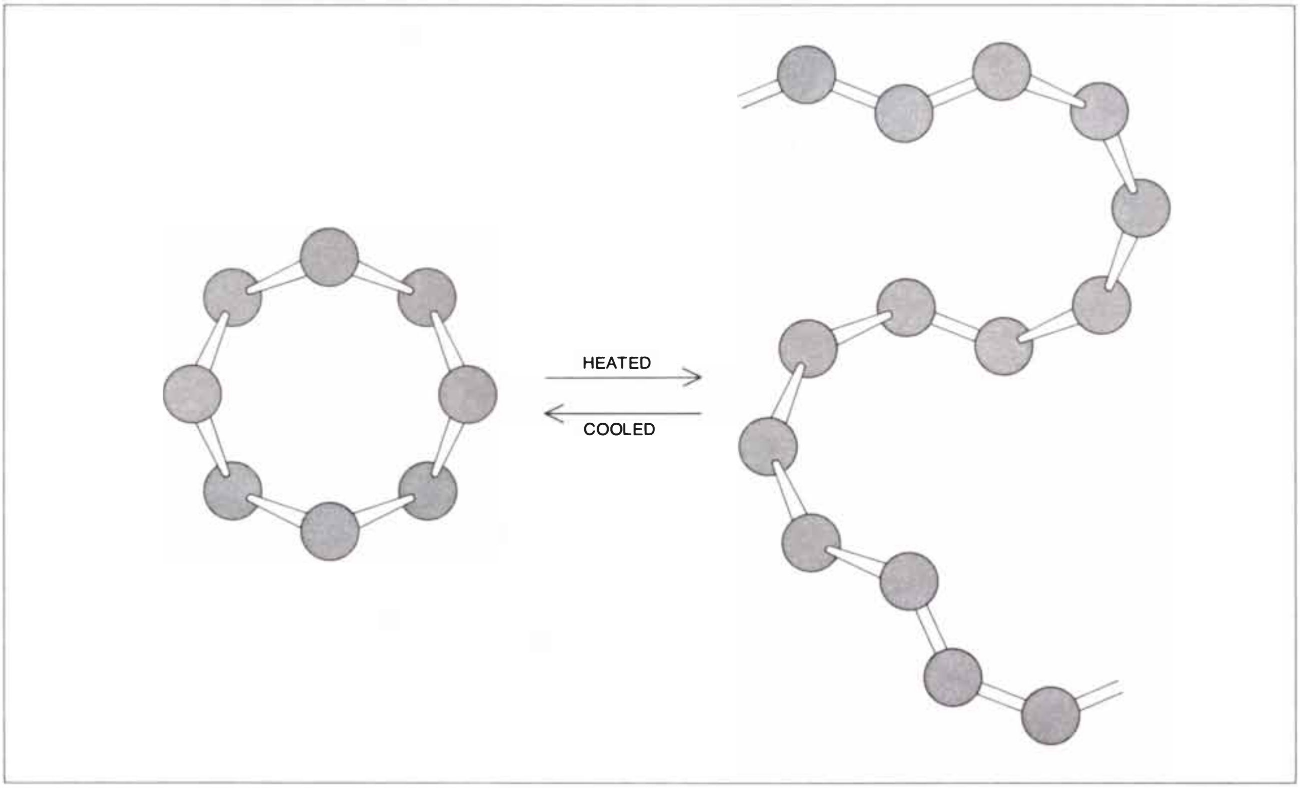
The high-polymeric form of sulfur can be "quenched," or isolated, by pouring the hot melt into water. When it is cooled, it forms rubbery lumps or fibers [*see illustration on page 74*]. The polymer is not stable at room temperature; it hardens slowly as the long chains crystallize and revert to the cyclic eight-member rings. Thus from a practical point of view polymeric sulfur has no uses. Nevertheless, its thermal behavior illustrates a characteristic feature of many inorganic polymer systems: rings can be converted to high-polymeric chains and at elevated temperatures this process tends to be reversed.

What one would like to synthesize is a polymer that resembles the inorganic glasses but that is altered so that the polymer chains exhibit flexibility at ordinary temperatures. As we have seen, the rigidity of glass and similar polymers is due largely to the presence of ionic side groups. If the charged side groups could be replaced by uncharged units, the flexibility of the chains should be increased.

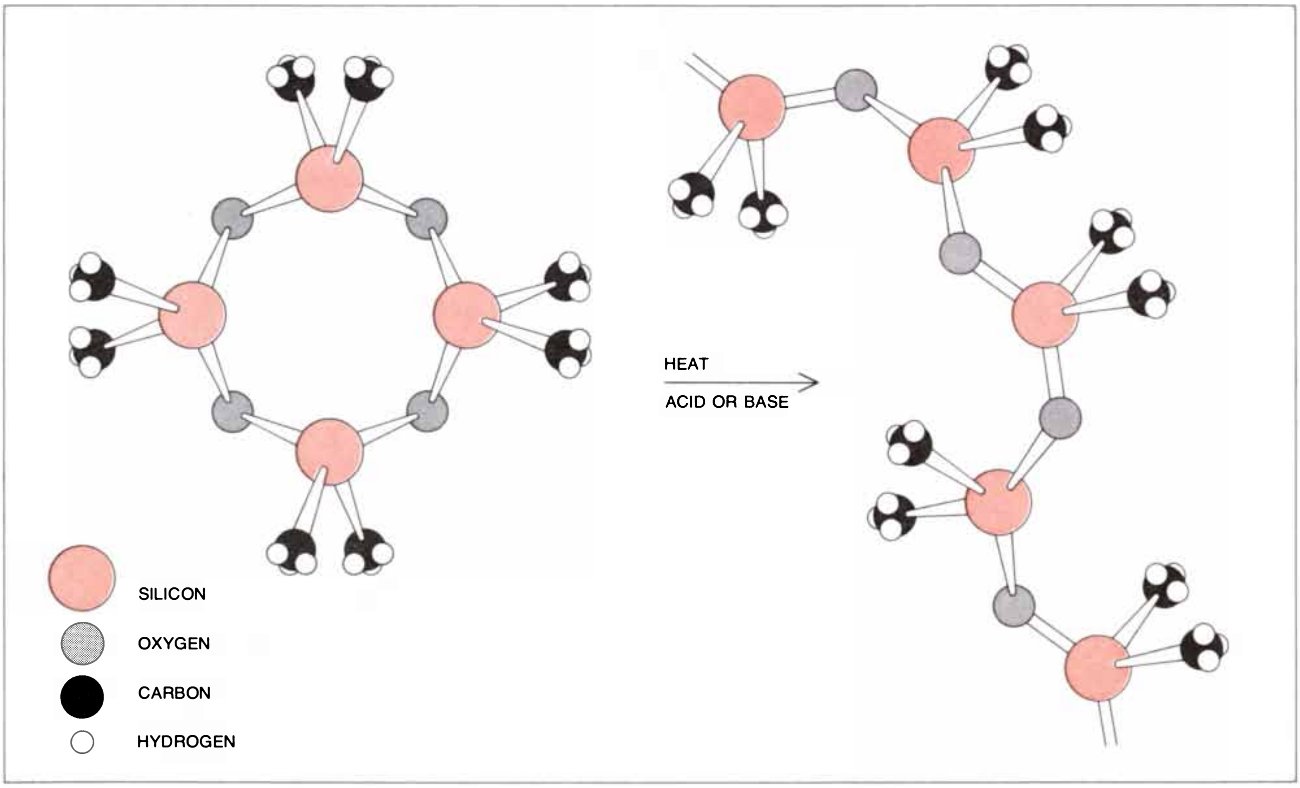
This concept is the idea behind the design of silicone polymers. The most widely used silicone, poly(dimethylsiloxane), consists of a chain of alternating silicon and oxygen atoms, with two methyl groups attached to each silicon. Like polymeric sulfur, silicones are made by polymerization of a cyclic ring compound with eight atoms in the ring. Here the starting material is a colorless, oily compound known as octamethylcyclotetrasiloxane, the molecule of which has a ring of four silicon atoms alternating with four oxygens. When the eight-member ring is heated above 100 degrees C. with a trace of acid or base, it is polymerized [*see bottom illustration on opposite page*]. The final polymer may consist of as many as 30,000 silicon-oxygen units. Silicone polymers are highly viscous liquids or gums. Silicone rubber is made from the gum by connecting the chains together in a cross-linking process. The cross-links prevent the chains from slipping past one another when the material is stretched.

Perhaps the most surprising feature of silicone high polymers is the fact that they are flexible and elastomeric over an extraordinary temperature range. Their glass-transition temperature is −130 degrees C. The temperature range of elasticity for silicone rubber is from −60 degrees C. to 250 degrees; the lower temperature marks the onset of crystallization. The flexibility of the bulk polymer is evidence of the ease with which the backbone bonds can bend [*see top*

© 1974 SCIENTIFIC AMERICAN, INC

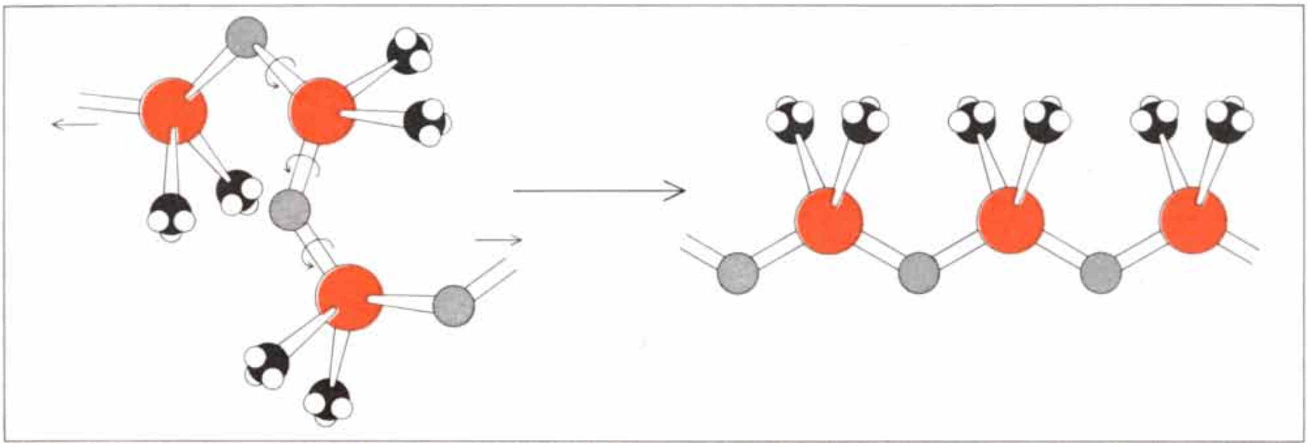


EIGHT-MEMBER RING OF SULFUR ATOMS characterizes ordinary crystals of rhombic sulfur. When heated above 140 degrees Celsius, the eight-member ring opens and joins with other opened rings to form polymeric plastic sulfur (*right*). When the polymer is cooled to room temperature, it slowly reverts back to the cyclic form. Quenching of polymeric sulfur in water is shown on page 74.

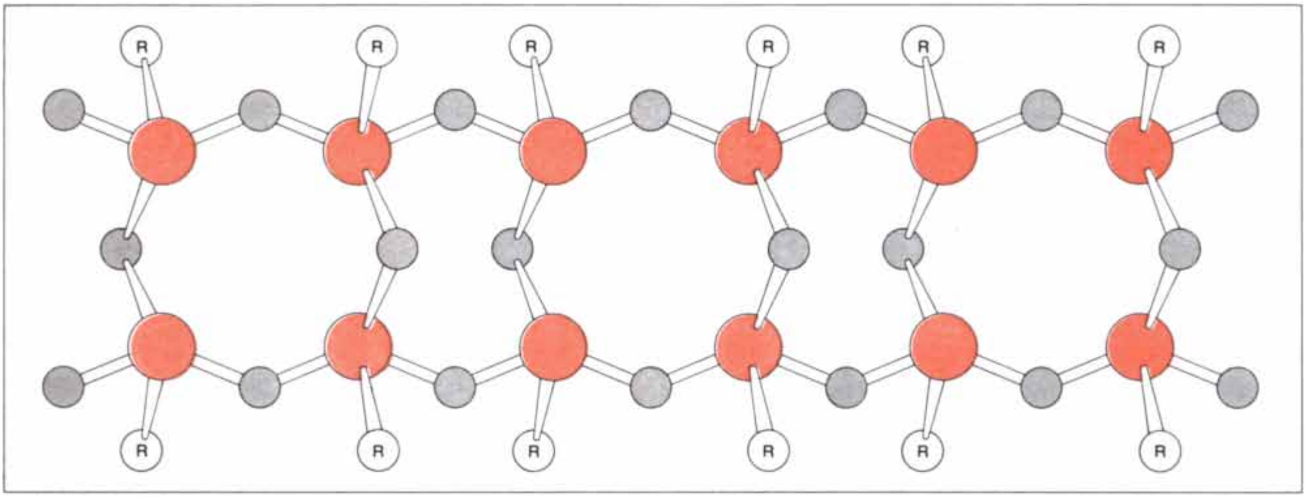


SILICONE POLYMER (*right*) is made by polymerization of octamethylcyclotetrasiloxane (*left*), which consists of eight-member rings in which four silicon atoms (*color*) alternate with four oxygen atoms. Each silicon atom carries two methyl (CH_3) side groups. The linear polymer can be converted to silicone rubber by forming cross-links between side groups. Cross-links prevent the chains from slipping past one another when the material is stretched. Silicone was the first commercial polymer with an inorganic backbone.

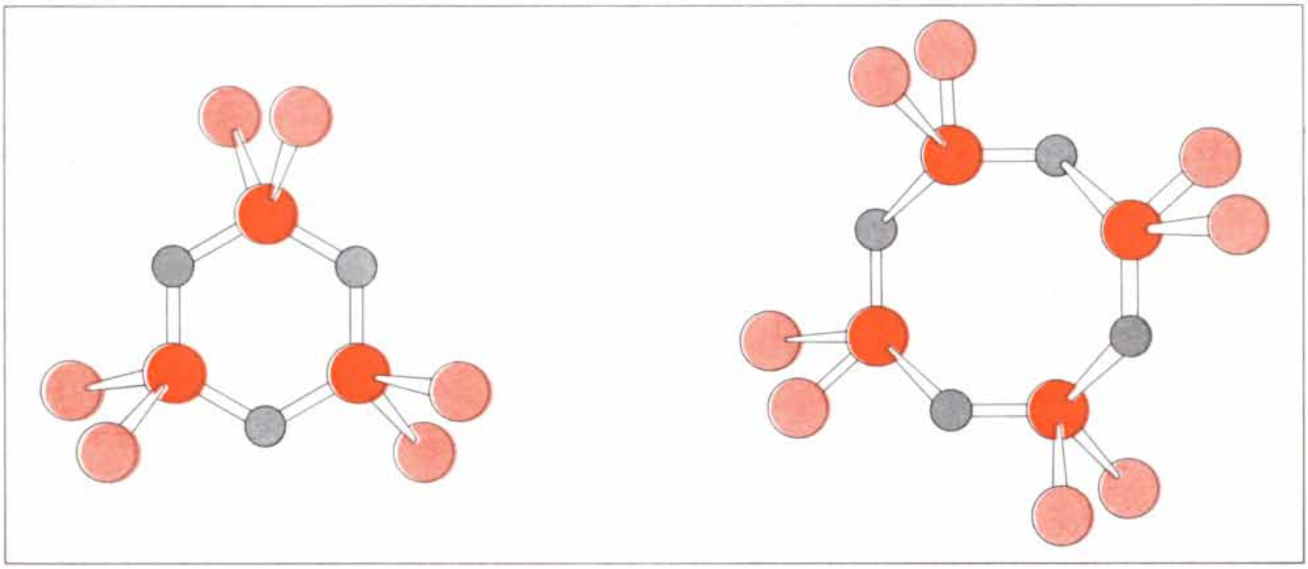
© 1974 SCIENTIFIC AMERICAN, INC



ELASTICITY OF POLYMER such as silicone rubber depends on the ease with which the chains can be elongated from a random coil. The elongation of the chains does not result from the distortion of the bond angles but rather from the unwinding of bonds.



SILICONE LADDER POLYMERS are nonionic analogues of the amphibole silicates illustrated on page 67. The double-strand polymer is more rigid than the single-strand one. Here *R* stands for a phenyl group (C_6H_5), a benzene ring minus one hydrogen atom.



TWO PRECURSORS OF INORGANIC RUBBER are rings in which phosphorus atoms (*dark color*) alternate with nitrogen atoms (*gray*); two chlorine atoms are attached to each phosphorus. The rings have the general formula $(NPCl_2)_n$, where *n* is 3 or any higher integer. The simplest member of the series, the trimer (*left*), has a ring containing six atoms. The tetramer (*right*) has an eight-member ring. On being heated the rings open and both compounds polymerize to form rubbery poly(dichlorophosphazene).

© 1974 SCIENTIFIC AMERICAN, INC

illustration on opposite page]. Indeed, silicone polymers are among the most flexible giant molecules known.

Their high mobility can be attributed to the lack of electric charge on the side groups and to the fact that the side groups are attached to every other skeletal atom instead of to every skeletal atom as in many organic polymers. The latter characteristic means that there are far fewer opportunities for the side groups to collide with each other or even to attract or repel each other as the backbone bonds go through their bending motions. It appears that the extreme flexibility of the silicone backbone makes silicone rubber highly permeable to oxygen. Thus silicone films have been tested in "artificial gill" devices that would extract dissolved oxygen from water for diving purposes.

Silicones also repel water strongly. Partly because of this property, silicones are used in car polishes, in antistick formulations for cooking purposes and in biomedical devices. For example, artificial heart valves and experimental heart-bypass pumps are often fabricated from silicone rubber because the polymer has a lower tendency than most organic polymers to trigger the clotting of blood or to irritate tissues.

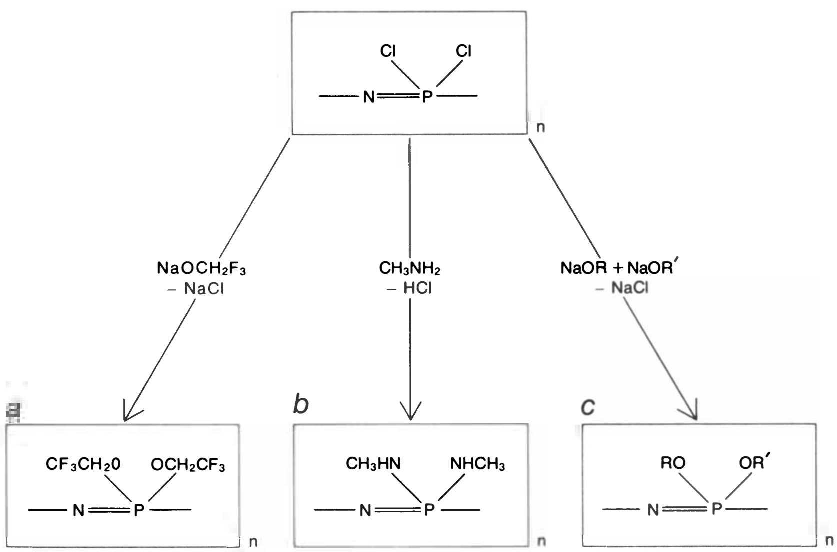
As I have indicated, one of the motivations for the development of polymers with inorganic backbones was the belief that these materials would be more stable at high temperatures than organic polymers. Although silicones are certainly more resistant to oxidation at high temperatures, they suffer from one drawback: at temperatures above 250 degrees C. the siloxane chains break down to form rings, so that the advantageous properties of the polymer are eventually lost. The depolymerization is similar to that observed with polymeric sulfur.

An obvious way around this problem is to design polymers that resemble the amphiboles, or double-chain silicates. Nonionic analogues of such polymers were first made at the General Electric Research and Development Laboratory, with phenyl groups in place of the charged side-group oxygen atoms. The resulting materials are called silicone ladder polymers [*see middle illustration on opposite page*]. As might be expected, the double-chain structure restricts the mobility of the silicon-oxygen bonds in the backbone; the polymers are high-melting, nonelastomeric materials. When they are dissolved in organic solvents, they yield viscous solutions. Polymers that swell with the addition of small amounts of solvents can be stretched and oriented. Phenylsilicone ladder polymers remain stable up to a temperature of 300 degrees C.

In our laboratory at Pennsylvania State University we have synthesized and studied members of a different, and probably more versatile, group of inorganic polymers called polyphosphazenes. These polymers have alternating phosphorus and nitrogen atoms in their backbone. It has been known for a century that two simple inorganic compounds, phosphorus pentachloride and ammonium chloride, react together to form a series of cyclic inorganic compounds. The products have the general formula $(NPCl_2)_n$, where n is 3, 4, 5, 6, 7.... The compounds have ring structures consisting of alternating nitrogen and phosphorus atoms with two chlorine atoms attached to each phosphorus. The cyclic trimer ($n = 3$) and tetramer ($n = 4$) are formed in the largest amounts [*see bottom illustration on opposite page*]. These two compounds are stable, white solids that melt respectively at 114 and 128 degrees C. and that are soluble in organic solvents. When the molten compounds are heated at 250 degrees C. for several hours, they polymerize to a transparent, rubbery high polymer. The resulting material is known as "inorganic rubber," or poly(dichlorophosphazene). Since it contains at least 15,000 phosphorus-nitrogen units in the chain, its molecular weight is more than two million.

Considering that poly(dichlorophosphazene) is made from purely inorganic materials, it is a remarkable compound. In its stress-relaxation behavior it is a better elastomer than natural rubber. Perhaps even more interesting is that it remains rubbery when it is cooled and hardens only when the temperature falls to near the glass-transition temperature: −63 degrees C. Thus, as in the case of silicone rubber, chains of inorganic atoms can certainly yield polymers with high chain flexibility.

Unfortunately poly(dichlorophosphazene) has one chemical property that rules out its use as an everyday plastic. It reacts slowly with atmospheric moisture to yield phosphoric acid, ammonia and hydrochloric acid. The elastomer then crumbles to a powder. We believed this hydrolytic breakdown was initiated by a reaction of the phosphorus-chlorine bonds with water rather than by a direct breakdown of the chain. Therefore we



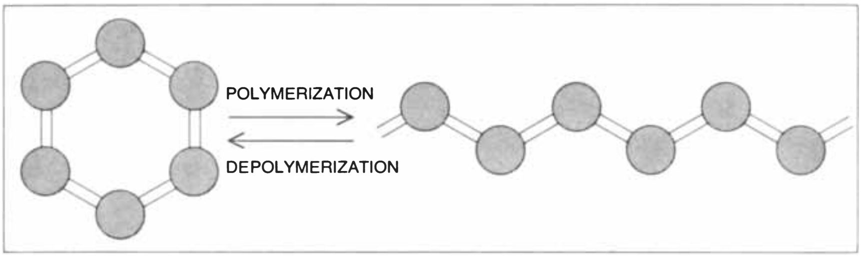
FILM-FORMING, FLEXIBLE, CRYSTALLINE THERMOPLASTIC

WATER-SOLUBLE POLYMER

AMORPHOUS ELASTOMER

SUBSTITUTED POLYPHOSPHAZENES have been made in the author's laboratory at Pennsylvania State University in a search for analogues of poly(dichlorophosphazene) that resist the action of moisture. The trick is to replace the chlorine atoms with more stable organic side groups. This can be done by treating poly(dichlorophosphazene) (*top*) with sodium salts of alcohols or phenols (*a*, *c*), in which case the chlorine is removed as sodium chloride, or with an amine such as methylamine (*b*), in which case the chlorine is removed as hydrogen chloride. The product of reaction *a* is a film-forming, crystalline thermoplastic; the product of *b* is a water-soluble polymer. Mixed substituent products from reactions of type *c*, where two different sodium alkoxides (NaOR and NaOR′) are employed, are usually amorphous elastomers. They are currently being developed for technological purposes.

© 1974 SCIENTIFIC AMERICAN, INC



DEPOLYMERIZATION, the conversion of a polymer chain to a ring, places an upper limit on the temperature at which many polymers created from inorganic ring compounds can be used. Both depolymerization and polymerization reflect the tendency of bonds of inorganic compounds to become "scrambled" when the compounds are subjected to elevated temperatures. The task of the polymer chemist is to find ways to enhance stability of chains.

reasoned that if the chlorine atoms were replaced by nonhydrolyzable groups, decomposition of the polymer should be prevented. The chief hurdle to be overcome was the need to devise a method for the synthesis of phosphazene high polymers that incorporated organic side groups.

We tried two approaches to the problem. The first was based on the argument that if the cyclic trimeric chlorophosphazene can be polymerized by heating, then perhaps cyclic phosphazenes that contained organic side groups might behave similarly. Accordingly we prepared a number of cyclic phosphazenes where the chlorine atoms had been replaced by organic groups such as phenyl, phenoxy, phenylamino, ethoxy and trifluoroethoxy. None of these compounds could be polymerized. We now believe there are both mechanical and thermodynamic reasons for this failure.

The second approach was more successful. It involved a search for a method for the direct replacement of the chlorine atoms in poly(dichlorophosphazene) by organic substituent groups. A careful investigation showed that high-polymeric poly(dichlorophosphazene) exists in two forms: a linear- or branched-chain structure without cross-links and a cross-linked modification. The cross-linked form swells in organic solvents, such as benzene, but it does not dissolve. Therefore this form cannot be used for substitution reactions. The un-cross-linked form dissolves in organic solvents to form viscous solutions. By careful control of the polymerization conditions, un-cross-linked polymer can be isolated in high yield, but prolonged polymerization brings about an almost total conversion to the cross-linked form. Avoiding this cross-linking reaction is the key to the entire process.

Solutions of the un-cross-linked polymer can be treated with reagents that will replace the chlorine. Sodium ethoxide, sodium trifluoroethoxide, sodium phenoxide and various amines, such as methylamine and aniline, can be used as reagents. Under the correct conditions all the chlorine atoms can be replaced [*see illustration on preceding page*].

The polymers prepared by this method are very resistant to hydrolysis. In our laboratory we have synthesized more than 40 different polymers by this method. Depending on the type of side group introduced, the polymers vary from low-temperature elastomers to flexible, film-forming thermoplastics or transparent glasses. Different side groups confer different surface properties, selective solubilities and, in some cases, unexpected solvent resistance on the polymers. In fact, it now seems likely that almost any set of required properties can be designed into the polymer by a judicious choice of the side groups. This almost unprecedented versatility of the phosphazene polymer system is its most remarkable feature.

Two examples will serve to illustrate this point. When poly(dichlorophosphazene) reacts with sodium trifluoroethoxide, the chlorine atoms are replaced by trifluoroethoxy (CF_3CH_2O–) groups. The polymer is called poly[bis(trifluoroethoxy)phosphazene], which has the formula $[NP(OCH_2CF_3)_2]_n$. The polymer can be fabricated into colorless, opalescent, flexible films, which superficially resemble polyethylene films in appearance. Solutions of the polymer can be extruded into a nonsolvent to yield flexible, slightly elastic fibers. The polymer has a low glass-transition temperature (−66 degrees C.), and it remains flexible from this temperature up to its melting point at 242 degrees C. The presence of the fluorinated side groups makes the polymer highly water-repellent, more so in fact than Teflon or silicones.

The opalescent appearance of films of this polymer, and indeed its thermoplastic properties, are a consequence of its high crystallinity: the regular packing together of sections of adjacent chains. Polymers that are noncrystalline tend to be elastomeric above their glass-transition temperatures. If they are crystalline, they form tough, flexible films or fibers.

The crystallinity of a polymer can usually be destroyed by the random introduction of two or more different substituent groups into the structure. If both trifluoroethoxy and, say, heptafluorobutoxy groups are attached to the phosphazene chain, the polymer will be a rubbery elastomer instead of a flexible thermoplastic. Elastomers of this type are currently under development at the Firestone Tire & Rubber Company for use as gaskets, O rings and low-temperature fuel lines and in other applications where resistance to solvents is required.

A second example of a polyphosphazene structure is provided by the polymer prepared by the reaction of poly(dichlorophosphazene) with methylamine. The product is a clear, transparent, film-forming thermoplastic with the formula $[NP(NHCH_3)_2]_n$. This compound differs from nearly all other synthetic polymers in being soluble in water. Amino substituent groups tend to generate hydrophilic (water-loving) properties rather than hydrophobic (water-hating) ones.

Polyorganophosphazenes, in common with silicones, suffer from one defect. Above 200 degrees C. the polymer chains break down to yield small rings, and particularly to form the cyclic trimer and tetramer. Because the useful properties of a polymer depend almost entirely on the presence of long chains, a depolymerization process such as this one is highly disadvantageous. Depolymerization places a serious upper limit on the temperatures at which many inorganic polymers can be used. The blocking of thermal depolymerization constitutes one of the main challenges for research with inorganic polymers.

The breakdown of inorganic polymer chains to form rings is a consequence of the tendency of many inorganic compounds to undergo "scrambling" of their bonds at elevated temperatures. In principle we can say that the polymerization of rings and the depolymerization of chains are simply different aspects of the same backbone-scrambling process [*see illustration on this page*]. The question one must ask is this: Which molecular feature in different inorganic systems might stabilize rings more than chains or vice versa? If one knew the answer to

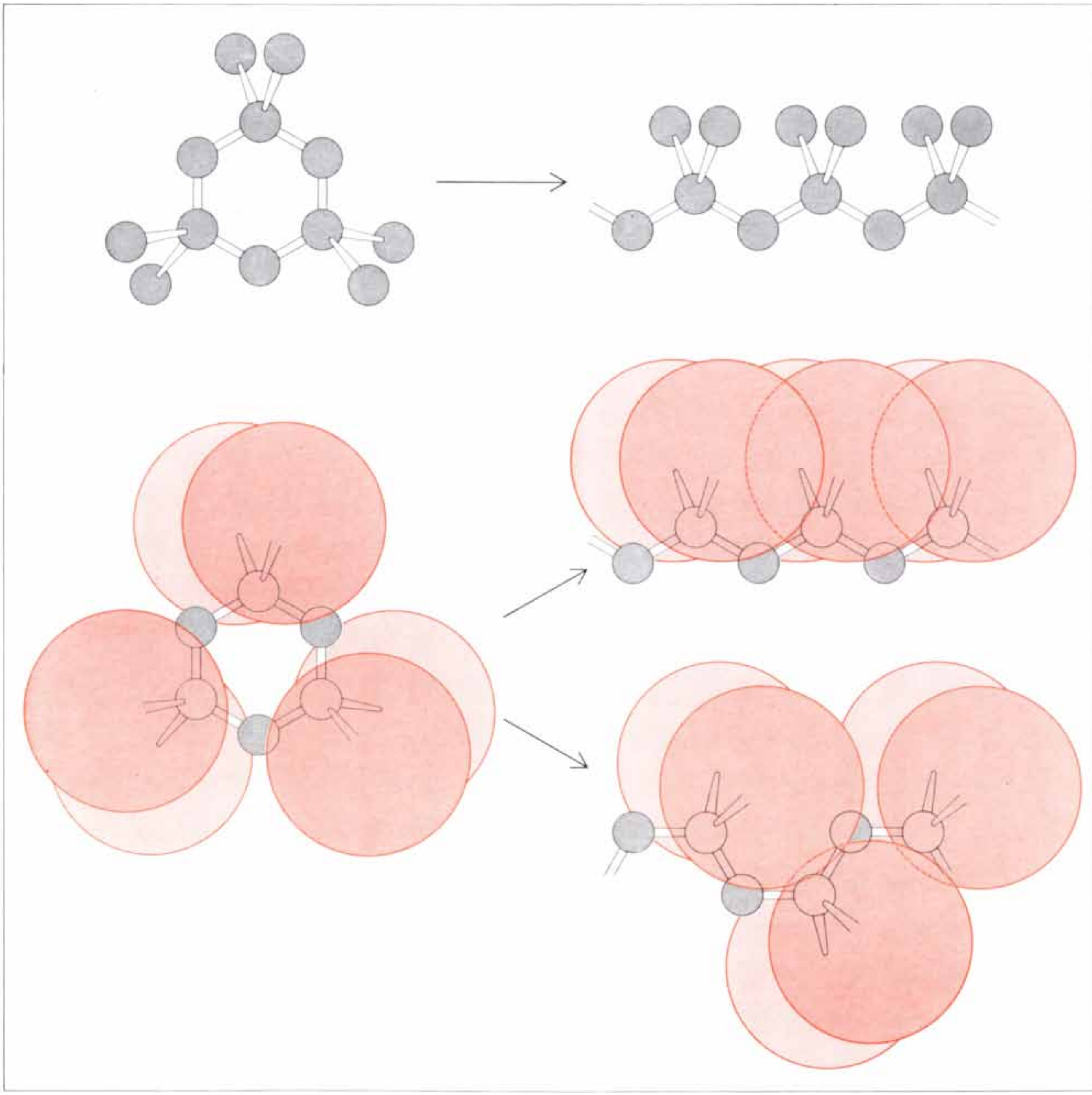
© 1974 SCIENTIFIC AMERICAN, INC

this question, one would be in a position to understand why some inorganic rings cannot be polymerized and why some inorganic chains are more thermally stable than others.

A great many factors actually bear on the question, but three influences are particularly important. Consider first some of the organic and inorganic ring compounds that so far have not been polymerized into linear chains. For example, benzene (C_6H_6) cannot be polymerized; in fact, the benzene ring is one of the most stable units known to chemistry. Similarly, *s*-triazines, borazines and the cyclic sulfur-nitrogen compounds known as thiazyl halides and sulfanuric halides appear to resist polymerization. One feature that all these ring systems have in common is that they are stabilized to a greater or lesser degree by an "aromatic" electron configuration. This arrangement is characteristic of electrons in benzene and related ring systems. Hence there is no inducement for the rings to open to form chains. Linear high polymers apparently cannot be stabilized as much by this effect.

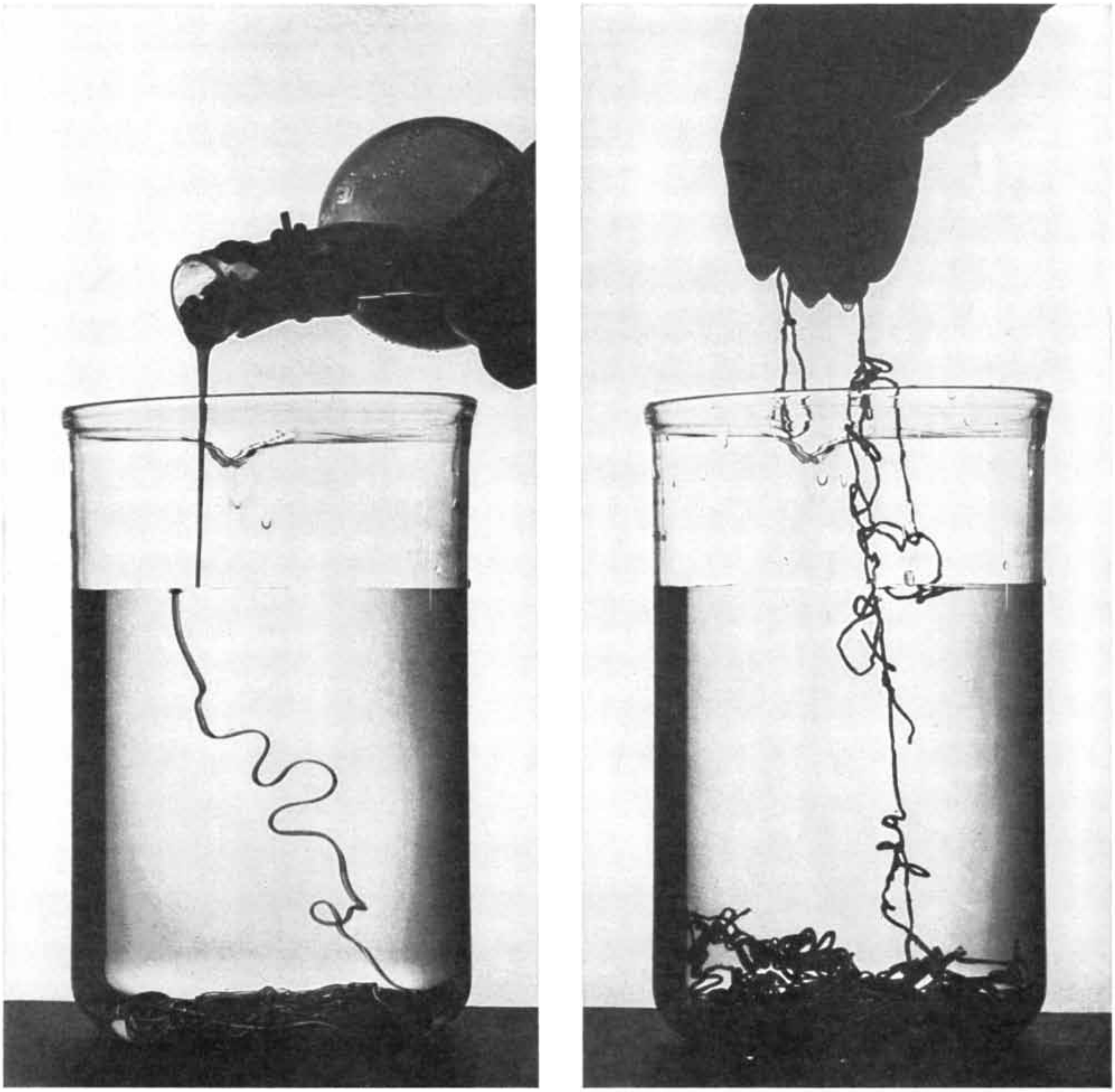
A second factor that bears on the problem involves the size of the side groups attached to the main chain. As we have already seen, the cyclic chlorophosphazenes polymerize to form long chains, whereas the organic-substituted rings so far have not been polymerized. Furthermore, experiments have shown that when organophosphazene polymers are made by the alternative route described above, they depolymerize to rings at a lower temperature than the chlorophosphazene high polymer does.

The tendency for certain polymers to



CHAIN STABILITY is favored by small side groups (*top*), which do not interfere with each other when a ring is polymerized into a chain. If a ring contains bulky side groups (*bottom*), polymerization becomes difficult if not impossible because the side groups interfere so strongly. Moreover, chains with bulky side groups tend to depolymerize at lower temperatures than chains with small ones.

© 1974 SCIENTIFIC AMERICAN, INC



POLYMERIC SULFUR can be prepared by heating ordinary rhombic sulfur to a temperature between 140 and 170 degrees C. If the polymer is quenched by pouring it into cold water (*photograph at left*), it will retain its flexible and elastic properties for a short time at room temperature (*right*). Eventually, however, the polymerized material will revert to its original ring form.

depolymerize depends largely on the size of the side groups, since these groups are more likely to interfere with each other when a compound is in the form of a chain than when it is in the form of a ring [*see illustration on preceding page*]. If the side groups are large organic units, they will collide with, or at least repel, each other strongly in the polymer, whereas they will avoid each other completely in a small cyclic molecule. Thus conversion of a trimer ring to a polymer will be difficult; conversely, if the polymer can be prepared by some other indirect route, it will be less stable than the rings. Given an opportunity, it will break down to form rings because this process readily allows the system to lower its energy. That is probably one of the more important reasons why organophosphazene polymers and some organosiloxane polymers depolymerize at high temperatures.

A third reason for the depolymerization of inorganic polymers and indeed many organic polymers as well is connected with the entropy of the system. The entropy is the degree of disorder of the system, and (particularly at high temperatures) systems change to generate the greatest degree of disorder. The breakdown of one polymer molecule to, say, 10,000 rings is a process that markedly increases the disorder of the system. Consequently high temperatures generally favor the conversion of long chains to small rings.

These facts illustrate that the synthesis of useful inorganic polymers is a complex and challenging problem. Yet those high polymers that have been made are often so strikingly different from most organic polymers that the search for new inorganic giant molecules will undoubtedly continue. At present this field is still in its infancy with many more elements and literally hundreds of elemental combinations yet to be explored.

© 1974 SCIENTIFIC AMERICAN, INC

These days, maybe you should...

GET A HORSE

...and a piece of the environment that goes with it.

As inflation eats into the value of our dollars, and the energy crisis halts our roaming, more people than ever are taking a fresh look at the special benefits for people who own their own ranchland in America's great southwest.

Here there are all kinds of great recreation and vacation activities. If you enjoy riding, fishing, hunting, boating, skiing, or just the great outdoors, you will want to give serious thought to the wisdom of using some of today's nestegg or earnings to assure tomorrow's way of life and enjoyment for your family.

Forbes Magazine owns the 260-square-mile Trinchera Ranch in southern Colorado and is offering for sale minimum five-acre ranch sites to those who love breath-taking mountain scenery, fresh air, wide open spaces and big country living. Prices start at $5,000 per ranch site. Cash or easy credit terms with interest at 6% *Annual Percentage Rate.*

For full details on this land offering, without obligation, please fill in the coupon and mail it today.

Sangre de Cristo Ranches Inc.

Division of Forbes Magazine
Box 2000, Old Chelsea Station, New York, N.Y. 10011

Name ______________________

Address ______________________

City __________ *State* __________ *Zip* __________

Telephone ______________________

Prefer: ☐ $5,000 ranches ☐ $7,500 ranches

A statement and offering statement has been filed with the Department of State of the State of New York. The filing does not constitute approval of the sale or lease or offer for sale or lease by the Department of State or any officer thereof or that the Department of State has in any way passed upon the merits of such offering. A copy of the offering statement is available, upon request, from the subdivider. NYA 74-31. Not available to residents of states where prohibited by law.

AD 10720(A)

Obtain HUD property report from developer and read it before signing anything. HUD neither approves the merits of the offering nor the value of the property as an investment, if any.

A780

© 1974 SCIENTIFIC AMERICAN, INC

VORTEXES IN AIRCRAFT WAKES

The advent of the "jumbo" jets has turned the well-known aerodynamic phenomenon of trailing-vortex air turbulence into a potentially serious hazard to smaller following aircraft

by Norman A. Chigier

Next time you watch a large, four-engine jet aircraft take off from an airport look carefully and you will probably see that the exhausts from the engines appear initially as four distinct smoke trails and that they subsequently merge into two trails, separated by a distance approximately equal to the wingspan of the aircraft. The smoke trails act as tracers, enabling one to visualize the complex aerodynamic process by which a pair of wing-trailing vortexes is formed in the wake of every aircraft. The presence of trailing vortexes can also be seen in the long white "contrails" that develop behind a jet aircraft flying at high altitude on a clear day. In this case the vortexes are made visible by the fact that water vapor in the engine exhausts freezes, forming ice crystals that are subsequently induced into the trailing-vortex system.

The phenomenon of wing-trailing vortexes in aircraft wakes has been known almost since the birth of aerodynamics as a separate discipline. Indeed, the formation of the vortexes has long been recognized as being directly associated with the lift generated by the wings. The potential danger of this phenomenon, on the other hand, has become apparent only within the past few years. The advent of the wide-bodied "jumbo" jets, such as the Boeing 747, the Lockheed C-5A and the McDonnell Douglas DC-10, coupled with an increase in the number of small, lightweight aircraft, has turned the problem of trailing-vortex air turbulence into a serious concern of the Federal Aviation Administration. The number of fatal crashes attributable to the interaction of aircraft with the vortex wakes of other aircraft has increased significantly in recent years.

Although the large majority of such crashes have involved small aircraft, the crash in 1972 of a DC-9 in the wake of a DC-10, which led to the total destruction of the former aircraft along with its training crew, demonstrated that medium-size aircraft are now also vulnerable. As a result of this crash the National Transportation Safety Board recommended to the FAA that "separation of controlled aircraft should be increased to at least three minutes whenever meteorological conditions conducive to vortices exist." Three minutes between aircraft is equivalent to a spacing of approximately seven miles.

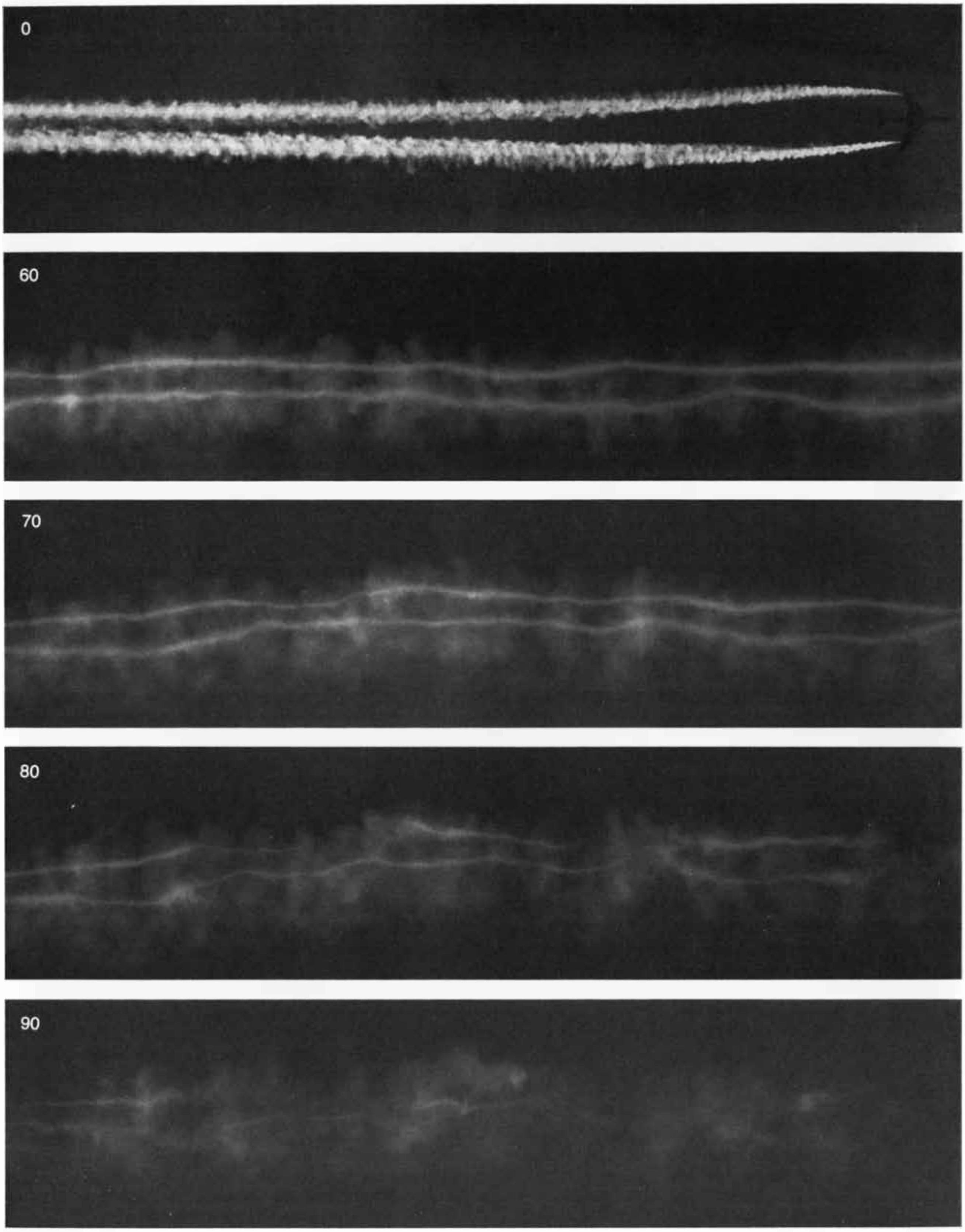
What can be done about this problem? The main objectives of current research programs have been (1) to ascertain the nature, strength and persistence of the vortexes for various kinds of aircraft; (2) to develop a practical means of reducing the hazard on existing aircraft by artificially inducing the vortexes to break up, and (3) to set up monitoring and control systems at airports in order to prevent aircraft from entering the vortex-wake turbulence of larger aircraft.

Physical models of the aerodynamic flow fields around aircraft have been developed on the basis of flow-visualization studies, theoretical predictions and direct measurements of velocity and pressure on the aircraft's surfaces. For an aircraft in flight the air approaching the aircraft is deflected upward, and the resulting curvature of the streamlines produces a net vertical force that gives the lift to the aircraft. At the wing tips a boundary layer is set up that separates from the wing surface and rolls up into a spiral vortex sheet over the top surface of the wing. The streamlines over the wing's surface converge so that, at some short distance behind the trailing edge of the aircraft, the dominant flow pattern is a pair of vortexes [*see illustration on page 80*]. Each of these vortexes is similar in form to the vortex set up when the drain of a bathtub is opened. In both cases the vortex consists of a central core region that is in "solid-body rotation," with the fluid particles rotating around the axis as though they were attached to a revolving disk. The maximum circumferential velocities are found at the outer edge of the core, and thereafter these velocities decrease exponentially with increasing radius from the axis. The strength of a trailing vortex is directly dependent on the weight of an aircraft; therefore, since the larger aircraft of today are approaching a weight on the order of a million pounds, the energy con-

VORTEXES ARE STUDIED at a test facility of the Federal Aviation Administration near Atlantic City, N.J. In the top photograph on the opposite page a Lockheed L-1011 "jumbo" jet is seen flying at an altitude of 200 feet past the far side of an instrumented test tower, which is emitting layers of colored smoke in a fairly steady downwind pattern. In the photograph at bottom the swirling smoke trails outline the turbulent vortex trailing from one of the aircraft's wing tips. Visible at the center of the large circular pattern is the long thin core of the vortex, which rotates at a very high velocity and expands only gradually with time. Moments later the vortex from the other wing tip will drift over and the smoke trails will suddenly rotate in the opposite direction. In addition to the smoke-visualization apparatus the 140-foot tower is equipped with hot-film air-velocity sensors at one-foot intervals; data from these devices are recorded on magnetic tape. Aircraft-wake-turbulence studies at the FAA experimental center are conducted under the direction of Leo J. Garodz.

© 1974 SCIENTIFIC AMERICAN, INC

© 1974 SCIENTIFIC AMERICAN, INC



ABRUPT DISSIPATION of a highly stable pair of wing-trailing vortexes behind an aircraft is believed to result from the growth of small "hydrodynamic" instabilities in the vortex cores. The number in the upper left-hand corner of each photograph gives the age of the vortexes in seconds. The photographs were made in the course of a recent test program conducted by the FAA to study the persistence and characteristics of the vortexes as a function of aircraft configuration, airspeed and altitude. In this particular test the aircraft, a Convair 880, was flown at an altitude of about 5,000 feet and at an airspeed of 160 knots. The vortexes are defined by vaporized oil added to the exhausts of the two outer jet engines. Time-sequence photographs such as these have suggested that if instabilities could be excited artificially in the wakes of large aircraft, they could hasten the breakup of the hazardous vortex structure.

© 1974 SCIENTIFIC AMERICAN, INC

tained within the vortexes can be considerable.

One can calculate the theoretical strength of an aircraft-wake vortex on the basis of standard aerodynamic theories, which derive their results from the distribution of pressure over the wing surfaces. Unfortunately these theories, which are mainly two-dimensional and usually do not take into account the effects of viscosity, are not able to provide information on the length and persistence of the vortexes or on the structure of the core region. In recent years measurements of the flow fields in aircraft wakes have been made in wind tunnels, water tanks and under flight conditions. The most alarming feature of the in-flight studies has been the persistence of the vortexes with apparently only small changes in strength up to distances as great as 15 miles behind large aircraft!

Measurements of the velocity field have also shown the presence of axial velocities within the core region. Under flight conditions this axial flow has been demonstrated by releasing a number of small helium-filled balloons from the ground while an aircraft was flying overhead. The helium balloons rose under the action of buoyancy and, as they entered the flow field behind the aircraft, began to rotate. Within the centrifugal flow field of the vortex the helium balloons were centrifuged, because of the lower density of helium compared with air, toward the central axis of the vortex. On reaching the axis the helium balloons were observed to follow the aircraft at a speed of approximately a tenth of the aircraft's flight speed. Axial velocities of this order of magnitude have also been measured in wind tunnels and under normal flight conditions. These axial flows are believed to play an important role in maintaining the highly stable conditions within the vortex. The vortexes do not remain in fixed positions in space but instead meander as they are buffeted by the surrounding atmospheric turbulence. This meandering behavior has presented a major difficulty in measuring the strength of the vortexes.

An attempt to estimate the strength of aircraft-wake vortexes was made by my colleagues and me at the Ames Research Center of the National Aeronautics and Space Administration using the 40-by-80-foot wind tunnel there and the rapid-scanning air-velocity sensor called a hot-wire anemometer. The measuring probe, which was capable of recording the three-dimensional components of air velocity, consisted essentially of three electrically heated wires each five microns long, mounted on the end of a rotating arm with a radius of eight feet. The rotating arm was made to pass through the center of the vortex, and the signals from the hot wires were recorded on magnetic tape. Using calibrations of the electrical signal from the heated wires as a function of wind speed, direction and temperature, the velocity components within the vortex could be measured. By making these traverses at various axial stations behind fixed models of aircraft in the wind tunnel the changes of vortex strength could be measured as a function of distance downstream.

A hot-wire anemometer was recently fitted to the boom of a small jet research aircraft by workers at Ames. By passing through the wake of larger aircraft under normal flight conditions, measurements of the velocity fields have been made at various distances behind the generating aircraft. A rapid-scanning laser-anemometer, also developed at Ames, enables one to make two-dimensional velocity-component measurements optically inside wind tunnels and water tanks without introducing any physical disturbance in the flow field. These measurements show that the tightly bound vortexes are highly stable and the rate at which their structure changes is very slow. The main changes that occur are an increase in the diameter of the core with a corresponding decrease in maximum circumferential velocities. The measured vortexes are separated by a distance approximately equal to the wingspan of the aircraft, and the diameter of the vortex core is approximately 3 percent of the wingspan. Measured maximum circumferential velocities are on the order of 30 percent of flight speed.

As aircraft approach airports it is now common for air-traffic controllers to give the warning: "Danger—wake turbulence." In many cases the pilot of the aircraft receiving this message can see no sign of the aircraft that generated the turbulence. He will attempt to follow the FAA operational instructions concerning wake turbulence, but he will be seriously handicapped by not knowing where the vortex is.

From a detailed analysis of a number of accidents involving a total loss of life and aircraft the mechanisms of the interaction between a vortex wake and a following aircraft now seem to be clear. The flow field in the vortex wake can be viewed as an induced upwash at the outer edge and an induced downwash in the region between the vortex pair. Aircraft flying along the same flight path as the generating aircraft and entering directly into the core of one of the vortexes experience a sharp, violent roll, which can easily exceed the roll control of the following aircraft. Severe structural damage can result to a very light aircraft, with complete loss of control of the aircraft by the pilot. When a pilot attempts to gain altitude during takeoff and enters the induced downwash between the vortexes from below, the aircraft can be forced downward.

The third possible form of interaction takes place when the following aircraft crosses the path of the vortex at right angles. The pilot experiences a sudden upwash, and by the time he has adjusted his controls he enters the downwash, which can force him to the ground.

A quantitative measure of the danger to the following aircraft can be given in terms of the strength of the trailing vortex. On the basis of either measured or calculated circumferential velocity distributions the effective rolling moment of the vortex can be calculated over the wingspan of a following aircraft for a variety of encounters. Wind-tunnel and water-tank studies are also being made in which these rolling moments are directly measured on small aircraft models held in the wake of larger aircraft models. These studies support the high danger ratings of aircraft-wake encounters.

In the case of the delta-wing aircraft designed for supersonic flight the vortexes form at the leading edge of the wing and grow in strength over the wing's upper surface. These vortexes contribute significantly to the lift of such aircraft. In-flight measurements are currently being made of the wake of the *Concorde,* the British-French supersonic transport, but little is known about the structure and persistence of this type of wake.

Under normal flight conditions a pair of aircraft-wake vortexes generally move downward and subsequently level off at a certain altitude. This is not always the case, however, and there is some evidence of vortexes maintaining the same altitude or even rising over distances of several miles. The examination of contrails shows that the vortex structure eventually disintegrates. Photographs of such contrails reveal regions of highly stable flow with very little dissipation, followed by a comparatively abrupt disintegration. Many theories have been postulated to explain this phenomenon with hardly any reliable experimental verification other than observation.

Four principal modes of vortex disin-

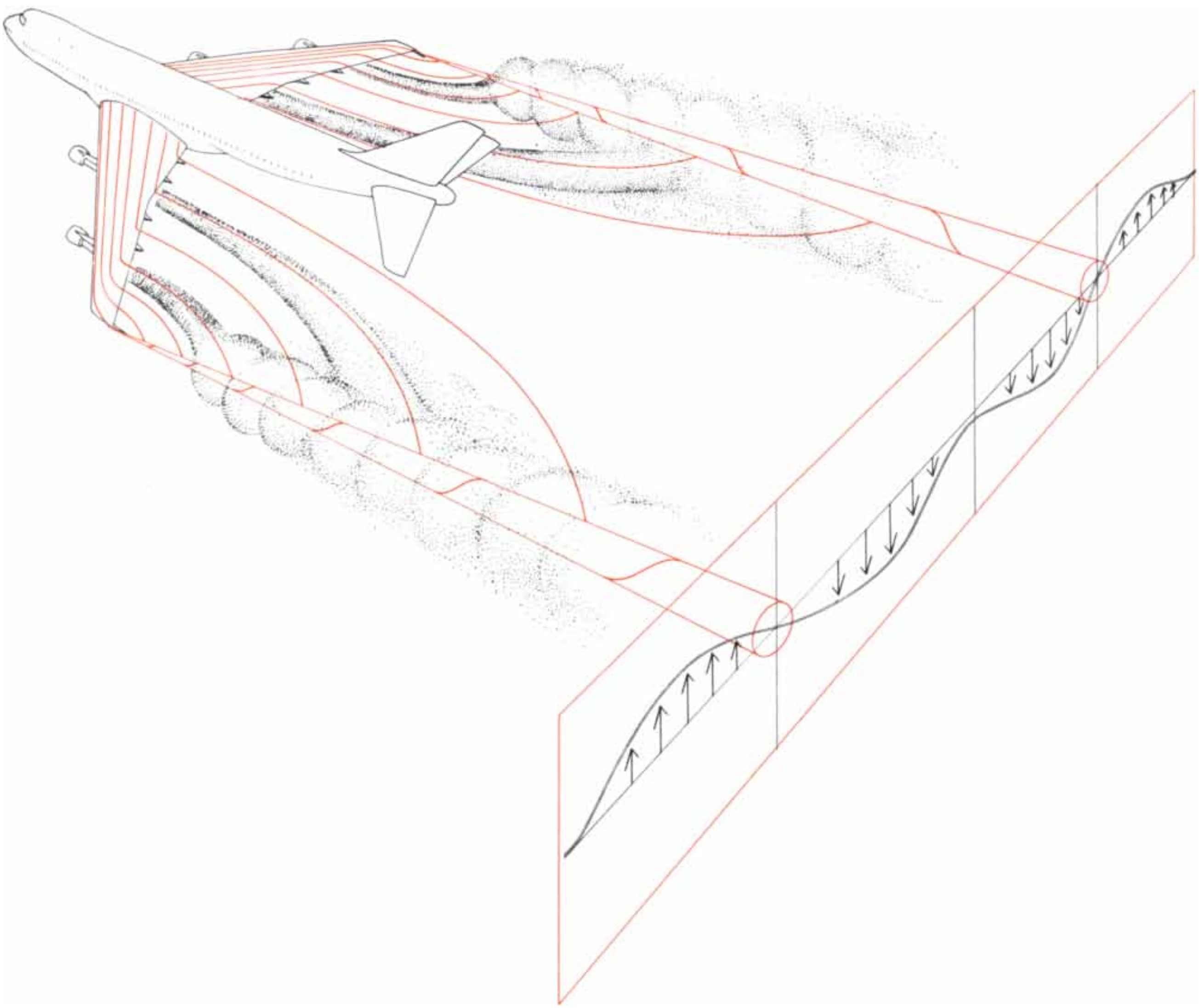
© 1974 SCIENTIFIC AMERICAN, INC

tegration have been identified so far: (1) vortex breakdown, (2) vortex instability, (3) viscous dissipation and (4) atmospheric interaction. Vortex breakdown is simply the sudden expansion of the core of the vortex. The vortex makes a sudden transition from a motion with a tightly centered angular momentum to a motion in which the radius of the core increases by an order of magnitude, with a correspondingly wide distribution of angular momentum. Large pressure gradients are set up across the breakdown, often resulting in the reversal of the flow along the axis of the core. This behavior is thought to be analogous to that of a shock wave. Vortex breakdown has been observed on photographs of liquid dye in the wake of a wing model towed in a water tank. Vortex breakdown has also been observed over the surfaces of delta wings at high angles of incidence.

In a recent study conducted in an Ames wind tunnel I measured the velocity distribution in a vortex that was in the process of breaking down in the wake of a slender delta wing. The measurements, which were made with the aid of a laser, showed substantial reductions in axial and circumferential velocity within the core and also the generation of turbulence. A method is currently being sought to induce vortex breakdown behind wide-bodied transport jets by fitting highly swept surfaces to the wings.

Under very stable conditions small instabilities are observed to grow in amplitude with increasing downstream distance. These "hydrodynamic" instabilities cause an enlargement of the core, and as the vortexes approach each other they begin to interact. In-flight photographs using smoke tracers show the linking together of the two vortexes and the formation of complex loops and rings.

Some degree of small-scale turbulence is always present in trailing vortexes, and this turbulence interacts with velocity gradients to produce local stresses.



PAIRED VORTEXES ARE FORMED in the wake of every aircraft by an aerodynamic process that is directly associated with the lift generated by the aircraft's wings. Air deflected upward by the leading edge of each wing establishes a boundary layer that separates from the wing surface and rolls up into a spiral vortex sheet (*colored outlines*). At some short distance behind the trailing edge of the wing the streamlines of the boundary flow converge, with the result that (regardless of the number or location of the engines mounted on the aircraft) the dominant flow pattern behind any aircraft is a pair of vortexes separated by a distance approximately equal to the wingspan of the aircraft; the diameter of the tightly bound vortex core is roughly 3 percent of the wingspan. Such vortexes are remarkably stable; some have been observed to persist with little change up to 15 miles behind large "jumbo" jets, such as the Boeing 747 depicted here. From the point of view of a following aircraft the flow field in the paired-vortex wake appears as an induced upwash at the outer edges and an induced downwash in the region between the vortex pair (*black arrows at right*).

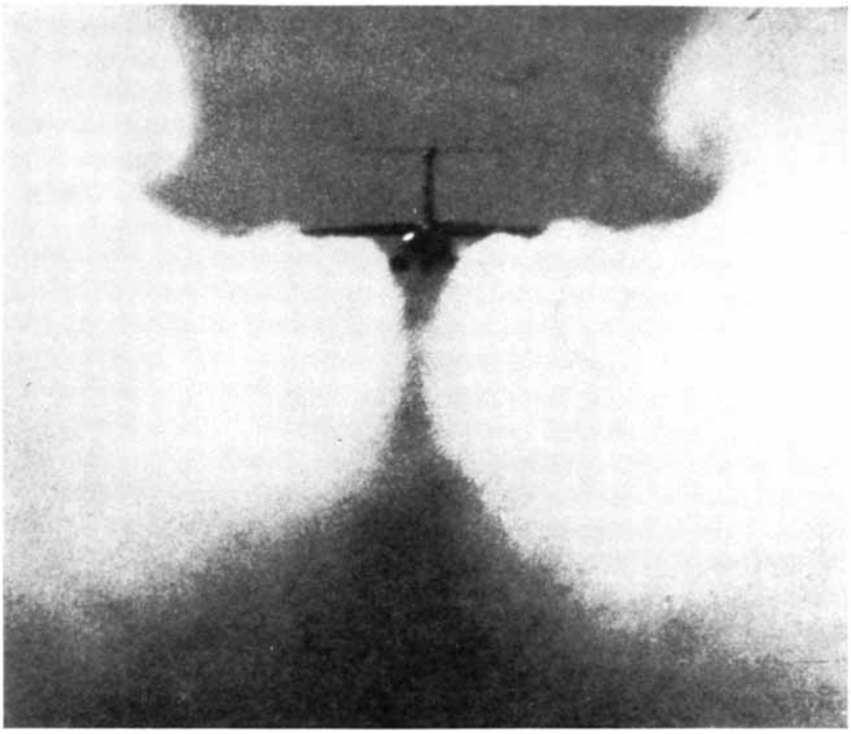
© 1974 SCIENTIFIC AMERICAN, INC

The rotating flow of the vortex has a stabilizing effect that inhibits diffusion and growth and can lead to a damping of the turbulence.

Vortexes are found to be most stable under calm conditions over the ocean. They disintegrate far more rapidly in turbulent atmospheric conditions over a mountainous region. Large-scale atmospheric disturbances will interact with the vortexes; the turbulence in the atmosphere can be diffused into the vortexes, causing them to disintegrate. These various modes of disintegration can proceed in combination. The dominant mode that leads to the final disintegration will depend on the local atmospheric and flight conditions.

In 1970 NASA initiated a basic research program to determine the feasibility of alleviating the vortex hazard. A year later the FAA asked NASA to accelerate its efforts in this direction. During the early stages of the research program opinions were divided between the theoretical aerodynamicists and a few individuals who felt that some means could be found to reduce the hazard. On general aerodynamic grounds it was known that the presence and formation of the vortexes were fundamentally linked with the generation of lift. Since aircraft would inevitably shed strong vortexes, the aerodynamicists in the program concluded that nothing could be done about the problem.

Preliminary studies carried out in wind tunnels demonstrated that the introduction of a large obstacle into the region of a vortex behind the aircraft broke up the concentrated vortex pattern. There was no effect on the lift or drag as measured by force balances on the wing model. It was thus demonstrated that breaking up the vortex from its orderly structured form would not affect the flow patterns upstream of the disturbance. It was also argued that the exact location of vortex disintegration under natural atmospheric conditions would not affect the performance of the aircraft. In practice there is very little loss of angular momentum within each vortex in a pair. The net angular momentum of the total wake, however, is zero, so that the possibility of two vortexes interacting with each other and diffusing their "vorticity" over a large cross-sectional area does not violate the fundamental physical laws of aerodynamics. From a practical point of view, if the vorticity is distributed over a much larger cross-sectional area, the magnitude of velocities becomes very small and the forces exerted on the following aircraft become negligible.



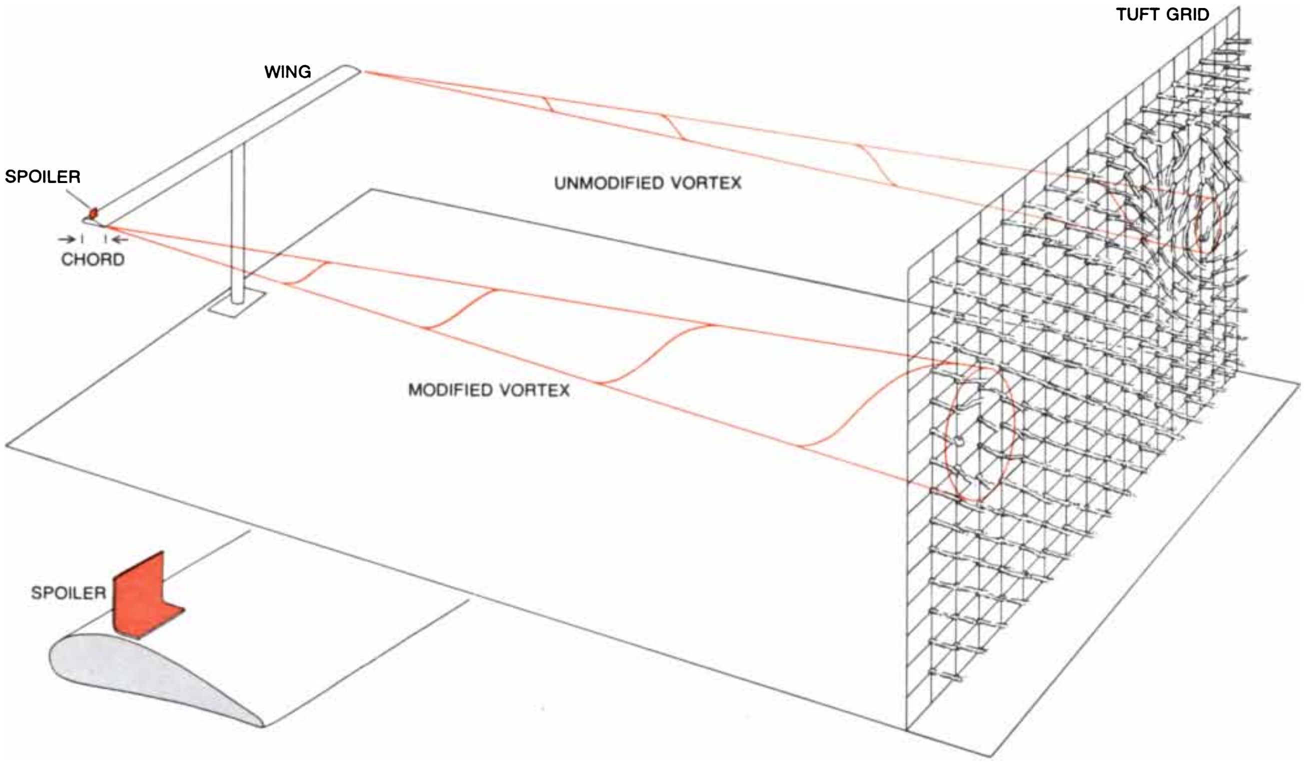
VORTEXES AT HIGH ALTITUDE are made visible by the formation of ice crystals from water vapor in the engine exhausts. The ice crystals are subsequently induced into the trailing-vortex system, generating the familiar white "contrails" visible from great distances on a clear day. The lead aircraft in this FAA test was an Air Force C-141 transport. The photograph, a single frame from a 16-millimeter film strip, was made from a trailing aircraft.

The NASA research program concentrated on a practical means of accelerating the dissipation of the vortexes. The first method was a "brute force" one; it was to create enough physical disturbance within the vortex to break up the orderly stable flow system, which is symmetrical around the axis of the vortex. Wind-tunnel tests conducted at Ames in 1970 showed that by mounting a vertical "spoiler" plate on the top surface of a wing close to the wing tip where the vortex is generated one could reduce the maximum circumferential velocities by 63 percent, with a corresponding increase in the diameter of the core. The spoiler also caused a small reduction in lift and a significant increase in drag.

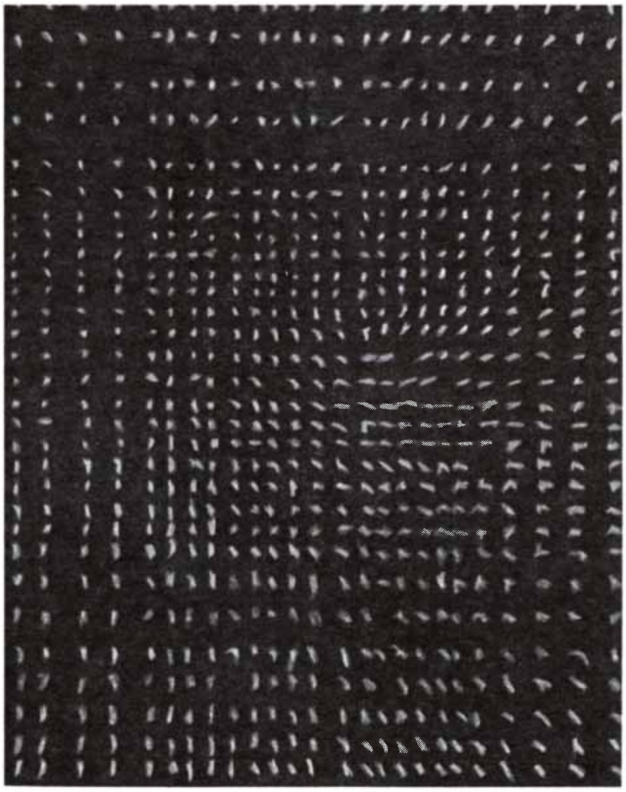
A spoiler was then fitted to the wing of a Convair 990 jet transport. A smaller probe aircraft (a Lear "executive" jet) was flown almost axially along the vortex while measurements were made of the aircraft's dynamic response to the vortex. In an attempt to get a "feel" for the passenger discomfort caused by such an interaction, I sat in the copilot's seat of the Lear jet as penetrations were made at an altitude of 12,000 feet. With the aid of smoke added to the exhaust from the Convair 990 we first positioned ourselves above the vortexes and then descended toward them. The experience was quite different from that of clear-air turbulence, in which aircraft have been reported to drop suddenly. During our flight it appeared to me as though someone had suddenly grabbed the wings and twisted the aircraft violently. Although I was strapped into the copilot's seat, my head hit the roof of the cockpit. The pilot completely lost control of the aircraft. We were thrown out of the vortex, but in a few seconds we entered the adjacent vortex and experienced a similar jolting roll in the opposite direction. Neither the pilot nor I could distinguish significant differences when penetrations were made at distances of five miles, three miles and one mile behind the larger aircraft.

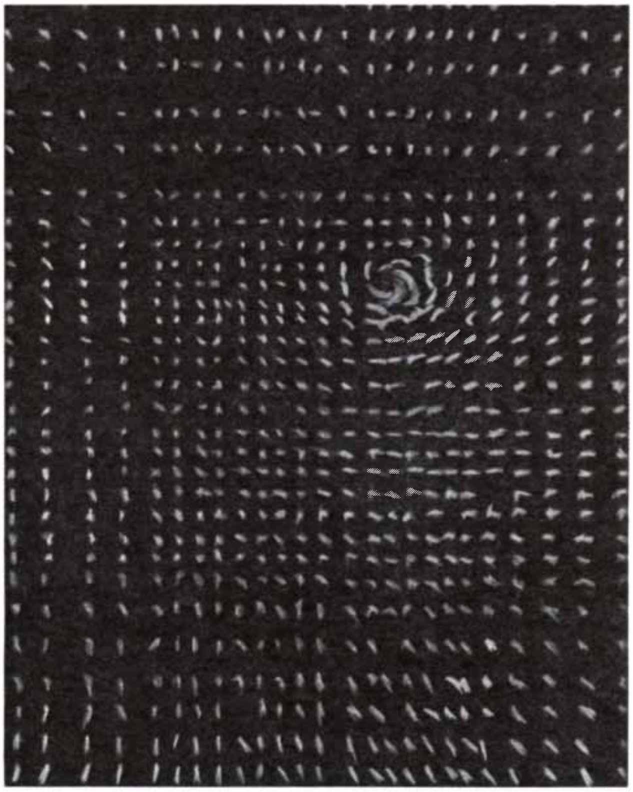
Variations in the deflection of the ailerons, the roll rate and the sideslip angle were recorded in the probe aircraft. They showed that the pilot could maintain control for a slightly longer period and with less disturbance in a vortex generated by a wing fitted with a spoiler. As far as I was concerned, I felt equally sick in both vortexes and would consider

© 1974 SCIENTIFIC AMERICAN, INC



EXPERIMENTAL VORTEX DISSIPATOR, a small vertical "spoiler" plate mounted on the upper surface of a wing near the wing tip, was tested by the author and his colleagues in a wind tunnel at the Ames Research Center of the National Aeronautics and Space Administration. The effect of the spoiler was studied by means of smoke visualization and with the aid of a tuft grid (*right*).





TUFT-GRID PHOTOGRAPHS revealed that under certain conditions the installation of a spoiler could reduce the maximum circumferential velocity of a vortex substantially with a corresponding increase in the diameter of the core. The spoiler also caused a small reduction in lift and a significant increase in drag. The modified vortex pattern appears at left, the unmodified vortex pattern at right. The distance from the wing to the tuft grid in both cases was 40 times the chord of the wing. In-flight tests of spoiler-type devices have so far proved inconclusive. Moreover, there is some evidence that the vortex may re-form downstream from the disturbance.

© 1974 SCIENTIFIC AMERICAN, INC

that any dissipating effect that had been caused would not be sufficient to be considered acceptable for normal passenger flying.

Since these preliminary tests were made a variety of devices have been fitted to aircraft models and tested in wind tunnels and water tanks. The alternatives tested include blowing jets into the vortex core, fitting drogues (small "air anchors") behind the trailing edge of the wing, fitting fixed panels at a variety of locations over the surface of the wing and fixing flaps at various angles at the trailing edge. All these devices generate turbulence, and the convection and diffusion of this turbulence within the vortex cause the core to spread out. Even though the experiments are still in progress, there is evidence that the use of flaps or slats to offload the wing tips and to produce ripples in the wingspan loading reduces the rolling moment of the wake. Such devices will introduce additional drag, but this may not necessarily be prohibitive. Water-tank tests show that if the vortex is spread out near the wing, it does not decay as fast as an unmodified vortex does, so that at an equivalent distance of five miles behind an aircraft the maximum vortex-induced velocities are the same. What this means is that a considerable improvement may be achievable for distances of less than five miles.

Attempts to introduce instabilities in the vortex may ultimately provide a more promising solution. Oscillating alternate flaps sinusoidally, so that there is no net reduction in lift, causes the vortexes to oscillate and interact with one another in such a way as to ultimately form loops at distances shorter than if the flaps were not actuated. One pilot has been able to achieve rapid disintegration of the vortexes by oscillating the control lever in the generating aircraft; in this case, however, passenger discomfort has simply been transferred to the generating aircraft!

Still another approach that has been tried is to change the wingspan loading, a procedure that involves changing the shape of the wing and thereby the distribution of pressure over its surfaces. Theoretically it can be demonstrated that it is possible to design a wing in which the streamlines will not roll up into a pair of concentrated vortexes. The C-5A airplane already has uprigged ailerons and segmented trailing-edge flaps (both for structural reasons); these are two practical ways of producing the "tailored" loadings capable of alleviating the vortex hazard.

Meanwhile pilots are reporting encounters with vortex-wake turbulence with disturbing frequency, and litigation cases against the FAA, the agency considered responsible for encounters in the vicinity of airports, continue to mount. Accidents involving loss of life or serious injury have now exceeded 100. Assuming that the easing of the current "energy crisis" allows recent trends to resume, then as more "jumbo" jets come into operation and as more private individuals purchase light aircraft, the situation will become worse. The present custom at U.S. airports is to allow aircraft to take off and land on a "first come, first served" basis, although it is now more common for air-traffic controllers to warn pilots of local traffic movements of the "jumbo" jets. Increasing congestion at the larger airports causes pressure on air-traffic controllers to get aircraft moving in and out of airports as rapidly as possible by trying to reduce the headway between aircraft.

The FAA is now concentrating its efforts on means of detecting and monitoring vortexes in the vicinity of airports. Doppler radar, laser anemometers and acoustic detectors are being developed for measuring both the strength and the location of vortexes. The quantity of smoke and heat in a vortex is too small to cause reflections from standard radar instruments. At present one of the more promising instruments is an acoustic-radar detector, in which a noise signal is transmitted from the ground and is deflected by interaction with the vortex. The deflection of the signal is measured by a detector situated at some distance from the noise generator. A number of such devices have been installed at O'Hare International Airport in Chicago and are being installed at other airports on an experimental basis. The range of detection of these acoustic devices is limited to a few hundred feet, and the noise they generate is annoying.

Detailed designs have been made of computer-assisted control systems, which would be installed in air-traffic-control towers. On the basis of expected traffic movements computer predictions will be made of danger zones, taking into account aircraft weight and flight configurations as well as local wind movements. These predictions will be "updated" as the vortex enters into the range of the detection device. Vulnerable aircraft approaching or leaving the airport will then be directed away from the danger zones. The costs of such a system are large and, if the system were to be considered sufficiently reliable to allow aircraft to fly in closer proximity, many more millions of dollars will have to be spent on developing the system. One school of thought holds that the statistical chance of aircraft interacting with the relatively small vortex-core regions is so remote that a "general avoidance" flight-path directive to pilots would be sufficient, and that it would be necessary to accept the occasional crash with consequent loss of life. If the number of accidents increases, that solution will obviously not be acceptable.

The trend of future developments in aircraft is toward even larger "jumbo" jets. For intercity air travel STOL (short takeoff and landing) and VTOL (vertical takeoff and landing) aircraft are being developed. Preliminary flight tests with "augmentor wing" STOL aircraft indicate that vortex velocities are attenuated faster than would be expected for the equivalent circulation from a wing without powered lift. The most recent types of "jumbo" jets, the DC-10 and the Lockheed L-1011, are three-engined jets with one engine in a central position over the fuselage. Flight tests behind these trijets showed that when the flaps were up, the vortexes were even more intense than they were behind a four-engine aircraft of equivalent weight, thus showing that the location of engine exhaust jets can also affect the stability of the vortexes. Supersonic aircraft flying in the highly stable stratosphere could produce wakes of much greater length than those produced in the more turbulent lower atmosphere. If STOL and VTOL aircraft, including helicopters, become accepted as a major means of intercity transport, traffic density will increase substantially. Although little attention has been given to the wakes of these aircraft, the large lifting forces associated with them should in principle produce strong vortexes. Coherent vortex patterns have been observed to persist for several minutes after a helicopter has passed.

Forecasts of future developments in aircraft design and air-traffic density show that the present hazardous situation will grow worse. Existing efforts to find some means of reducing the strength of the vortexes will need to be intensified. It has already been suggested that the huge amounts of energy the vortexes embody should be harnessed to drive turbines and generate power. As in many other fields of technology, as soon as people become sufficiently concerned for their personal safety, funds will be made available to find a technological solution to the problem.

© 1974 SCIENTIFIC AMERICAN, INC

THE CRY OF THE HUMAN INFANT

The sound spectrograph reveals the clear-cut differences between the crying of normal infants and the crying of abnormal ones. Such analysis may be useful in the early detection of infant disorders

by Peter F. Ostwald and Philip Peltzman

What is it that makes the cry of an infant so compelling? The sound stimulates strong feelings and distinct reactions from almost everyone within earshot. Undoubtedly much of the effect of the infant's cry is biologically determined in order to guarantee that the infant receives care and nutrition. Crying is one of the first ways in which the infant is able to communicate with the world at large. A mother quickly becomes aware of differences in the cry signals of her infant, although she is not always right in interpreting the meaning of the cry. In our research at the Langley Porter Neuropsychiatric Institute in San Francisco we have sought to develop more precise and objective methods of analyzing the sounds made by infants. With the aid of various acoustical techniques we have been able not only to identify the distress cry of normal infants but also to detect the presence of certain abnormalities and diseases by the characteristic cry of the affected infants.

One of the first objective reports of the acoustic structure of infant cries was published in 1832 by William Gardiner. In his book *The Music of Nature* he described the cries and calls of both humans and animals by means of musical signs. "Children," he wrote, "have no difficulty in expressing their wants, their pleasures, and pains, by their cries, long before they know the use or meaning of a word." His descriptions included "the spiteful voice of one child wantonly teasing another" and "the puling cry of a spoiled child." According to Gardiner, the tones of infant crying generally lie between the notes *A* and *E* in the middle of the piano keyboard. The initial expiratory component is usually the most prominent feature of the cry; it lasts about a second on the average and has an up-and-down melodic pattern. The inspiratory component of the cry is much shorter.

Charles Darwin was also interested in the crying of infants. In his book *The Expression of the Emotions in Man and Animals,* published in 1872, he presented a series of photographs showing the grimaces associated with tearful emotional expressions in children. When Thomas Edison invented the phonograph, it became possible to study a given cry repeatedly and determine its structure. In 1906 Theodor S. Flatau and Hermann Gutzmann of Germany recorded the crying of infants on Edison wax cylinders and produced a classic description of the acoustic properties of infant cries. They used both musical notation and the phonetic alphabet to characterize the cries' vocalic and consonantal features. Among the infants whose crying they recorded was one who had trouble breathing; this they immediately noticed because the pitch of crying was about one octave higher than that of normal infants.

For the next several decades the study of infant sounds was continued essentially by the method developed by Flatau and Gutzmann. The next major step did not come until the development of the sound spectrograph at the Bell Telephone Laboratories in the 1940's. This instrument dissects complex sounds into bands of frequencies and presents the bands in a horizontal record. In 1951 A. W. Lynip of Houghton College showed by means of sound spectrograms that there is a difference between the hunger cries and the attention-getting cries of infants. The sound spectrograph quickly became a standard piece of equipment for the study of infant sounds.

The examination of infant cries with sound-analyzing devices reveals that the fundamental frequency of a newborn infant's cry normally lies in a band between 400 and 600 hertz (cycles per second). Resonance at higher frequencies tends to peak between 1,200 and 4,800 hertz. With expiration cries caused by pain the fundamental tone may rise momentarily by several octaves or break abruptly into a wide-band noise. Of course, newborn infants not only produce distress vocalizations; they also hum and coo. Such sounds are initially heard only when the infant is relaxed or feeding at the breast, but it is likely that they play a significant role in later emotional expressions and in communication. Cooing is probably the first sound of agreement made by the infant. Its duration and quality are much closer to phonetic sounds than crying is, and many linguists believe that cooing is the true onset of speech.

The fact that infants produce typical cries and quickly learn to coordinate their behavior with that of the mother has led a number of investigators to search for normal and abnormal developmental patterns of crying, cooing and other vocalizations during the first year of life. Samuel Karelitz of the Downstate Medical Center of the State University of New York has made a phonograph recording that demonstrates how the cries of infants with certain brain disorders or genetic anomalies differ from the cries of normal infants. Henry M. Truby of the University of Miami used spectrograms to relate the sounds of the cry to movements of the infant's mouth and throat. The investigations of Ole Wasz-Höckert and his co-workers in Finland and of John Lind in Sweden have provided strong evidence that spectrographic analyses of infant cries have definite value for the diagnosis of a number of infectious, metabolic and developmental conditions.

Our sound-recording laboratory is lo-

© 1974 SCIENTIFIC AMERICAN, INC

cated in a hospital nursery so that we can study infants as soon as possible after they are born. The infant is placed in a recording chamber where the temperature is held between 80 and 85 degrees Fahrenheit. On the average 25 cries per infant are recorded. When we work with sick or premature infants who must be kept in an incubator, the entire incubator can be rolled into the recording chamber without disturbing the child. To record the infant's cries a sensitive microphone is suspended above him. The sounds are recorded on tape and sound spectrograms are made from the recording. The frequency and amplitude characteristics of the sounds can also be directly analyzed by means of a digital computer programmed for processing of such data.

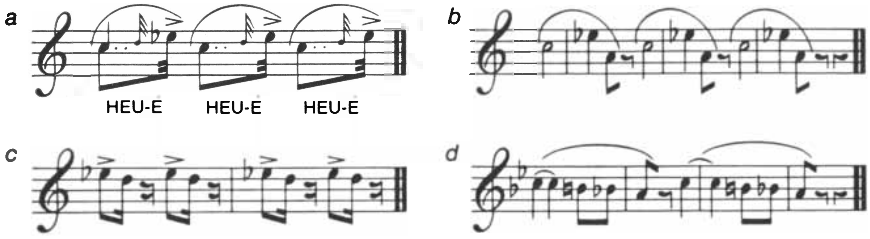
Our attention has been focused on the distress cry of infants. We have recorded numerous distress cries resulting from routine medical procedures such as the taking of blood samples and minor surgery such as circumcision. The distress cry is louder, longer and noisier than the hunger cry. It also tends to be irregular, with more interruptions and gagging.

There are times when unusual crying provides the first indication that something may be wrong with the infant. A nurse noted a peculiarly shrill cry from an infant who was considered to be completely healthy at birth. There was nothing unusual about the infant's weight, breathing, color or other behavior. On closer inspection, however, the nurse found that the infant had extremely wide fluctuations in body temperature. Examination by a neurologist led to the discovery that the infant had a grave congenital defect: he had no cerebral cortex.

In another instance we had great difficulty obtaining any cries from a two-day-old infant. The nurse told us that the infant rarely cried, which is most unusual in healthy infants unless they are under sedation. When the infant did cry, the sound was rather high-pitched. Although a pediatric examination at that time revealed no clinical evidence of disease, it became obvious during the next few weeks that the infant was not developing normally. A thorough physical examina-

CRIES OF INFANTS are recorded in a soundproof chamber. To keep the infant comfortable, air warmed to 80 degrees Fahrenheit is circulated through the chamber. The infants are usually placed on a soft mattress, although with sick or premature infants who must be kept in an incubator the entire incubator can be rolled into the recording chamber. A sensitive microphone is suspended from six to 12 centimeters above the infant's mouth, the height being adjusted according to the age and crying loudness of each infant. The sounds are first recorded on a high-quality tape recorder whose acoustical characteristics are well known and then analyzed.

© 1974 SCIENTIFIC AMERICAN, INC



NOTATION OF CRIES BY MUSICAL SIGNS was introduced in 1832 by William Gardiner in *The Music of Nature,* in which he described and analyzed the cries and calls of both human beings and animals. Some examples of Gardiner's notations are shown here: a child sobbing (*a*), a child crying (*b*, *c*) and "the puling cry of a spoiled child" (*d*).

tion disclosed a serious defect in the blood circulation between the infant's heart and lungs.

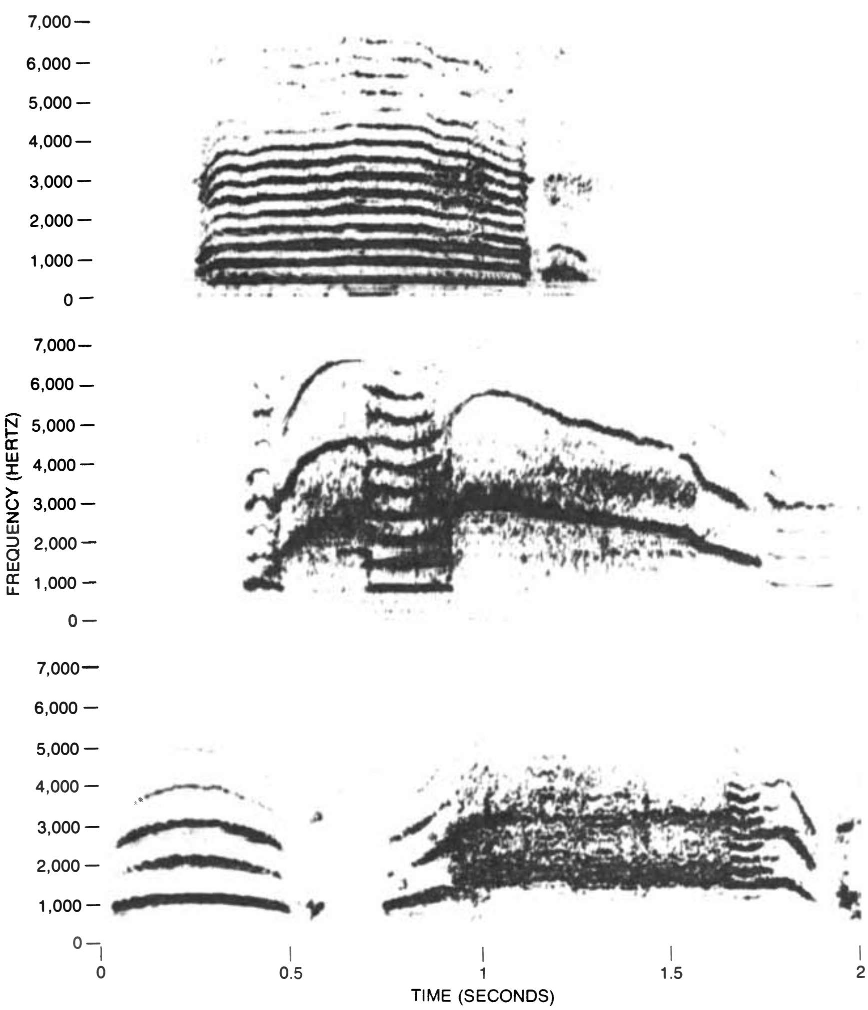
In order to determine if the duration of crying and the pitch of the cry were related to clinical diagnostic ratings of infants, we recorded the distress cries of 13 infants in the hospital nursery. We checked the clinical records of each infant, noting the details of prenatal history, birth and postnatal development. Five infants were normal; five were possibly abnormal, although no specific disease had been diagnosed; three were definitely abnormal. There was no significant difference in the duration of the cry of the infants in the three clinical groups. There was a fundamental difference, however, in the pitch patterns of the normal, possibly abnormal and abnormal infants. We found that the possibly abnormal and the abnormal infants had cries of higher pitch than the normal ones [*see illustration on opposite page*].

Another promising diagnostic technique for detecting abnormalities in infants, in addition to the spectrographic analysis of the cry, is the measurement of auditory evoked potentials, that is, changes in brain waves in response to sounds. Electrodes on the infant's scalp sense the small changes in electric potential in the brain that take place in response to sounds presented to the infant by means of a miniature earphone. The electrical changes are then analyzed by computer. Visual evoked potentials can also be studied by substituting a flashing light for the sound stimuli.

The usefulness of this electrophysiological technique in combination with spectrographic analysis has been demonstrated in an unusual clinical case. A pair of twin girls were born joined at the head. It was uncertain whether or not the two infants had independently functioning nervous systems. When we made our first recordings of the infants' cries at the age of three months, the cries of Infant *A* were louder and more sustained than those of Infant *B*. Infant *A* would also continue to cry for longer periods than her twin, whose crying was feeble and irregular.

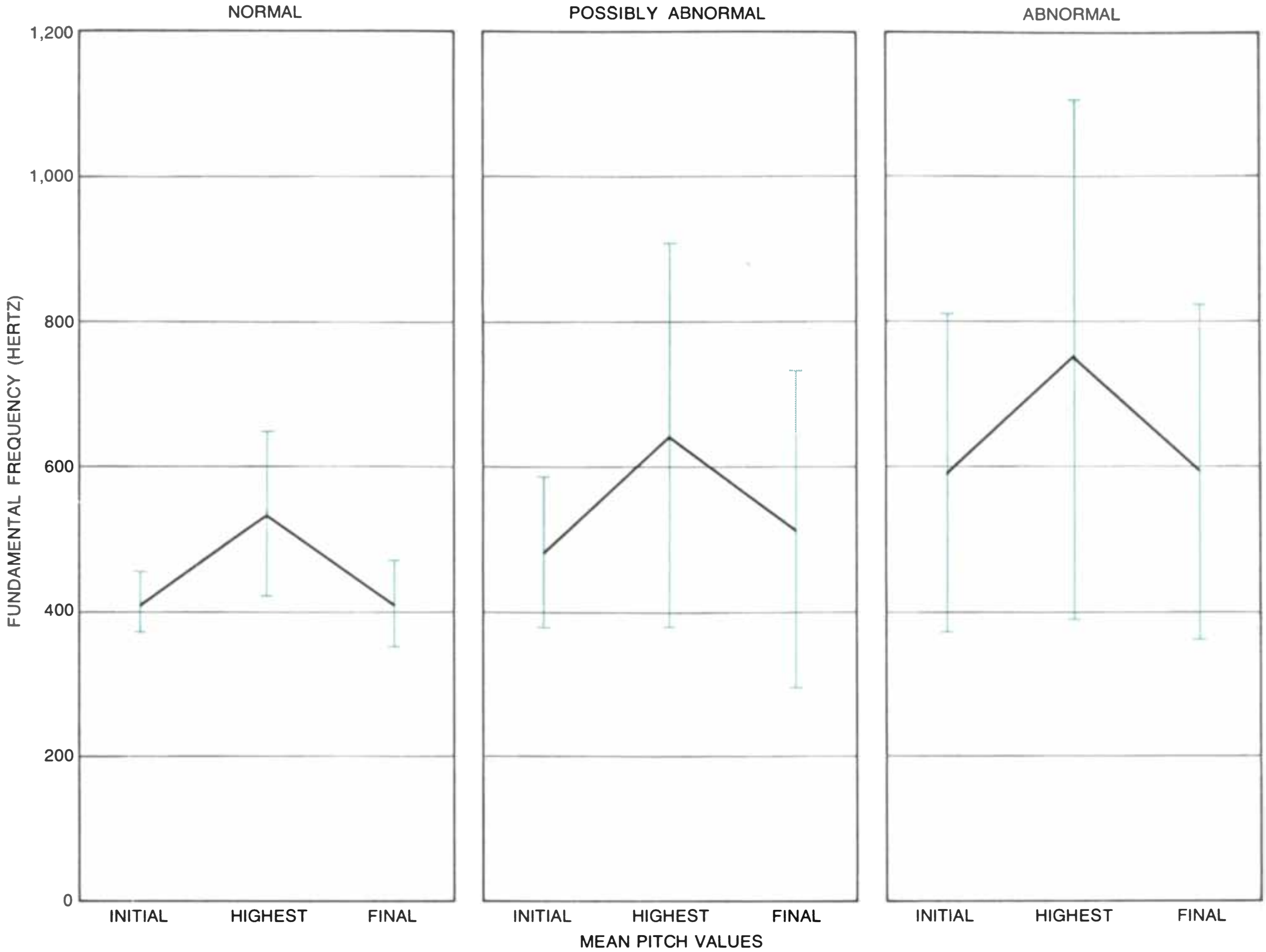
The sound spectrograms of Infant *A*'s cries looked exactly like what we have come to regard as being typical of a normal infant. The tone rose in pitch, remained relatively steady and then fell off smoothly, all within one exhalation lasting roughly one second. On the other hand, Infant *B*'s sound spectrograms were clearly abnormal [*see illustrations on page 88*]. There frequently were interruptions in the middle of an exhalation, so brief that they were inaudible unless one listened very closely. Moreover, Infant *B* produced more short and more long cries than Infant *A*. Comparison of the total sound intensity of the cries showed that Infant *B*'s cries had significantly less acoustic energy than Infant *A*'s.

The results of the acoustic analysis indicated that Infant *A*'s nervous system was apparently functioning normally but that Infant *B*'s nervous system might be impaired in some way. We know that vocalization is partly regulated by how



SOUND SPECTROGRAMS OF INFANT DISTRESS CRIES show the difference in the acoustical structure of the cries of normal and abnormal infants. The distress cry of a normal infant has a characteristic prolonged expiratory component followed by a brief inspiratory gasp (*top spectrogram*). The distress cry of an infant diagnosed as possibly abnormal was more irregular and higher pitched (*middle*). This infant later showed generalized delay in motor and mental development. An infant diagnosed as abnormal gave high-pitched, irregular cries, some without turbulence and some with turbulence (*bottom*).

© 1974 SCIENTIFIC AMERICAN, INC



PITCH PATTERNS of cries of infants from three clinical groups (normal, possibly abnormal and abnormal) show marked differences. A total of 356 distress cries from 13 infants were analyzed. The mean initial pitch, the mean highest pitch and the mean final pitch were determined for each clinical group. In addition the range around each mean is given by the colored vertical lines. These show that the cries of abnormal infants reach a height in pitch that is out of the range of normal infants. The findings suggest the possibility of developing an early screening test for detecting certain kinds of abnormality in infants by analyzing their cries.

we hear ourselves, so that it was possible Infant *B* was deaf. Each of the infants was therefore given a series of test sounds while her brain waves were being recorded. When the sounds were presented only to Infant *A*, we obtained evoked potentials from her side of the joined head and no response from Infant *B*'s side. When we presented the sounds to Infant *B*, we obtained no evoked responses from either side of the joined head. This lack of response could have been due to an impaired auditory capacity in Infant *B*, but in the light of other evidence we felt that it could be attributed to the infant's highly variable and fluctuating background brain-wave readings, which probably masked the more subtle changes that occur in response to acoustic stimuli. When we used flashes of light to test the visual evoked responses of the twins, we found that the brain activity of each infant was unmistakably independent.

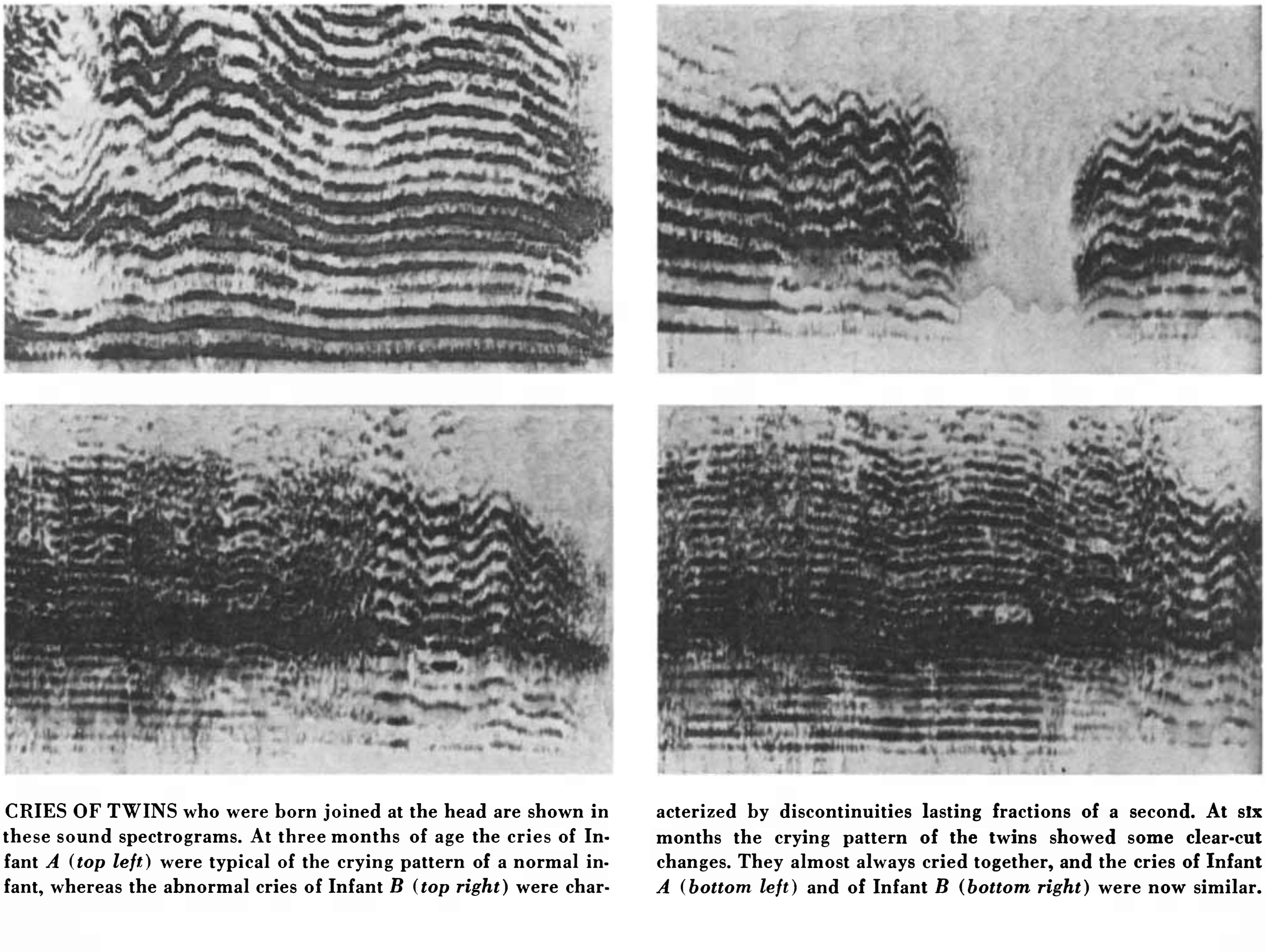
When the twins were six months old, we tested them once again and found that there had been an unexpected change. The cries of both infants now appeared to be exactly alike. Whenever Infant *A* cried, Infant *B* would soon join in, and the twins would vocalize with remarkable homophony. It proved to be extremely difficult to obtain recordings of individual cries. When we did obtain sound spectrograms of individual cries, no significant differences could be discerned. The abnormal, discontinuous cries of Infant *B* had disappeared. Soon afterward an attempt was made to separate the twins surgically. Both died during the operation.

How could infants with formerly distinctive "vocal personalities" develop virtually identical cries in only three months? The explanation, we suggest, is that the twins influenced each other's crying behavior. Since Infant *A* had a strong, normal cry, it is likely that Infant *B* learned or at least in some way acquired her normal cry from her conjoined twin. From observations of other twins it is known that under certain circumstances twins can influence each other so strongly as to develop a kind of secret language that only they understand.

Crying is one of the first social acts of an infant. It marks a shift on the infant's part from being silently dependent in the mother to being able to communicate with the world at large. Although cases of fetal vocalization *in utero* have been reported (we ourselves have never detected it), one usually considers the respiratory actions of an infant following birth to be the onset of human sound production. James F. Bosma of the National Institute of Dental Research has shown that birth sounds are usually combinations of gasping, coughing and efforts to breathe as the newborn infant struggles to adapt to his new environment. Other

© 1974 SCIENTIFIC AMERICAN, INC

TWIN GIRLS BORN JOINED AT THE HEAD, each with a cleft lip and a cleft palate, were tested by the authors to determine if each infant had an independent nervous system. Analysis of crying patterns suggested that the nervous system of one infant probably was impaired. Additional auditory and visual evoked-response studies indicated that the twins had completely separate nervous systems. Infants died when an effort was made to separate them surgically. Independence of the two brains was confirmed at autopsy.



CRIES OF TWINS who were born joined at the head are shown in these sound spectrograms. At three months of age the cries of Infant *A* (*top left*) were typical of the crying pattern of a normal infant, whereas the abnormal cries of Infant *B* (*top right*) were characterized by discontinuities lasting fractions of a second. At six months the crying pattern of the twins showed some clear-cut changes. They almost always cried together, and the cries of Infant *A* (*bottom left*) and of Infant *B* (*bottom right*) were now similar.

© 1974 SCIENTIFIC AMERICAN, INC

investigators have shown that the early vocal behavior of newborn infants is affected by the kind of anesthetic drugs given the mother during labor and by how quickly the umbilical cord is clamped after delivery.

Careful study of the infant's early auditory and vocal behavior, we believe, may reveal precisely how communication between the infant and the people in his environment develops. The problem is to find aspects of infant behavior that can be measured objectively and then to relate these measurements to developmental processes in the infant and to events in the environment. In some instances attempts to find such relations have been successful. In several diseases caused by chromosomal aberrations the affected infants have characteristic cries. In one disease, which results in a small, weak larynx, the infant has a high-pitched mewling cry so characteristic that the disease is called the *cri du chat* (cat cry) syndrome.

Recently our research efforts have turned to the study of brain processes associated with the infant's auditory performance. In one study we attempted to find out if a fetus would respond to sounds during the last trimester of pregnancy. Many mothers have observed that there are changes in fetal behavior as a result of a loud noise. A number of psychologists have attempted to condition the fetus to respond to sounds, but it is not clear whether the observed responses are the result of fetal conditioning or of the mother's behavior.

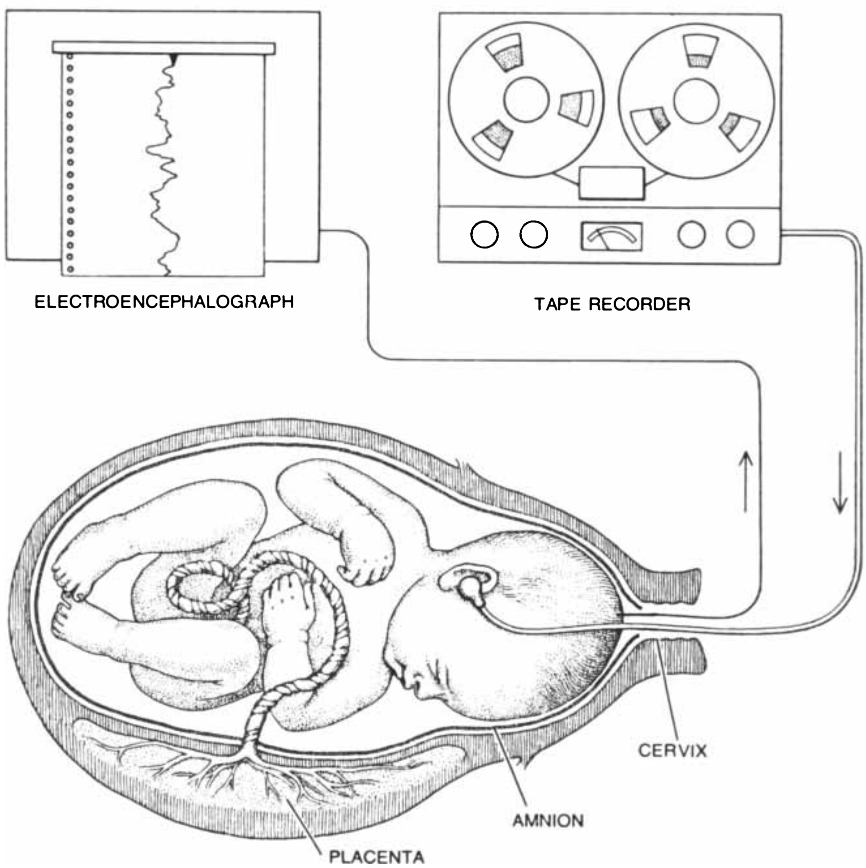
It is likely that a fetus does have some kind of auditory capacity before it is born because of the advanced development of its hearing system. By the fifth month of gestation the middle-ear and inner-ear structures have attained full adult size. The ability of this prenatal hearing system to process sounds, however, has yet to be convincingly demonstrated.

In our study six pregnant women volunteered to have their infants monitored during the first stages of labor. A miniature earphone was introduced into the uterine cavity through the partially dilated cervix and directed alongside the fetal head to a region near the ear. An electrode was then gently clipped to the fetal scalp and the electrode lead wires were connected to the brain-wave recording apparatus [*see illustration at right*].

The sound stimuli delivered to the fetus consisted of bursts of a 1,000-hertz tone at a moderately high level. We found that only one fetus showed a consistent evoked response to the sound. It is worth noting that the mother of the responding fetus was the only one who had not been given a drug to stimulate uterine contractions.

Electronic monitoring of the fetus during labor has been pioneered by Edward H. G. Hon of the University of Southern California School of Medicine. He has demonstrated that records of the fetal heart rate and of intrauterine pressure provide clinically valuable information about the condition of the fetus. In our own work we not only have examined fetal sound responses but also have monitored fetal heart rate, fetal brain waves and uterine-contraction patterns; the monitoring has been done by Phillip J. Goldstein of the San Francisco General Hospital. A computer receives the various signals and computes the integrated amplitude of the fetal electroencephalogram, the mean fetal heart rate and the mean intrauterine pressure for every five-second period. The values are printed out by a teletypewriter. The computer also produces a graphic display showing how the fetal encephalogram, the fetal heart rate and the intrauterine pressure are related. These methods provide the obstetrician with precise and relatively objective numerical and two-dimensional displays of fetal cerebral and cardiovascular function during labor. Conditions in labor that might affect the infant adversely and perhaps cause brain damage are thus more readily apparent.

A recent development in our work is a laser-beam optical system that produces a graphic display on a television screen of the fetal brain waves minutes after they have been recorded. With this system we have been able to show that mepivacaine, a drug frequently given during labor to block the mother's pain, depresses the amplitude of the brain waves of the fetus. In addition we have developed a method of transmitting physiological data from a fetus from one hospital to another over standard telephone lines. This has been demonstrated in a link between Mercy Hospital in Red-

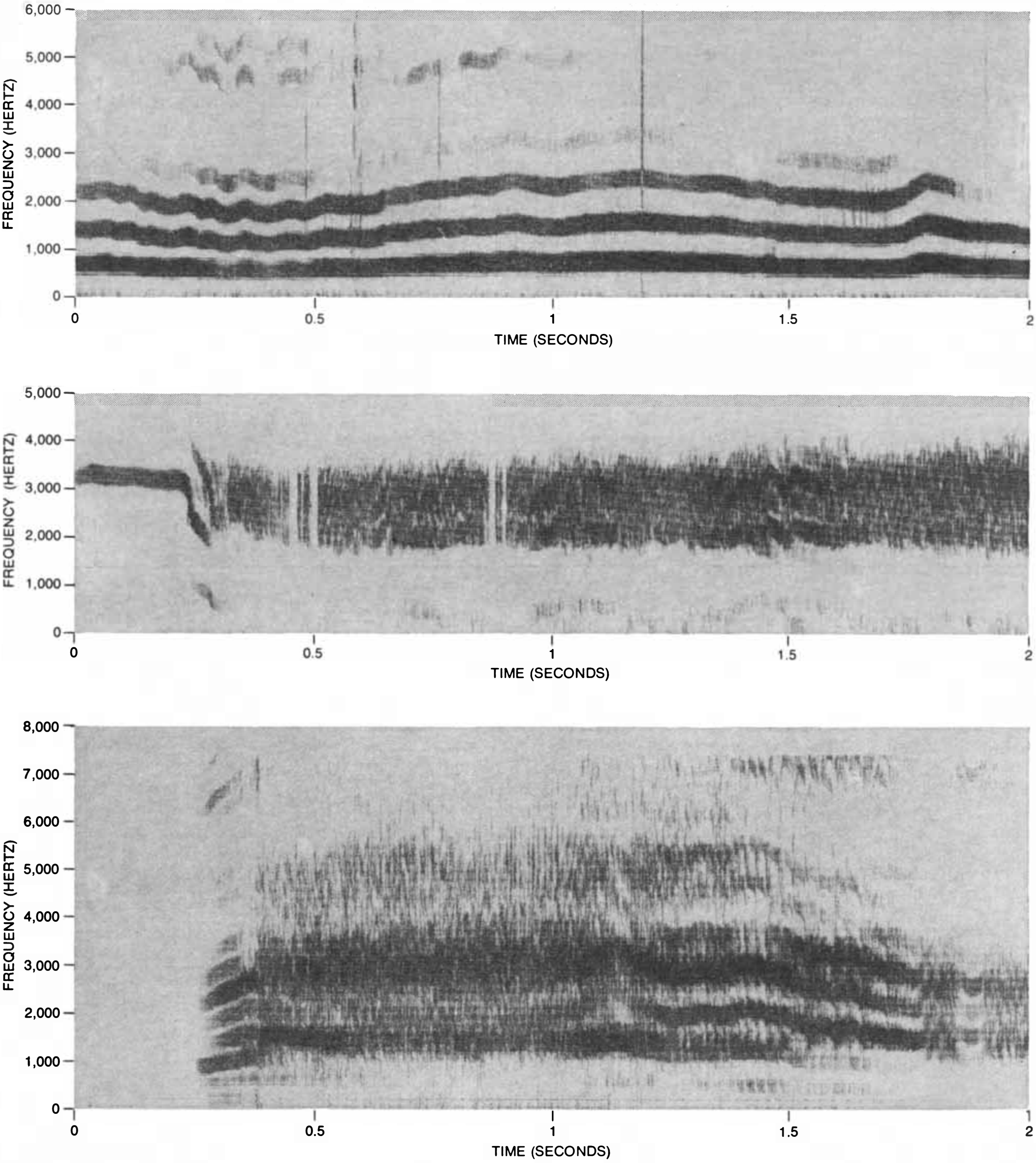


APPARATUS FOR MONITORING FETAL RESPONSES during labor is shown here in schematic form. A flexible stainless-steel electrode is introduced into the uterine cavity through the dilated cervix after labor has begun. The electrode is attached to the scalp and picks up changes in brain-wave activity. A miniature earphone also is introduced through the cervix and positioned near the fetal ear. In a study with six volunteer pregnant women only one fetus had a consistent evoked response to sound delivered through the earphone.

© 1974 SCIENTIFIC AMERICAN, INC

ding, Calif., and the San Francisco General Hospital, which are about 200 miles apart. The system makes possible instant consultation and action in cases of fetal distress.

In some cases neurological signs of brain damage may not be recognizable a week or two after birth, making it difficult to identify the infants who will not develop normally. If infants are suspected of having some generalized disorder on the basis of electronic monitoring during labor, perhaps some kind of routine screening would serve to detect more specific signs of brain damage. As we have seen, the spectrographic analysis of the infant's cry and the measurement of evoked auditory and visual potentials show considerable promise of providing a quantitative basis for such screening.



ABNORMAL CRYING of infants suffering from various severe disorders is shown in these sound spectrograms. A chromosomal disorder known as the *cri du chat* (cat cry) syndrome results in a characteristic mewling cry (*top*). A three-day-old male infant with hypoglycemia had an unusual cry with almost no low-frequency components (*middle*). Sounds of an infant who cried almost steadily were high in pitch and had a strangulated quality (*bottom*). The infant died of undetermined causes at the age of eight months.

© 1974 SCIENTIFIC AMERICAN, INC

Can technology solve the problems caused by technology?

Not long ago the prime goal was to provide more food, better living conditions, a longer life. For much of the world technology has successfully met these problems. But some people, looking at the way advances have disrupted the environment, say industrialization is a curse.

Chiyoda, as Asia's largest engineering firm, feels that it is through better technology that man will solve the problems caused by industrial progress. And it is to such better technology that we at Chiyoda are devoting our efforts.

Here are a few of the things we've done and are doing.

Desulfurization of Fuels.

To meet widely varying environmental and other requirements, Chiyoda has built desulfurization plants based on all the best known processes. We've engineered and constructed 15 of the 29 fuel oil desulfurization plants in Japan. Now we're working on the world's largest 50,000 B/D HDS plant for Idemitsu Kosan Co., Ltd. It will reduce the sulfur content of heavy fuel oil to 0.3%. Processing plants of this type help both to minimize air pollution and to alleviate the worldwide shortage of low-sulfur fuels.

Flue Gas Desulfurization.

The first major result of our efforts to develop new environmental control technology was the Chiyoda THOROUGHBRED 101 Flue Gas Desulfurization Process, which removes both particulates and sulfur dioxide to help keep the air free of pollutants from power plants, steelworks, refineries, and other industrial complexes. The Chiyoda THOROUGHBRED 101 has already been installed in four commercial plants. Six more installations will be completed by July 1974, among them a large capacity 750,000 Nm^3/hr. (467,000 scf/m) system for the Hokuriku Electric Power Company.

Removal of Fine Particles.

In addition to conventional dust removal systems, Chiyoda has recently introduced a new indoor fume controller that can be installed directly on the roof without the need for any conveying duct. It offers high efficiency in separating fine particles ranging in size down to 0.5 microns. The unit is economical to install and operate, and easy to maintain.

Water Pollution Control.

Another recent Chiyoda development was a water clarification plant with a flow rate several times faster than conventional systems. The whole plant is integrated into one single compact unit, resulting in a low construction cost. Applications include the pretreatment of boiler feed water, treatment of waste from petrochemical plants, removal of oils in waste water from petroleum refineries, and recovery of raw materials.

We have also recently perfected the Chiyoda THOROUGHBRED 242 Continuous Activated Carbon Waste Water Treatment Process, which removes organic substances from waste water. It is expected to play an important role in the water treatment systems of the oil refining, petrochemical, chemical and steel industries.

We don't claim Chiyoda has all the answers. But we do feel our 1,800 engineers and scientists and those at other technology-minded companies around the world provide hope. With community support, we can make technology solve the problems caused by technology.

RESPONSIBLE, GLOBAL ENGINEERING

CHIYODA CHEMICAL ENGINEERING & CONSTRUCTION CO., LTD.

Yokohama, Japan. Telex: CHIYO J47726

© 1974 SCIENTIFIC AMERICAN, INC

Ferdinand Braun and the Cathode Ray Tube

The "indicator tube" developed by Braun in the 1890's measured electrical quantities by the deflection of an electron beam. It was an ancestor of the oscilloscope and television picture tube

by George Shiers

Two signal events in the history of physics were the outcome of experiments with cathode ray tubes: the discovery of X rays in 1895 by Wilhelm Konrad Röntgen and the discovery of the electron in 1897 by J. J. Thomson. In the same period a German experimenter, Karl Ferdinand Braun, contrived a refined cathode ray tube capable of measuring certain electrical quantities. Braun called his invention a cathode ray indicator tube; in it the cathode rays were confined to a narrow beam and directed in such a way that they produced a luminous pattern on a fluorescent screen at one end of the tube.

Braun intended his device for the visual study of alternating electric currents of relatively low frequency. It has since become an instrument of outstanding usefulness in many fields of science and engineering. The lineal descendant of the indicator tube is the laboratory oscilloscope, but many other devices now commonplace in electronics technology also derive from Braun's invention. Among them are the radar display screen and the television picture tube.

The effects of an electrical discharge in a rarefied gas were well known in the 1890's. Preceding Braun's experiments were almost 200 years of enchantment and mystification with the changing patterns of color and light observed in exhausted globes and tubes.

In 1705 Francis Hauksbee produced a "fine purple light" inside an exhausted globe of glass by rotating and rubbing the globe. From Hauksbee's whirling sphere were derived the powerful friction machines that became the standard generators of electricity throughout the 18th century. Experiments with these machines in the 1740's created interest in many electrical phenomena, including demonstrations of luminous displays inside partially evacuated flasks and globes.

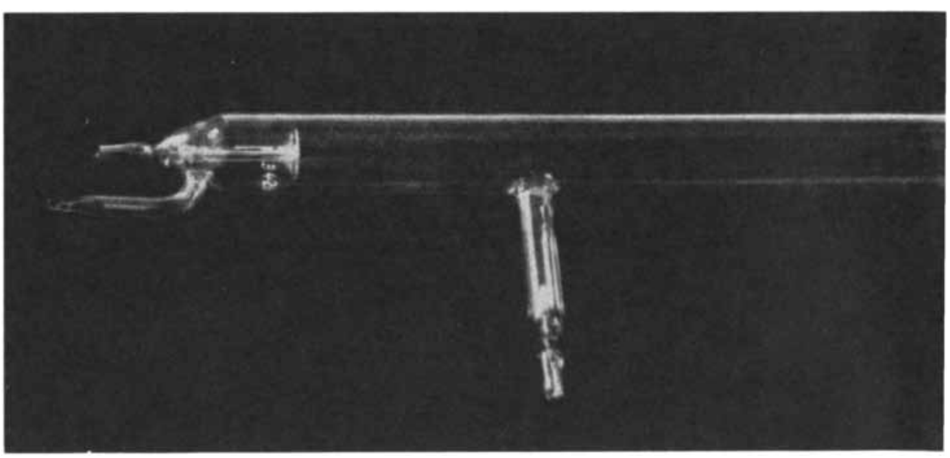
In an attempt to discover how well a vacuum conducts electricity William Watson in 1751 constructed an evacuated glass tube containing two metal plates, or electrodes. He reported that the discharge observed during his experiments was "a most delightful spectacle... of a bright silver hue." In another experiment with a curved glass tube he observed an "arch of lambent flame."

In succeeding years it was demonstrated that the color of the glow depends on the kind of gas in the tube and on its pressure. It was found that with a suitable degree of vacuum the continuous body of luminescence breaks into striations, the bright regions separated by darker ones. A bright light around the negative electrode, or cathode, was also noted [*see illustration on page 94*].

Experiments at the lowest pressure then obtainable indicated that a vacuum is a nonconductor. When air was admitted to the tube, the invisible discharge became a "beautiful green," which, as more air entered, changed to blue and then to indigo and violet. These were important observations, but the more substantial discoveries that led to applications and to a theoretical understanding of these phenomena could not be made until better vacuum pumps and improved glass tubes became available a century later.

In 1855 Heinrich Geissler, a German instrument maker and glassblower, invented a pump in which a column of mercury acted as a piston, producing a better vacuum than had previously been possible. Two years later he fabricated closed glass tubes of high vacuum with wires leading from enclosed electrodes sealed permanently in the glass envelope.

Geissler worked with Julius Plücker, a physicist and mathematician, at the University of Bonn. Plücker quickly applied the new sealed tubes to the investigation of electrical discharges in gases at low pressure and made the important discovery that a magnet influences the glow. When a magnet was placed near



BRAUN INDICATOR TUBE, made in 1897, is a slightly modified version of the instrument employed by the inventor in his initial experiments. The cathode is at the extreme left; the anode is in the small tube protruding from the neck. The viewing screen, a sheet of trans-

© 1974 SCIENTIFIC AMERICAN, INC

the negative electrode, the light spread out in a pattern similar to that assumed by iron filings in a magnetic field. To a large extent it was the joint work of Geissler and Plücker that led, before the end of the century, to the discovery of X rays and the discovery of the electron and to Braun's beam-deflection instrument.

Related experiments were made at the same time in England by John Peter Gassiot, often in collaboration with Michael Faraday, in some cases with tubes made by Geissler. Taken together, these investigations showed that the cathode was the source of certain emanations, although their character was not yet understood. These "negative rays" behaved like a flexible conductor in the way they reacted to a magnetic field; they caused a rarefied gas to glow, and they produced a blue or green light where they struck the glass of the enclosing tube.

Plücker's research was continued by his colleague Johann Wilhelm Hittorf, who described other characteristics of the discharge in 1869. With the better vacuum then available he observed further changes in the discharge as lower pressures were achieved, as well as characteristics of the glow in the vicinity of the cathode. Hittorf also discovered that an object placed in front of the cathode cast a shadow in the fluorescent glow on the opposite wall of the tube. This was a fundamental observation, since it indicated that the "glow rays," as Hittorf called them, were propagated in straight lines from the cathode.

In the 1870's Eugen Goldstein began a series of investigations with discharge tubes that he continued for many years. He showed that the impact of cathode rays on coated plates mounted inside the tube produced chemical changes. He studied cathode materials; more important, he demonstrated that the rays were emitted perpendicularly to the surface of the cathode and that they could be made to converge (that is, they could be focused) by employing a concave cathode. In 1876 Goldstein introduced the term *Kathodenstrahlen,* or cathode rays.

Interest in electrical discharges in vacuum tubes reached a peak in 1879, when William Crookes gave a series of novel demonstrations. Using more than a dozen kinds of tubes, including a "shadow tube" that became famous, he illustrated all the features of the discharge previously known, and some new ones, such as the ability of the rays to do mechanical work and to produce heating.

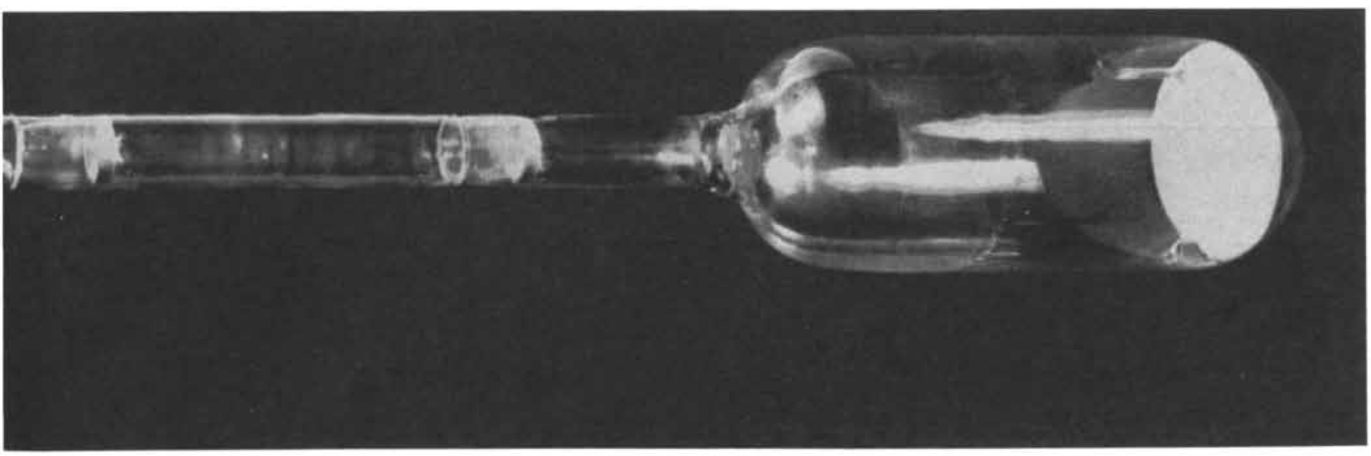
In some of his tubes Crookes reduced the pressure to a millionth of an atmosphere, so that the discharge was no longer visible. He showed that matter in this rarefied state behaves in a special way. The emission from the cathode was, he believed, a "torrent of molecules"; he called it "molecular light." He announced a theory of a fourth, ultragaseous state of matter, which he termed "radiant matter." Crookes's belief that the radiant matter consisted of negatively charged particles streaming from the cathode exacerbated a conflict between the proponents of two opposing views: those who believed the emissions were particles and those who thought they were waves.

During the 1880's the range of investigations into gas discharges was extended by a series of ingenious experiments. One surprising discovery, made by Heinrich Hertz, was that cathode rays could penetrate thin materials, such as gold leaf. In 1894 Philipp E. A. Lenard confirmed the result of Hertz's experiment by constructing a tube with a window of thin aluminum foil at the end opposite the cathode. He found that the rays passed through this window, and he was able to detect them several centimeters away in air. The following year Jean B. Perrin, experimenting with a tube containing open-ended concentric cylinders, proved that the rays carried negative charges of discrete magnitude. The study of electrical phenomena in discharge tubes had become a frontier of physical science.

By the early 1890's many kinds of vacuum tubes were to be found in laboratories and lecture rooms. Called Geissler tubes, Plücker tubes, Hittorf tubes or Crookes tubes, they were used mainly to demonstrate the novel visual effects of the discharge.

The experiments of a few scientists, however, had objectives more serious than the creation of colorful displays. In 1895 Röntgen was working with a Crookes tube enclosed in black cardboard. He noticed that a nearby screen of fluorescent crystals glowed when the tube was energized. In the next few weeks he studied the invisible radiations from the face of the tube and found that they passed through some opaque substances and could expose photographic emulsions. Late in December Röntgen announced his discovery of the new kind of rays, which he called X rays.

Braun, then director of the Physical Institute at Strassburg (now Strasbourg), turned his attention to X rays soon after he received a copy of Röntgen's paper. Born at Fulda in Germany in June, 1850, Braun had studied at the University of Marburg and had become professor of physics there in 1876. He



lucent mica coated with a luminescent compound, is at right, at the end of the bulbous portion of the glass envelope. Two diaphragms, each with a small aperture in the center, restrict the breadth of the electron beam. (The first indicator tubes had only one diaphragm.) The deflection coil and other devices that were mounted externally are not shown in the photograph. The tube was about 20 inches long.

© 1974 SCIENTIFIC AMERICAN, INC

had spent the next 20 years in various posts at Strassburg, Karlsruhe and Tübingen, and had returned to Strassburg in 1895.

The cathode ray tube was of intense interest during the early months of 1896 as hundreds of workers repeated Röntgen's experiments. Braun was among those enthusiasts, but unlike most of them he was more concerned with the source of the X rays than with the radiation itself. What intrigued him were the emanations from the cathode and their power to excite fluorescence in the glass envelope of the tube.

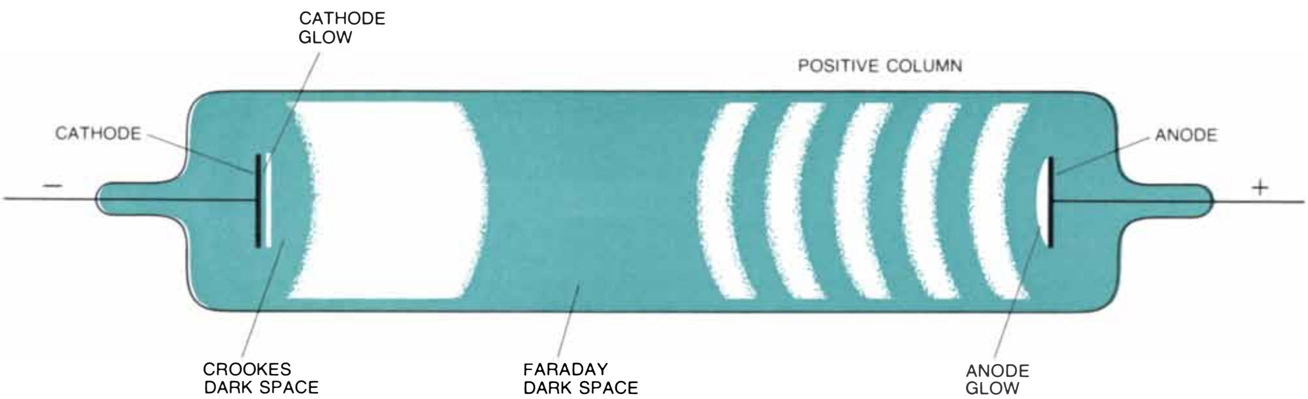
No one had yet controlled these rays in order to make use of the fluorescence, although in 1894 Albert Hess had proposed that they could be of value in magnetic studies. Braun saw in the discharge tube the possibility of a new instrument, a visual indicator of oscillatory and transient phenomena in electrical circuits. The need for such an instrument was already apparent.

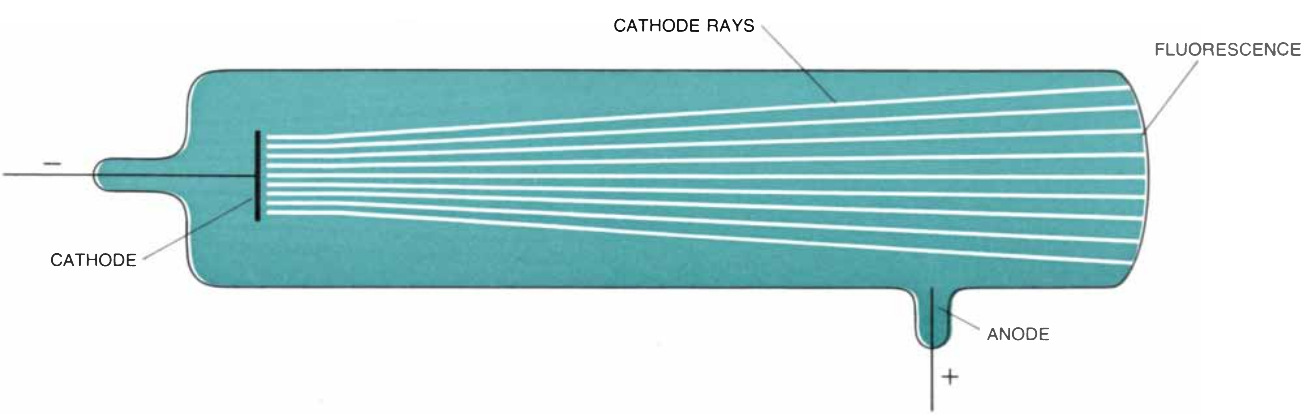
By the 1890's central generating stations were providing electric power to many cities and towns. Some newer systems, including the one at Strassburg, supplied alternating current. The periodic nature of this current, which reverses its polarity with every cycle, presented a number of problems of measurement.

The wave form of an alternating current is a curve that shows the instantaneous value of the voltage or current during one cycle. In power systems it is ideally a sine curve, that is, a curve proportional at every point to the sine of the angle through which the alternator shaft has turned. "Phase" is also of interest in power systems; it is the angular displacement between two or more periodically varying quantities, such as voltage and current or two voltages [*see illustration at top left on page 96*]. At that time precise knowledge of these swiftly changing quantities was difficult to obtain.

One laborious method for mapping wave forms consisted in noting a series of voltmeter readings and the corresponding angular positions of the alternator shaft. More refined techniques then being developed employed rotating or vibrating mirrors that displayed the wave form as a curve traced by a beam of light on a screen. In the electromechanical oscillograph, for example, a tiny mirror is attached to the moving coil of a galvanometer, which oscillates in sympathy with the applied voltage. A beam of light reflected from this mirror moves to and fro, and the line of light thereby produced is opened up transversely, or scanned, by a second mirror moving at right angles to the first.

At the frequencies of commercial power systems (usually 50 or 60 cycles per second) the delineation of wave forms was within the capabilities of such early oscillographic instruments. Currents of much higher frequency were of growing interest, however. Special types of alternators furnished electricity of 5,000 or





EFFECTS OF AN ELECTRIC DISCHARGE in an evacuated tube depend on the amount of gas remaining in the tube. At pressures of about a ten-thousandth of an atmosphere (*top*) electrons emitted by the cathode collide with and ionize residual gas molecules, producing a glow in various regions within the tube. With further rarefaction these collisions become infrequent, the bright areas recede toward the anode and the Crookes dark space expands to fill the length of the tube. At about a millionth of an atmosphere (*bottom*) no ionization glow is visible; it is under these conditions that cathode rays are produced. The electrons travel at high velocity, evoking fluorescence in the glass when they strike the end of the tube. The electrons are emitted perpendicularly to the surface of the cathode but diverge slightly as a result of mutual repulsion. Their path is not influenced by the shape or the position of the anode.

© 1974 SCIENTIFIC AMERICAN, INC

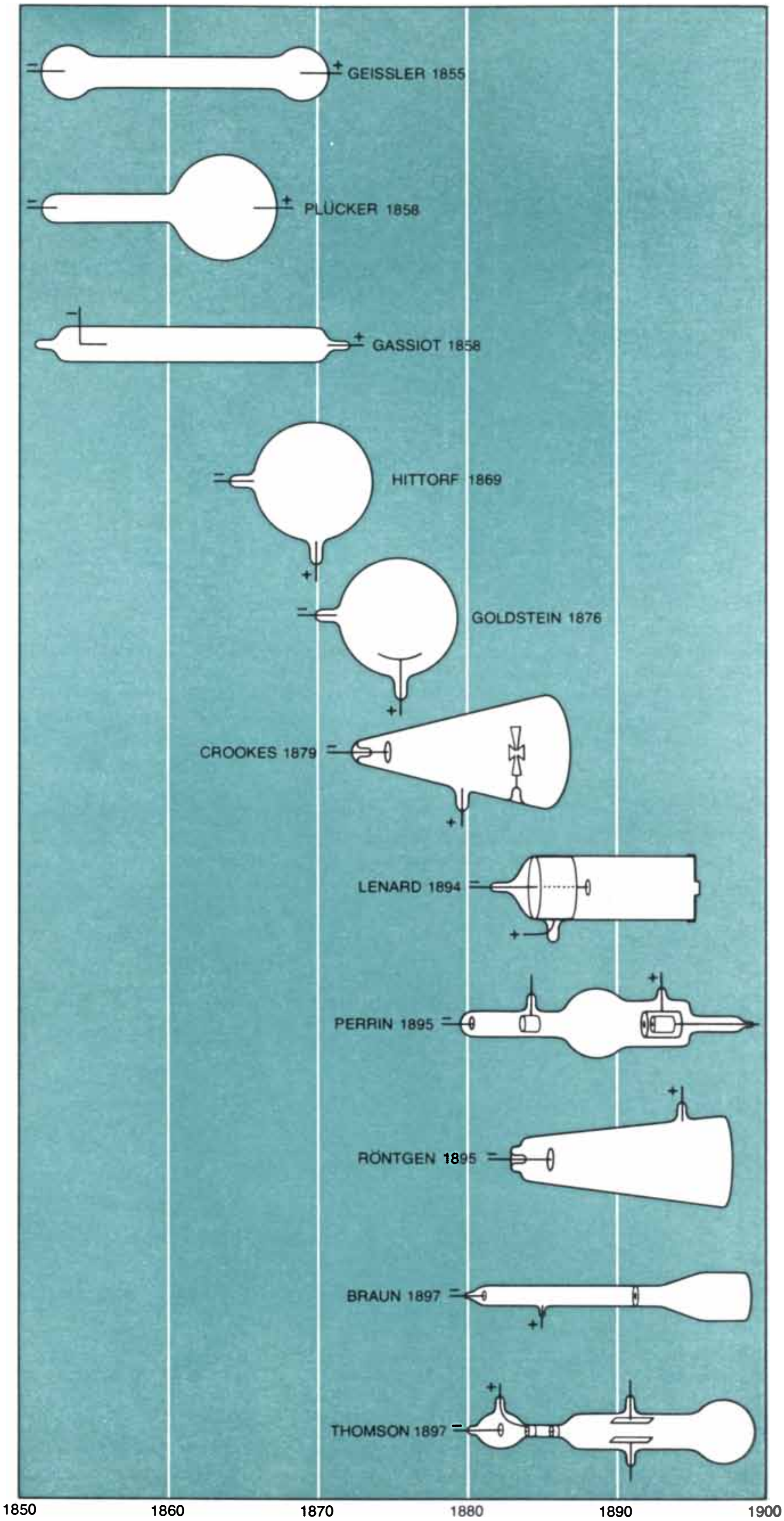
even 10,000 cycles per second. Still higher frequencies were generated by oscillatory circuits incorporating Leyden-jar capacitors and by discharges across spark gaps; these were already in use in the new art of wireless telegraphy. Because of the inertia of their moving parts and because of relatively low resonant frequencies electromechanical instruments could not record oscillations at these high frequencies.

Braun had a long-standing interest in oscillatory phenomena; his doctoral paper, written in 1872, was on the oscillations of strings and elastic rods. The new field of electromagnetic waves, introduced by the work of Hertz in 1888, and the high frequencies obtained from spark-gap generators, both of which attracted other notable investigators, were closely allied with Braun's favorite subject. It is not surprising that he developed an interest in them, nor is it surprising that the lack of a suitable instrument for the study of these phenomena attracted his attention. Braun had often designed original apparatus for his own investigations and for classroom use.

Braun readily perceived that certain features of the cathode ray tube were ideal for adaptation to a wave-form indicator. If the rays were constricted, their impact on a suitable fluorescent screen inside the tube would produce a spot of light instead of a broad luminosity across the entire face of the tube. Because the rays respond to a magnetic field by moving at right angles to the field, the moving spot could be made to oscillate in synchrony with a periodically varying field. The motion of the spot could therefore reflect changes in voltage or current and their relationship to time.

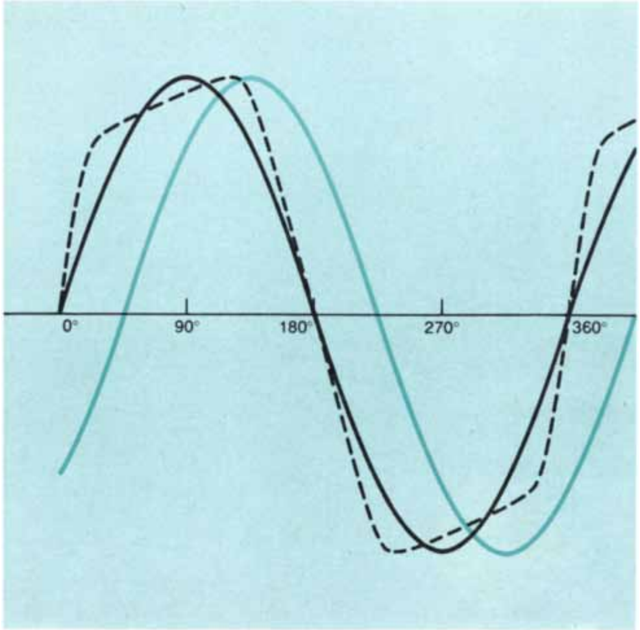
Because such a beam is virtually without inertia, it would respond almost instantaneously to changes in the deflecting force. The beam could also move in any direction across a plane surface and could trace intricate patterns. In these respects it is like light, but in its responsiveness to magnetism it offers an advantage over light.

If a device of this kind could be constructed, Braun perceived, it would raise the maximum frequency of graphic indicators by several orders of magnitude. Engineers and scientists would be brought into visual contact with the dynamics of oscillatory circuits. Even in crude form such an instrument would enable one to study short-lived and rapidly changing currents in greater detail and with greater precision than had pre-

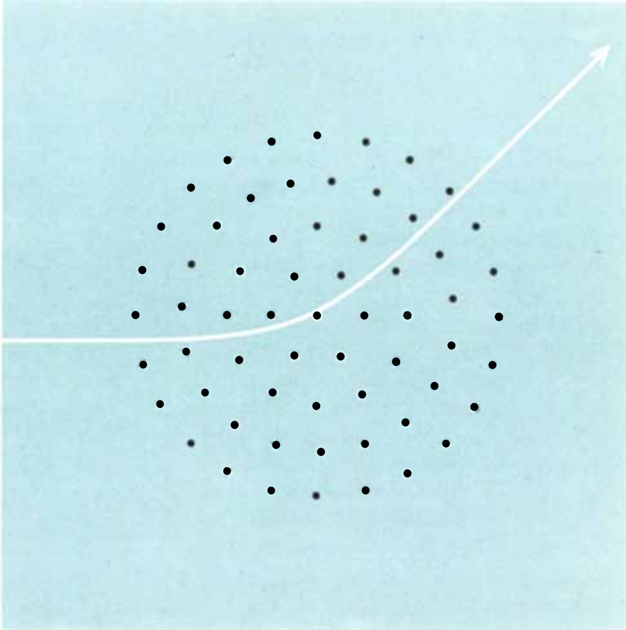


STUDY OF ELECTRICITY IN A VACUUM was a productive field of research in the 19th century. Vacuum tubes employed by some of the investigators are illustrated here. This work culminated in the discovery of X rays by Wilhelm Konrad Röntgen, the invention of the indicator tube by Karl Ferdinand Braun and discovery of the electron by J. J. Thomson.

© 1974 SCIENTIFIC AMERICAN, INC



SINUSOIDAL WAVE FORM is generated by an alternator when the instantaneous value of the voltage or current is proportional to the sine of the angle of rotation of the alternator shaft. The Braun indicator tube provided a means of observing such wave forms and of detecting distortion in them (*broken black curve*). It was also capable of measuring the phase relationship of two wave forms; here the solid black curve "leads" the colored curve by 45 degrees.



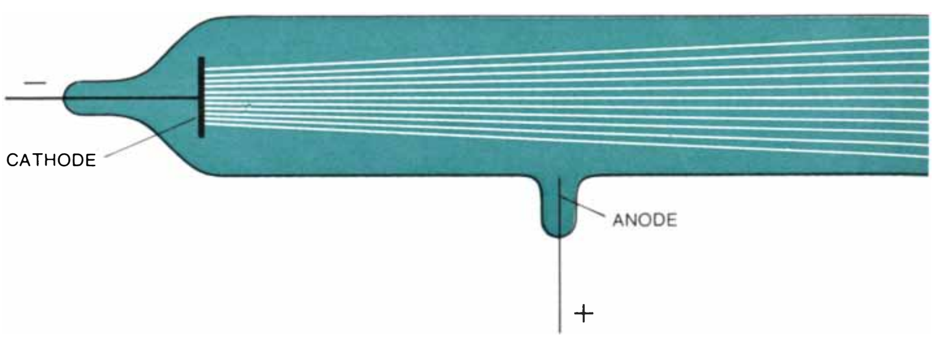
ELECTRON IN A MAGNETIC FIELD follows a path described by a segment of a circle. The field (*black dots*) is perpendicular to the page. While traversing it the electron (*white line*) is deflected onto a circular trajectory whose radius is determined by the strength of the field. The electron enters and leaves the field on straight lines that are tangent to the circle. It is neither attracted nor repelled by the magnet and remains in a single plane parallel to the page.

viously been possible. Such an instrument was, in fact, as essential to electrical science as the telescope, microscope and spectroscope were in other fields.

During the latter half of 1896 Braun designed a tube that would serve his purpose. The basic requirements of the design were already evident: the divergent rays from the cathode would have to be confined to a narrow beam and a fluorescent screen would have to be placed at the end of the tube as a viewing surface. The simplest way to obtain a thin beam was to block most of the rays with a diaphragm or baffle and pass only a small portion through a hole in the center. A tubular segment of the glass envelope would contain the cathode, the restricting diaphragm and an anode at the side; the other half of the tube would be expanded to accommodate the viewing screen. The screen would consist of a thin, translucent disk of mica, coated with a phosphor such as barium platinocyanide or zinc silicate. A coil outside the tube, placed close to the diaphragm and connected to the current being studied, would deflect the beam. Finally, as in other oscillographic methods, a rotating mirror in front of the phosphor screen would be required to translate the line of light produced by the moving spot into a two-dimensional curve.

Braun designed several tubes of varying geometry and had models made in Bonn by Franz Müller, a successor to Geissler. Braun's paper "A Method to Demonstrate and Study the Time Sequence of Variable Currents" was published in *Annalen der Physik und Chemie* in February, 1897. The paper includes a diagram of the tube he favored for experiments [*see illustration below*].

The tube was mounted horizontally in a cradle with the deflection coil near the



FIRST BEAM-DEFLECTION INSTRUMENT was constructed for Braun by Franz Müller in 1897. The deflection coil, mounted athwart the pierced diaphragm, bent the beam upward when energized in one direction and downward when the polarity was reversed. When an

© 1974 SCIENTIFIC AMERICAN, INC

diaphragm and perpendicular to the axis of the tube. The high voltage required to energize the tube was produced by a Töpler influence machine, a descendant of the hand-turned friction generators, or sometimes by an induction coil.

Each experiment required elaborate preparation by Braun and his assistants. When an unsealed tube was used, it had to be laboriously evacuated by a hand-operated mercury pump. Someone had to crank the influence machine or attend to the induction coil. The speed of the rotating scanning mirror had to be adjusted, and sundry connections in the electrical circuits had to be made as each experiment progressed. When all was in order and an alternating current was applied to the deflection coil, a glowing green line about an inch long appeared on the phosphor screen. In the rotating mirror this line was transformed into a faint and flickering curve.

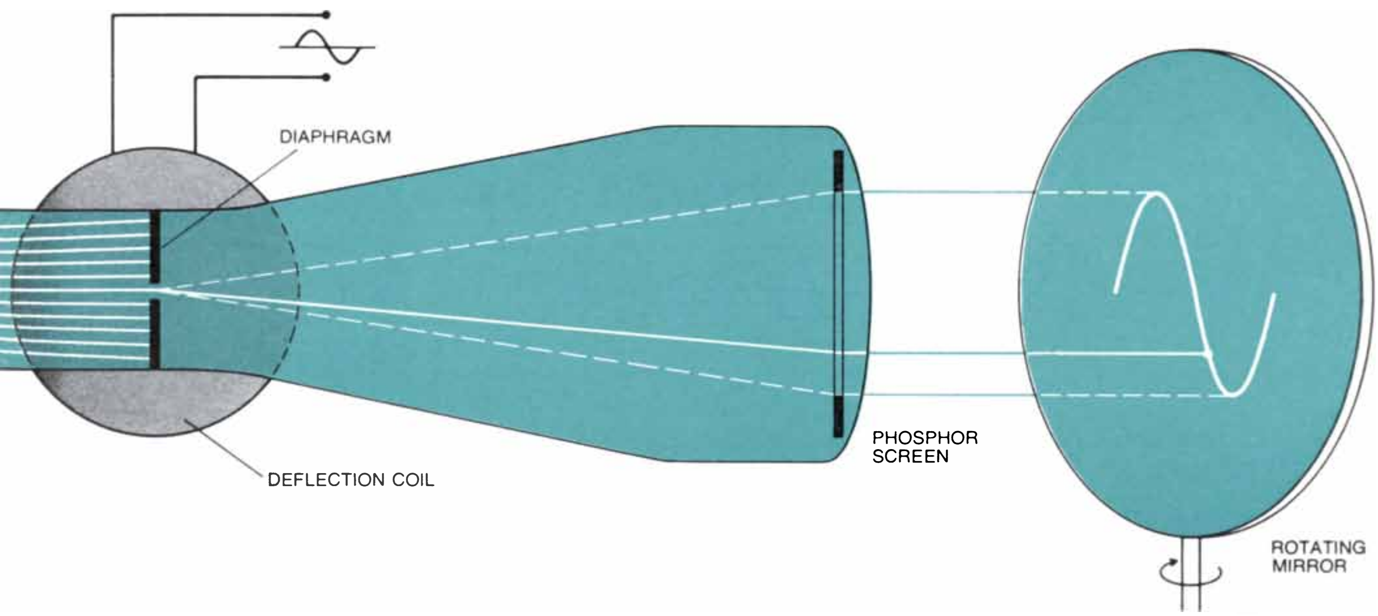
In his initial experiments Braun made use of the Strassburger Centrale power lines (120 volts, 50 cycles per second) to test his instrument. At that time the wave form of some alternators was a kinky curve distorted by the mechanical and magnetic peculiarities of the generating or distribution apparatus. Braun was surprised to discover that the local wave form was virtually sinusoidal. He compared the power-line wave form with the curve produced by a spot of light reflected from a mirror attached to one of the tines of a tuning fork, a standard method of producing a sine wave. In giving diagrams of both curves in his paper he may have been trying to show that his apparatus could perform as well as the better-known arrangement of tuning forks and mirrors.

Having verified the basic performance of the indicator tube, Braun investigated the more complex wave forms of currents in an induction coil. He compared the wave forms in the primary and secondary windings and observed the wave form of the secondary current under various conditions: when the circuit was open, closed or connected to a capacitor. While studying these patterns he noted that the brightness of the spot varied with its velocity across the phosphor screen. A slow trace produced a bright line, whereas a fast movement of the spot resulted in a line so faint that it was not easily seen even in a fully darkened room.

In further experiments Braun showed that the glowing spot could produce Lissajous figures directly on the phosphor screen without an interposed rotating mirror. These closed curves (named for Jules Antoine Lissajous) are created by the simultaneous displacement of the spot by two simple harmonic motions at right angles to each other. Braun obtained a second deflecting force by rotating a magnetized rod in a horizontal plane below the tube. With alternating current from the power lines energizing the deflection coil he observed changing patterns by varying the rate at which the magnet was spun. He readily obtained certain simple patterns, such as a figure eight, which represents a frequency ratio of two to one. He could not demonstrate the higher ratios, however, because at high rotational speeds it became difficult to synchronize the two frequencies.

In a subsequent series of experiments Braun was able to operate the tube as an all-electric instrument, without any moving mechanical parts. Again the scanning mirror was removed, but instead of a rotating magnet a second deflection coil, at right angles to the first, was installed. When the currents in the two coils were in phase (when their peaks and nulls were simultaneous), the resultant forces acting on the beam produced on the viewing screen an oblique line. A small difference in phase produced a narrow ellipse; as the difference increased, the ellipse broadened and, when the two currents were 90 degrees out of phase, became a circle. A phase difference of 180 degrees produced an oblique line perpendicular to the first line. By this method Braun was able to determine the phase displacement between the primary and the secondary winding of an induction coil.

These patterns are modified when the circuit energizing one coil includes more inductance or capacitance than that energizing the other. With one coil in a circuit containing only resistive elements (and the inductance of the deflection coil itself), Braun connected the other coil to an inductor with an air core, to an inductor with an iron core and to a

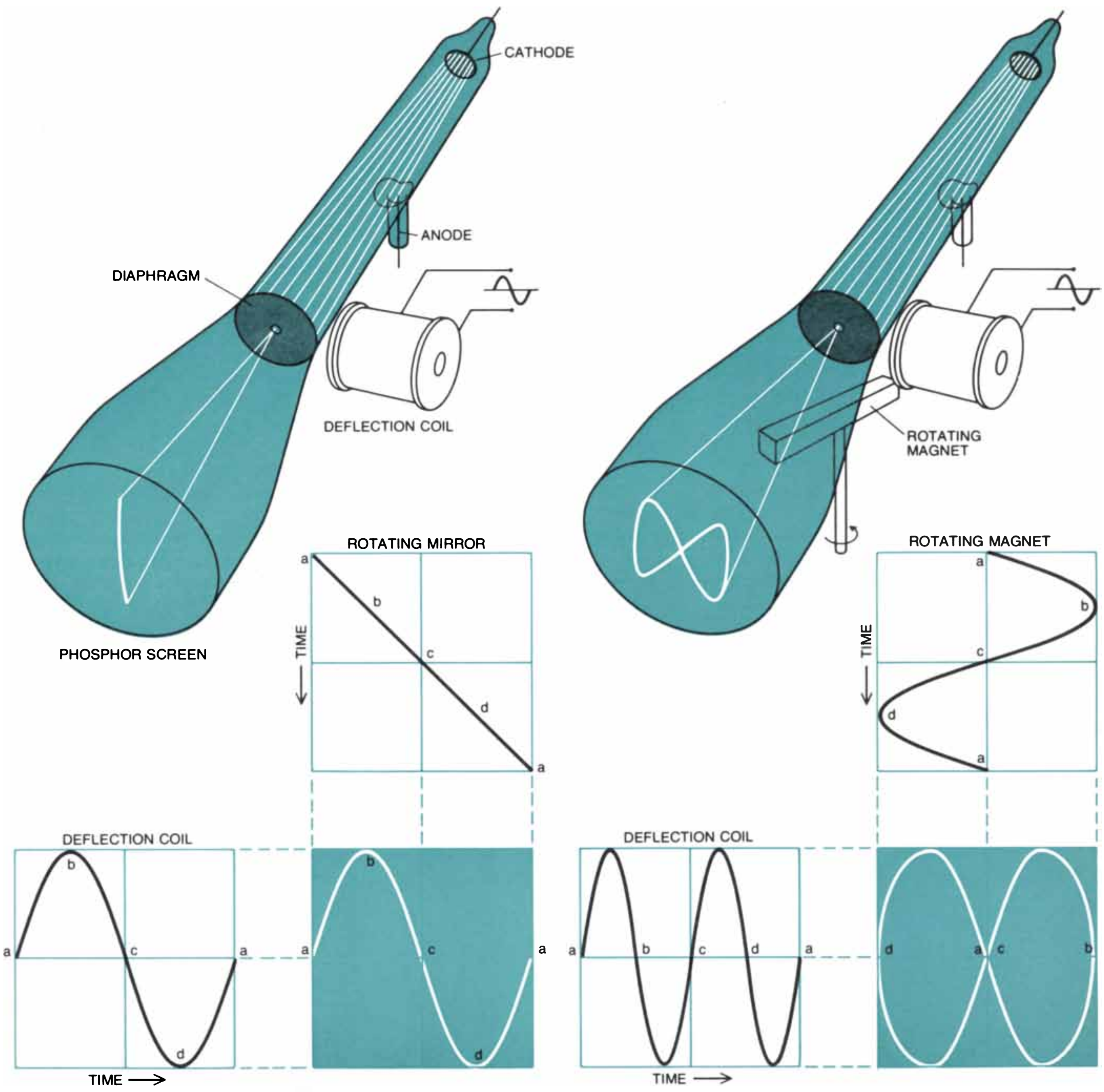


alternating current was applied to the coil, the resulting motion was too rapid to be detected by the eye and appeared on the phosphor screen as a vertical line about an inch long. In the rotating mirror, however, the line was transformed into an image of the original wave form; the rotation of the mirror added a horizontal component to the motion of the luminous spot, "scanning" it transversely.

© 1974 SCIENTIFIC AMERICAN, INC

capacitance, separately and in combination. The resulting figures graphically confirmed the fundamental theory of reactive electric circuits. When a capacitance was introduced, the patterns were distorted, with peaks and indentations. Braun recognized that these distortions were caused by the superimposition of oscillations of higher frequency generated by the combination of inductance and capacitance in a resonant circuit.

The versatility of the instrument was confirmed in additional experiments on phase displacements with polarized electrolytic cells and on the magnetism induced in long iron bars by direct current and alternating current in movable coils. The response of the tube to a voltage that changes extremely fast was also investigated. When a capacitor is discharged through an inductor, the initial

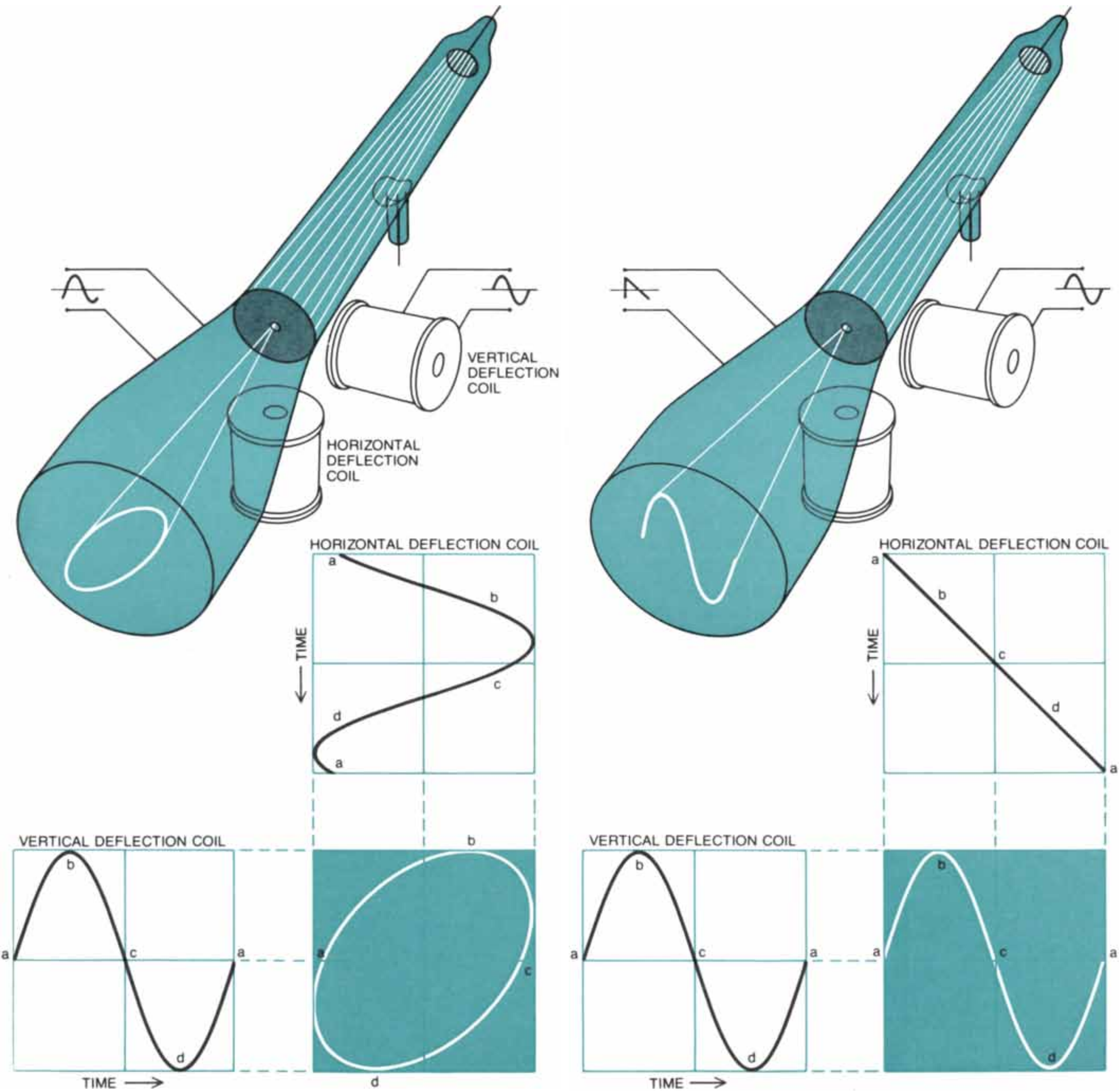


MODES OF OPERATION of the indicator tube are depicted in the drawings on these two pages. In his first experiments (*left*) Braun deflected the beam only in the vertical plane, relying on a rotating mirror, which is not shown, to scan the image horizontally. The resulting wave form can be analyzed into two components. The vertical component consists of the sinusoidal current applied to the deflection coil; the horizontal component is the motion of the mirror: the image moves across it at a uniform rate, by convention from left to right, returning to its starting position with each half-revolution. In the graph each point on the resulting wave form is determined by the corresponding points on the component curves. For his subsequent experiments Braun eliminated the rotating mirror. By spinning a magnet under the tube while maintaining an alternating current in the vertical deflection coil he was able to obtain Lissajous figures directly on the phosphor screen (*second from left*). These figures are produced whenever two waves interact per-

© 1974 SCIENTIFIC AMERICAN, INC

current is followed by a series of oscillations that is rapidly extinguished. Braun discharged a Leyden jar through the deflection coil and noted that the fluorescent spot moved about half an inch up and then about the same distance down. The rapidity of this motion proved, at least qualitatively (he did not time the discharge), an essential characteristic of cathode rays: their infinitesimal inertia.

In his paper Braun did not discuss the theory of the indicator tube, nor did he explain how he had arrived at its design; the process seems to have been largely intuitive. Following German custom he employed the term "cathode rays" (most English writers called them "negative rays") at a time when the nature of the rays was a matter of speculation and controversy. The issue was soon to be resolved by Thomson, who was then near-



pendicularly to each other; if the ratio of their frequencies is integral, a closed figure is formed. Here the magnet is rotating at one-half the frequency of the deflection coil signal; consequently the figure has two cusps. When a second deflection coil was added to the tube, Braun was able to study phase relationships (*third from left*). The currents are equal in frequency and amplitude but the signal applied to the vertical deflection coil leads the other by 45 degrees; the resulting trace is an ellipse. A further refinement, added by Jonathan Zenneck, one of Braun's assistants, made it possible to display the wave form of one of the deflection signals directly on the phosphor screen, without a rotating mirror (*right*). Zenneck's "time base" scanning system generated a signal that swept the beam of electrons across the screen horizontally at a uniform rate, returning it to the extreme left at the end of every cycle. As the graph of the wave form suggests, the scanning signal is the exact electrical equivalent of the motion of the spinning mirror.

© 1974 SCIENTIFIC AMERICAN, INC

ing the climax of years of work with electricity in gases at the Cavendish Laboratory of the University of Cambridge. Ten weeks after Braun's paper was published Thomson announced his discovery that cathode rays consist of "corpuscles" less massive than atoms and carrying identical negative charges, in other words, electrons.

The success of Braun's pioneering experiments is remarkable in view of the limitations of his apparatus and techniques. Because the cathode was not heated, as it is in most modern tubes, a potential of as much as 50,000 volts was required between cathode and anode to produce cathode rays of sufficient intensity. This large potential caused undesirable leakage currents inside the tube and electrostatic charges on the glass envelope. It also diminished the sensitivity of the tube, that is, the angular displacement of the electron beam for a given deflecting force. As the voltage is increased the velocity of the electrons, and thus their momentum, also increases, so that a larger force must be applied to deflect the beam.

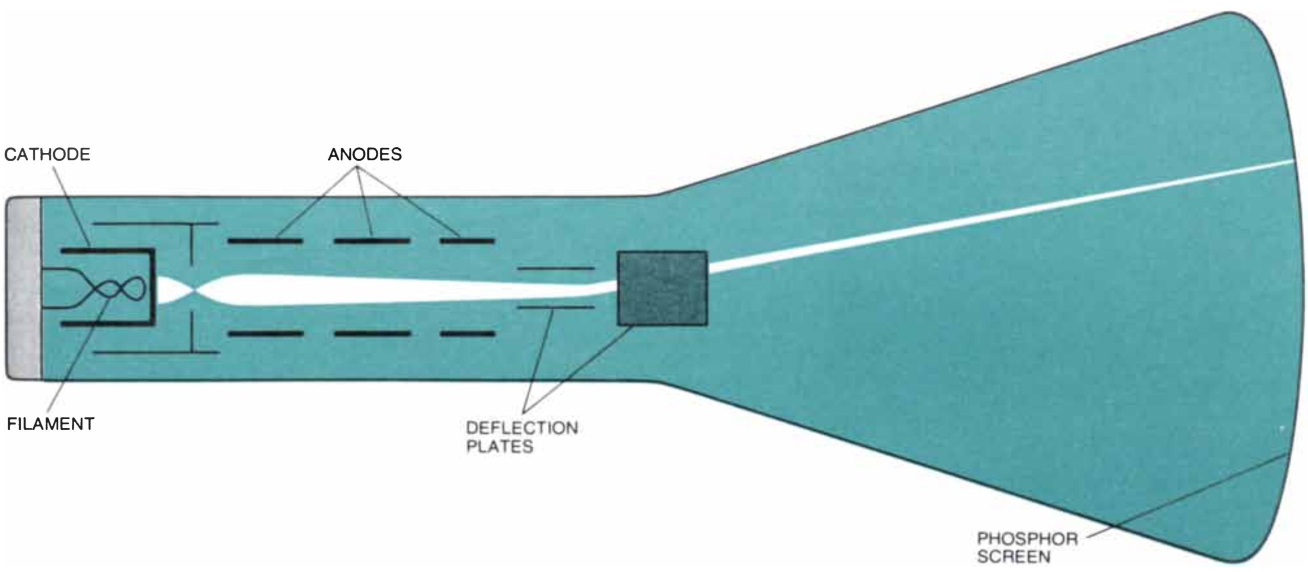
These factors, as well as the inconstant output of the high-voltage source (influence machine or induction coil), caused unstable operation, flickering of the trace and vagueness of the fluorescent patterns. The image was further degraded by the crude method employed to deflect the beam. With a coil on only one side of the tube the magnetic field was asymmetrical; this dispersed the beam so that it produced an elliptical patch rather than a circular spot on the screen. The aperture in the diaphragm governed the size of the spot, but Braun could not control the sharpness of its outline, or definition, since the tube did not provide a means of focusing the beam. Therefore the spot was fuzzy and the trace indistinct, particularly in comparison with the crisp patterns then obtainable by optical methods.

Braun was aware of some of these shortcomings in his instrument. He pointed out that the method required "strong forces," hinted at possible sources of error and suggested certain precautions, such as surrounding the tube with a grounded shield. He nevertheless expressed confidence that the tube offered "definite advantages for many scientific purposes."

Braun's work on the indicator tube was continued by Jonathan Zenneck, one of his assistants. By 1899 Zenneck had designed an improved tube, made photographic records of various traces and invented an electrical method for recursively deflecting the beam on the horizontal axis at a constant rate. This "time base" method of horizontal scanning enabled the instrument to display wave forms without the rotating mirror. Although it did not eliminate all the moving mechanical parts (a motor was still required in the circuit that generated the scanning signal), it enabled the fully constituted trace to be displayed on the phosphor viewing screen inside the tube. During the next few years Zenneck explored the possibilities of the tube in a variety of applications, including the infant art of signaling without wires.

In 1897 public demonstrations in England and Italy by Guglielmo Marconi brought wireless telegraphy to the attention of the world. By the end of that year the new method of communication was being investigated in Germany. Braun's experience with alternating current and his knowledge of oscillatory circuits led him naturally to the new field. He made several contributions of fundamental importance, among them a circuit to couple a spark generator with an antenna system that increased the "energy of action," or radiated power, and greatly extended the range of transmission. He also introduced antennas for directional transmission and reception.

Braun and Marconi were jointly awarded the Nobel prize for physics in 1909 for their contributions to wireless telegraphy. In his acceptance speech Braun described experiments with the cathode ray tube and also spoke of his early investigations of crystalline materials, such as galena and the pyrites. Recalling his own work with these minerals in 1874, and the fact that certain crystals allow current to pass more freely in one direction than in the other, Braun had recognized about 1900 that these substances could rectify high-fre-



MODERN OSCILLOSCOPE TUBE is almost identical in principle with Braun's instrument but is much changed in detail. Electrostatic deflection is substituted for the electromagnetic coils Braun employed, and the single wire anode has been replaced by a series of cylindrical anodes that accelerate the electrons, focus them into a narrow beam and control the brightness of the image. The heated cathode emits larger numbers of electrons at smaller potentials and thus allows the deflecting forces to be reduced without diminishing the tube's sensitivity. The modern tube also maintains a better vacuum than could be achieved with the apparatus Braun employed.

© 1974 SCIENTIFIC AMERICAN, INC

quency currents and thus serve as detectors in radio receivers. His psilomenan detector, introduced in 1901, was the prototypical crystal detector and the first practical electronic device to employ semiconductor materials.

Early in 1915, a few months after war began in Europe, Braun visited the U.S. on behalf of the Telefunken Company; the visit was to bring a melancholy end to a brilliant career. He stayed until it became impossible for him to return to Germany. Living with his son Konrad in New York, but deprived of his laboratory and with little means to follow his interests, he was forced to spend his last years divorced from his scientific life.

When Braun died in April, 1918, the indicator tube had been improved and was being applied to investigations that prefigured the science of electron optics. He had also seen his tube perform in a novel role: as a receiver in early television systems. Max Dieckmann and Gustave Glage had patented the first picture-transmission system that incorporated a Braun tube in 1906. The following year Boris Rosing had introduced another television system that incorporated a Braun tube in the receiver. In 1908 Alan A. Campbell Swinton had suggested the use of "two beams of Kathode rays" as a solution to the almost insuperable problems of mechanical television. Three years later he had made the first detailed proposal for an electronic television system based on the "ingenious oscillograph, invented a few years ago by Braun."

All of this took place before World War I. The rapid progress made in radiophysics and electron tubes during and after the war was manifested in the sudden emergence—and popularity—of commercial radio broadcasting at the beginning of the 1920's. In 1922 the Western Electric Company introduced a gas-focused cathode ray tube that operated on low voltages with simple auxiliary circuits. This oscilloscope, a direct descendant of Braun's curve-tracing device, was portable and convenient, the ideal instrument for the booming radio industry.

Braun did not patent his indicator tube, nor did he make a concerted effort to promote it. In many respects it was a premature invention. Braun's model was operating more than seven years before the first tube of importance to radio was patented, yet the indicator tube required decades of research and technological evolution before it became the indispensable device it is today.

Do you want to observe in comfort; carry your "portable observatory" with one hand? or—

Photograph, in sharp detail, wildlife near at hand or at inaccessible distances?

Or have you an industrial requirement, a need for the finest optical system available built into special ruggedized equipment?

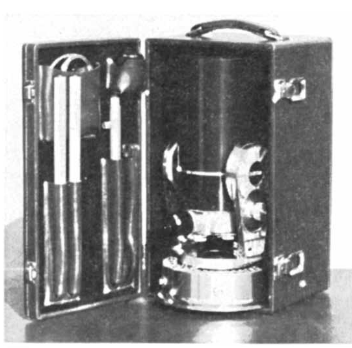
Other optical systems are performance-tested against Questar; why not come right to the source?

QUESTAR
DOES IT ALL

© Questar Corporation 1973
All rights reserved

Questar started out in life as a fine personal telescope in the days when you either ordered a conventional, custom-built refractor, or built a reflector yourself if you had happened to come across the wonderful book "Amateur Telescope Making" by the now-legendary A. G. Ingalls. But Questar's originator foresaw that we were entering upon a period which might almost be known as "the optical age," when this superfine optical system would be in demand not only as a multi-purpose telescope but for all sorts of modern engineering developments, to permit miniaturization and portability that would be otherwise impossible.

Because of its world-wide reputation for quality and versatility, Questar is the *specified* optical system, and it is on-the-shelf here, ready for many special applications. Our ruggedized versions save you engineering time and money. Why not let us help solve your problems?



QUESTAR, THE WORLD'S FINEST, MOST VERSATILE TELESCOPE, IS PRICED FROM $865. SEND FOR OUR BOOKLET IN COLOR WITH 150 PHOTOGRAPHS BY QUESTAR OWNERS. $1 TO COVER MAILING AND HANDLING COSTS ON THIS CONTINENT; BY AIR TO SOUTH AMERICA $2.50; EUROPE AND NORTH AFRICA, $3; ELSEWHERE, $3.50.

QUESTAR

BOX 520, NEW HOPE, PENNSYLVANIA 18938

© 1974 SCIENTIFIC AMERICAN, INC

MATHEMATICAL GAMES

Reflections on Newcomb's problem: a prediction and free-will dilemma

by Martin Gardner

This department's topic for July, 1973, Newcomb's paradox, produced an enormous outpouring of letters. Robert Nozick, who first wrote about the paradox in a paper published in 1970, agreed to look over the correspondence and put down his reactions. Nozick is a philosopher at Harvard University and the author of *Anarchy, State and Utopia,* a book that will be published this summer by Basic Books. William A. Newcomb, the man who discovered the paradox, is a theoretical physicist at the Lawrence Livermore Laboratory of the University of California.

What follows is the communication I received from Nozick in October. May I urge readers who wish to write again not to do so until they have read Nozick's original paper? It goes into considerably more technical detail than my first article or Nozick's present comments.

Newcomb's problem involves a Being who has the ability to predict the choices you will make. You have enormous confidence in the Being's predictive ability. He has already correctly predicted your choices in many other situations and the choices of many other people in the situation to be described. We may imagine that the Being is a graduate student from another planet, checking a theory of terrestrial psychology, who first takes measurements of the state of our brains before making his predictions. (Or we may imagine that the Being is God.) There are two boxes. Box 1 contains $1,000. Box 2 contains either $1 million or no money.

You have a choice between two actions: taking what is in both boxes or taking only what is in the second box. If the Being predicts you will take what is in both boxes, he does not put the $1 million in the second box. If he predicts you will take only what is in the second box, he puts the million in the second box. (If he predicts you will base your choice on some random event, he does not put the money in the second box.) You know these facts, he knows you know them and so on. The Being makes his prediction of your choice, puts the $1 million in the second box or not, and then you choose. What do you do?

There are plausible arguments for reaching two different decisions:

1. The expected-utility argument. If you take what is in both boxes, the Being almost certainly will have predicted this and will not have put the $1 million in the second box. Almost certainly you will get only $1,000. If you take only what is in the second box, the Being almost certainly will have predicted this and put the money there. Almost certainly you will get $1 million. Therefore (on plausible assumptions about the utility of the money for you) you should take only what is in the second box [*see illustration on opposite page*].

2. The dominance argument. The Being has already made his prediction and has either put the $1 million in the second box or has not. The money is either sitting in the second box or it is not. The situation, whichever it is, is fixed and determined. If the Being put the million in the second box, you will get $1,001,000 if you take both boxes and only $1 million if you take only the second box. If the Being did not put the money in the second box, you will get $1,000 if you take both boxes and no money if you take only the second box. In either case you will do better by $1,000 if you take what is in both boxes rather than only what is in the second box [*see illustration on page 104*].

Each argument is powerful. The problem is to explain why one is defective. Of the first 148 letters to *Scientific American* from readers who tried to resolve the paradox, a large majority accepted the problem as being meaningful and favored one of the two alternatives. Eighty-nine believed one should take only what is in the second box, 37 believed one should take what is in both boxes—a proportion of about 2.5 to one. Five people recommended cheating in one way or another, 13 believed the problem's conditions to be impossible or inconsistent and four maintained that the predictor cannot exist because the assumption that he does leads to a logical contradiction.

Those who favored taking only the second box tried in various ways to undercut the force of the dominance argument. Many pointed out that if you thought of that argument and were convinced by it, the predictor would (almost certainly) have predicted it and you would end up with only $1,000. They interpreted the dominance argument as an attempt to outwit the predictor. This position makes things too simple. The proponent of the dominance argument does believe he will end up with only $1,000, yet nevertheless he thinks it is best to take both boxes. Several proponents of the dominance principle bemoaned the fact that rational individuals would do worse than irrational ones, but that did not sway them.

Stephen E. Weiss of Morgantown, W.Va., tried to reconcile the two views. He suggested that following the expected-utility argument maximizes expectation, whereas following the dominance argument maximizes correct decision. Unfortunately that leaves unexplained why the correct decision is not the one that maximizes expectation.

The assumptions underlying the dominance argument, that the $1 million is already in the second box or it is not and that the situation is fixed and determined, were questioned by Mohan S. Kalelkar, a physicist at the Nevis Laboratories of Columbia University, who wrote: "Perhaps it is false to say that the Being has definitely made one choice or the other, just as it is false to say that the electron [in the two-slit experiment] went through one slit or the other. Perhaps we can only say that there is some amplitude that B2 [second box] has $1 million and some other amplitude that it is empty. These amplitudes interfere unless and until we make our move and open up the box.... To assert that 'either B2 contains $1 million or else it is empty' is an intuitive argument for which there is no evidence unless we open the box. Admittedly the intuitive evidence is strong, but as in the case of the double-slit electron diffraction our intuition can sometimes prove to be wrong."

Kalelkar's argument makes a version of the problem, in which the second box is transparent on the other side and someone has been staring into it for a week before we make our choice, a significantly different decision problem. It seems not to be. Erwin Schrödinger, in a

© 1974 SCIENTIFIC AMERICAN, INC

famous thought experiment, imagined a cat left alone in a closed room with a vial of cyanide that breaks if a radioactive atom in a detector decays. Must a disciple of Niels Bohr's assert that the cat is neither alive nor dead, Schrödinger asked, until measurements have been made to decide the case? Even if one accepts the Bohr interpretation of quantum mechanics, however, what choice *does* one make, in Newcomb's problem, when one knows that others can see into the box from the other side and observe whether it is filled or empty?

Many who wrote asserted that the dominance argument assumes the states to be probabilistically independent of the actions and pointed out that this is not true for the two states "The $1 million is in Box 2" and "The $1 million is not in Box 2." The states would be probabilistically independent of the actions (let us assume) in the matrix for the utility argument, which has the states "He predicts correctly" and "He predicts incorrectly." Here, however, there is no longer dominance. Therefore it appears that the force of dominance principles is undercut. "It is legitimate to apply dominance principles if and only if the states are probabilistically independent of the actions. If the states are not probabilistically independent of the actions, then apply the expected-utility principle, using as the probability-weights the conditional probabilities of the states given in the actions." The quotation is from my original 1970 essay, which formulated this position, then went on to reject it as unsatisfactory for the following reasons.

Suppose a person knows that either man *S* or man *T* is his father but he does not know which. S died of some very painful inherited disease that strikes in one's middle thirties and *T* did not. The disease is genetically dominant. S carried only the dominant gene. *T* did not have the gene. If S is his father, the person will die of the dread disease. If *T* is his father, he will not. Furthermore, suppose there is a well-confirmed theory that states a person who inherits this gene will also inherit a tendency toward behavior that is characteristic of intellectuals and scholars. S had this tendency. Neither *T* nor the person's mother had such a tendency. The person is now deciding whether to go to graduate school or to become a professional baseball player. He prefers (although not enormously) the life of an academic to that of a professional athlete. Regardless of whether or not he will die in his middle thirties, he would be happier as an academic. The choice of the academic life would thus appear to be his best choice.

Now suppose he reasons that if he decides to be an academic, the decision will show that he has such a tendency and therefore it will be likely that he carries the gene for the disease and so will die in his middle thirties, whereas if he chooses to become a baseball player, it will be likely that *T* is his father, therefore he is not likely to die of the disease. Since he very much prefers not dying of the disease (as a baseball player) to dying early from the disease (as an academic), he decides to pursue the career of an athlete. Surely everyone would agree that this reasoning is perfectly wild. It is true that the conditional probabilities of the states "S is his father" and "*T* is his father" are not independent of the actions "becoming an academic" and "becoming a professional athlete." If he does the first, it is very likely that S is his father and that he will die of the disease; if he does the second, it is very likely that *T* is his father and therefore unlikely that he will die of the disease. But who his father is cannot be changed. It is fixed and determined and has been for a long time. His choice of how to act legitimately affects our (and his) estimate of the probabilities of the two states, but which state obtains (which person is his father) does not depend on his action at all. By becoming a professional baseball player he is not making it less likely that S is his father, therefore he is not making it less likely that he will die of the disease.

This case, and others more clearly including a self-reference that this case may seem to lack, led me to think probabilistic nonindependence was not sufficient to reject the dominance principle. It depends on whether the actions influence or affect the states; it is not enough merely that they affect our judgments about whether the states obtain. How do those who reject the dominance principle for Newcomb's problem distinguish it from those other cases where dominance principles obviously apply even though there is probabilistic nonindependence?

But one must move carefully here. One cannot force a decision in a difficult case merely by finding another similar case where the decision is clear, then challenging someone to show why the decision should be different in the two cases. There is always the possibility that whatever makes one case difficult and the other clear will also make a difference as to how they should be decided. The person who produces the parallel example must not only issue his challenge; he must also offer an explanation of why the difficult case is less clear, an explanation that does not involve any reason why the cases might diverge in how they should be decided. Interested readers can find my additional parallel examples where dominance is appropriate, plus an attempt to explain why Newcomb's case, although less clear, is still subject to dominance principles, in my original essay, "Newcomb's Problem and Two Principles of Choice," in *Essays in Honor of Carl G. Hempel*, edited by Nicholas Rescher, Humanities Press, 1970.

This obligation to explain differences in the clarity of parallel examples in order to show that no different decision should be made also rests on those who argued in their letters for taking only what is in the second box. For example, it rests on Robert Heppe of Fairfax, Va., who said that the situation "is isomorphic with one in which the human moves first and openly," and on A. S. Gilbert of the National Research Council of Canada, who called the Newcomb case "effectually the same as" one where you act first and an observer attempts to communicate with a "mindreader" in the next room who then guesses your choice, using a payoff matrix identical with Newcomb's.

A large number of those who recommended taking only the second box performed the expected value calculation and concluded that, provided the probability that the Being was correct was at least .5005, they would take only the second box. Not only did they see no problem at all; they either maximized expected monetary value or made utility linear with money in the range of the problem. Otherwise the cutoff probability would be different. William H. Riker

	HE PREDICTS YOUR CHOICE CORRECTLY	HE PREDICTS YOUR CHOICE INCORRECTLY
TAKE BOTH	$1,000	$1,001,000
TAKE ONLY SECOND	$1,000,000	$0

Payoff matrix for expected-utility argument

© 1974 SCIENTIFIC AMERICAN, INC

of the department of political science at the University of Rochester suggested that people making different decisions merely differed in their utility curves for money. Such persons, however, need not differ in their choices among probability mixtures of monetary amounts in the standard situations in order to calibrate their utilities.

Those who favored taking both boxes made almost no attempt to diagnose the mistakes of the others. An exception is William Bamberger, an economist at Wayne State University. He wrote that the proponent of choosing only the second box "computes not the alternative payoffs of choosing one or two boxes for a given individual, but the average payoff of those who choose two as opposed to the average observed payoff of those who choose one." The problem, of course, is how to compute the probability for a given individual of his payoff for each choice. Should one use the differing conditional probabilities, or ignore them because dominance applies only when the states are probabilistically independent of the action (and so when for each state its conditional probabilities on each act is the same), or ignore them since the conditional probabilities of the state on the acts are to be used only when they represent some process of the act's influencing or affecting which state obtains?

A number of respondents said their choice would depend on whether the predictor made his prediction after they had at least started to consider the problem. If so, they would do their best to decide to take only the second box (so that this data would be available to the predictor), and some added that they hoped they would change their mind at the last minute and take both boxes. (They gave the predictor too little credit.) On the other hand, if the predictor made his prediction before they even considered the problem, these writers believed they would take both boxes, since there was no possibility of their deliberations affecting the prediction that had been made.

Several respondents maintained that if the conditions of the problem could be realized, we might be forced to revise our views about the impossibility of backward causality. Newcomb himself seems to think that special difficulties arise for proponents of backward causality if the predictor writes some term designating an integer on a slip of paper in the second box, with the understanding that you get $1 million only if that integer is a prime. Of course, the predictor writes a prime if, and only if, he predicts that you will take the second box. How can your choice determine whether a number is prime or composite? The advocate of backward causality need not think it does. What your choice affects, in his view, is what term the predictor writes down (or wrote down earlier), not whether the integer it designates is prime or composite.

The reasoning of some of the letters indicates it would be useful to specify precisely the conditions whereby we could discover in which time-direction causality operates. Might one even say that some conditions universally preceding certain decisions are part of the effects of the decision (by backward causality) rather than part of the cause?

Not everyone was willing to choose one or the other action. Among the five respondents who suggested some form of cheating, Robert B. Pitkin, editor of *American Legion Magazine*, speculated that Dr. Matrix, the numerologist, would walk in with a device to scan the contents of the boxes, take the boxes with the money in them and never open an empty box. "He quite naturally succeeded in getting all the money, for the rule of bridge that one peek is worth two finesses applies here too.... By introducing a choice which the Being has not anticipated, and is not permitted to take into account, he achieves a stunning victory for free will." (What prevents the Being from taking this into account?)

Other letter writers also struck blows for free will. Nathan Whiting of New York would take both boxes but would open only the first one, leaving the second box unopened. Ralph D. Goodrich, Jr., of Castle Rock, Colo., would take only the first box. Richard B. Miles of Los Altos, Calif., also recommended a "creative" solution: Turn to another person before you make your choice and offer to sell him for $10,000 the contents of whatever box or boxes you choose.

Isaac Asimov wrote: "I would, without hesitation, take both boxes.... I am myself a determinist but it is perfectly clear to me that any human being worthy of being considered a human being (including most certainly myself) would prefer free will, if such a thing could exist.... Now, then, suppose you take both boxes and it turns out (as it almost certainly will) that God has foreseen this and placed nothing in the second box. You will then, at least, have expressed your willingness to gamble on his nonomniscience and on your own free will and will have willingly given up a million dollars for the sake of that willingness—itself a snap of the finger in the face of the Almighty and a vote, however futile, for free will.... And, of course, if God has muffed and left a million dollars in the box, then not only will you have gained that million but *far more important* you will have demonstrated God's nonomniscience. If you take only the second box, however, you get your damned million and not only are you a slave but also you have demonstrated your willingness to be a slave for that million and you are not someone I recognize as human." (No one wrote to argue for taking only the second box on the grounds that either it results in getting $1 million or it demonstrates the Being's fallibility, either of which is desirable.)

Those who held that the conditions of the problem could not be realized were of two types. There were those who believed the situation to be physically impossible because the Being could not predict all the information input of every light signal that would arrive at your eyes in the appropriate time interval. ("To gain such knowledge the Being must have a physical agency for collecting information that travels faster than the speed of light," wrote George Fredericks, a physicist at the University of Texas.) And there were those who argued that if the room is closed, the problem reduces to that of Maxwell's demon—a suggestion made by Fredericks and by John A. Ball of the Harvard College Observatory.

Those who believed the conditions of the problem to be inconsistent as well as physically impossible said that the almost certain predictability of decisions was inconsistent with free will, and therefore with making choices, yet the problem assumed that genuine choices could be made. This is a hard argument

	HE PUT $1,000,000 INTO BOX 2	HE DID NOT PUT $1,000,000 INTO BOX 2
TAKE BOTH	$1,001,000	$1,000
TAKE ONLY SECOND	$1,000,000	$0

Payoff matrix for dominance argument

© 1974 SCIENTIFIC AMERICAN, INC

FAR EAST & AUSTRALIA

10 Express flights a week by SAS, Thai International and Aeroflot.

connections from all over Europe to Copenhagen/ Frankfurt	Monday	Tuesday	Wednesday	Thursday Friday Saturday	Saturday	Sunday	TRANS-SIBERIAN EXPRESS Thursday	TRANS-SIBERIAN EXPRESS Saturday
	SAS	SAS	THAI	SAS	THAI	Aeroflot	Aeroflot	SAS
departures:								
LONDON	11.20	15.10	10.05	8.15	10.50	11.20	8.15	8.15
COPENHAGEN	15.00	16.25	12.05	12.05		15.00	12.05	12.00
FRANKFURT	12.50	18.30	14.05	9.30		12.50	9.30	9.30
	Tashkent	NONSTOP	NONSTOP	Tashkent	NONSTOP	Moscow	Moscow	Moscow
arrivals, following day:								
DELHI		7.00				5.05		
BANGKOK	9.20	12.50	6.55	6.25	6.00	11.30		
HONG KONG[1]	14.10	17.25	11.25	11.25	11.25	19.40		
KUALA LUMPUR		15.50						
SINGAPORE	12.55	17.10	10.35[3]	13.15/11.05/10.35	10.35[3]	15.00		
JAKARTA	14.25							
SYDNEY			22.20[3]	22.20[2]	22.20[3]			
TOKYO							9.35	9.10

[1]) Connecting flights from Bangkok. [2]) Thai International from Bangkok, Sunday.
[3]) Connecting flight from Bangkok, through-going aircraft Europe – Sydney.

FIRST AND ONLY NONSTOP SERVICES
FRANKFURT—DELHI (SAS)
FRANKFURT—BANGKOK (THAI)
LONDON—BANGKOK (THAI)

ONLY SERVICES WITH NEXT-DAY ARRIVAL IN SYDNEY – JUST ONE NIGHT EN ROUTE!

SAS SCANDINAVIAN AIRLINES
Representing THAI International

© 1974 SCIENTIFIC AMERICAN, INC

to drive through because it appears to be the choices that are predicted. The relevant connections are difficult to get straight. Predictability of decisions does not logically imply determinism under which the decisions are caused (for example the possibility of backward causality where an uncaused decision causes an earlier prediction, or "seeing ahead" in time in a block universe).

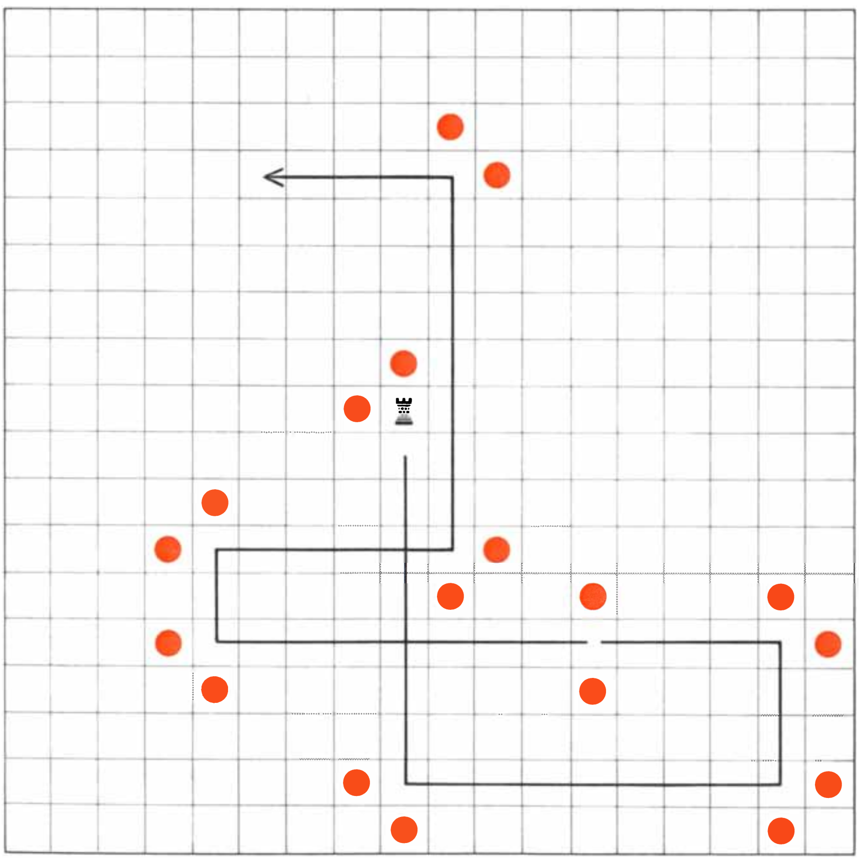
Nor, we should note in passing, does determinism entail predictability, even in principle. Events could be fixed in accordance with scientific laws that are not recursive. Is determinism incompatible with free will? It seems to many to be so, yet the argument that determinism is incompatible with responsibility for action, which free will implies, depends on a notion of responsibility insufficiently worked out to show precisely how the connections go. Some say merely that a free act is an uncaused one. Yet being uncaused obviously is not sufficient for an act to be free; one surely would not be responsible for such an action. What other conditions, then, must be satisfied by an uncaused act if it is to be a free one? The literature on free will lacks a satisfactory specification of what a free action would be like (given that "uncaused" is not enough). Perhaps if we were given this specification of additional conditions, they would turn out to be sufficient apart from the action's being uncaused.

Another problem will help to exhibit some complicated relations between free will and determinism. It has been asserted (by C. S. Lewis, for instance) that no determinist rationally can believe in determinism, for if determinism is true, his beliefs were caused, including his belief in determinism. The idea seems to be that the causes of belief, perhaps chemical happenings in the brain, might be unconnected with any reasons for thinking determinism true. They might be, but they need not be. The causes might "go through" reasons and be effective only to the extent that they are good reasons. In the same way it might be a causal truth about someone that he is convinced only by arguments that constitute specified types of good reasons (deductive, inductive, explanatory and so on).

Some philosophers have argued recently that we know some statement p only if part of the cause (or more broadly the explanation) of our believing p is, if we pursue the story far enough, the fact that p is true. You know this magazine is before you now only if its being there is part of the explanation of why you believe it is there. If psychologists are stimulating your brain to create the illusion that you are seeing a magazine, you would not really know there is a magazine before you even if a psychologist happened to have left one on the table in front of you. The magazine's being there would not play the proper causal role in the story of your belief. If we do not mind our beliefs being caused by the facts, and indeed find it somewhat plausible to think we have knowledge only to the extent that they are, then we may also find it less disturbing that our actions are caused by certain types of facts holding in the world, for example, the fact that it would be better to do one thing rather than another. To say this, of course, is not to present a theory of free action; it is merely to hint that it may be possible to remove the sting of determinism. This approach is a comfortable one when we act correctly, but it is difficult to see how it can be extended plausibly to wrong acts where questions of responsibility are particularly pressing.

Proponents of the C. S. Lewis position might reply that the determinist should not feel so comfortable. Even though he says he is caused to believe in determinism (and anything else) by what are good reasons, he must also maintain that he is caused to believe that such reasons are good reasons. He may have a second set of reasons for believing the first set of reasons are good. Now, however, his opponent can raise the same question as before. Why does he believe the second set of reasons? The determinist must end either by finding self-supporting reasons (which say of themselves that they are good reasons) or by admitting that the best explanation of why he believes they are good reasons is that they are. This surely leaves his opponent unsatisfied, and the match seems to be a draw.

Those who believe in free will find themselves in similar dilemmas. Kurt Rosenwald of Washington wrote: "When I was 19 or 20, I thought about the free-will problem... and I came to this conclusion: If we make an exhaustive study of that problem, and finally arrive at the result that our will is free, we still will not know whether our will is indeed free or our mind is of such a nature that we have to find our will to be free, although it is not, in fact, free. This became one of my reasons for studying not philosophy but the natural sciences. Thinking about it now, 50+ years later, it still seems to me that I was right." But does not the possibility that we are caused to



How to trap a rook with two quads per move

© 1974 SCIENTIFIC AMERICAN, INC

Fly Smithsonian.

Have you ever been to a place where they teach eagles to be wild?
Have you ever journeyed deep into the mind of a man who single-handedly revolutionized Man's entire perception of the universe?
Have you ever penetrated to the heart of the Cretan Maze, to confront the Minotaur?

Visit some of the world's most enlightening destinations.
Fly with SMITHSONIAN Magazine for the next 12 months... for less than one dollar a month.

We invite you to spread the wings of your imagination every month, and come with us into many surprising worlds of nature, art, history and science. And we invite you to bring your entire family along.

When did Man first set foot on the North American continent—and think about when Man is likely to first set foot on Mars. Is rainmaking becoming respectable? Why is it essential to save the prairie dog from extinction? What *really* happened the night that Paul Revere made his famous ride?

What happens when the gold-bedecked chieftains of the Akan pay a state visit to their new King? What is currently the rarest stamp in the world? Were dinosaurs the only warm-blooded reptiles that ever lived? What causes the brilliant colors of butterfly wings? How close are we to directly harnessing the power of the Sun?

If these questions intrigue you, you'll be intrigued and pleased with SMITHSONIAN. And if you are attracted and moved by beauty, you'll love SMITHSONIAN the more . . . because it is one of the handful of magazines in the world today that is almost a work of art in itself. Here every month you will find a dazzling display of beauties both ancient and modern—from the most recently discovered Pompeiian frescoes to the finest works of modern masters (many of them never seen outside private collections before).

What's more, as beautiful, informative, intelligent and entertaining as SMITHSONIAN Magazine is, it's *only one* of the knowledge-enhancing benefits you and your family receive when you become a National Member of the Smithsonian Associates.

Much more than a magazine

In addition to SMITHSONIAN every month, your membership in the Smithsonian Associates entitles you and your family to:

■ **Special travel programs** organized for Members only—nearly 50 opportunities each year to visit historic sites, art centers and scientific installations (many of which are not open to the general public).

■ **Discounts on quality books** dealing with history, art, science and nature.

■ **Discounts on unusual gifts** from the museum shops of the Smithsonian Institution . . . fascinating arts and crafts from all corners of the globe.

■ **Special welcome** when you and your family visit the Smithsonian in Washington, D.C. including use of the Member's Lounge and the privilege of dining with the Smithsonian staff.

The contribution you are making to knowledge

Beyond the extensive, almost unlimited, resources and opportunities for learning that your Membership will bring to you and your family, a substantial portion of your annual $10 Membership Fee goes *to help increase the fund of Man's knowledge.*

It goes to help support the many research and educational programs of the Smithsonian Institution's private sector which depends on the private support of individuals for its activities.

Join today— return the coupon below

If you are interested in exploring many new regions of the world and of the mind, if you respond to the historic and the beautiful and the innovative, if you would like to help increase the range of Man's mind, join the Smithsonian today. All it takes is a few brief moments to fill out and mail the coupon below to spread the wings of your mind.

Mail this coupon to:

SMITHSONIAN Membership Service
P.O. Box 2606 • Greenwich, Conn. 06830

Please enroll me as a National Member of the Smithsonian Associates, and enter my subscription for the next year (12 issues) of SMITHSONIAN.

Bill me for the annual dues of $10.

Name *(please print clearly)*

Address

City

State Zip

MEMBERSHIP BENEFITS
■ SMITHSONIAN Magazine — 12 issues
■ Travel Program — Domestic and Foreign
■ Book and Gift Discounts
■ Reception Center at Smithsonian Institution in Washington, D.C.

P6-34

© 1974 SCIENTIFIC AMERICAN, INC

believe in false conclusions apply also to conclusions in the natural sciences? And to the verdict of 50+ years later?

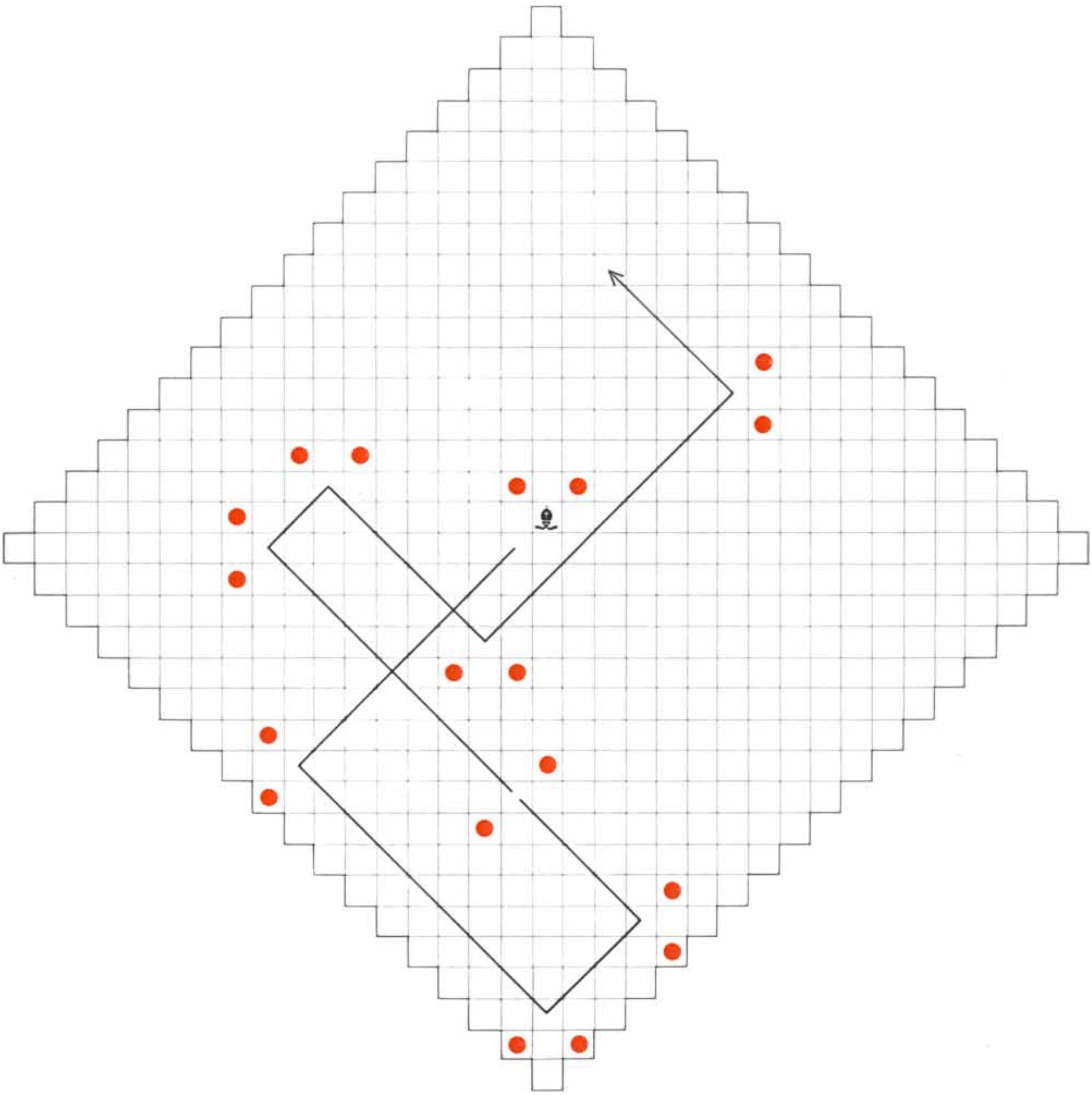
I published my original essay after thinking about Newcomb's problem intermittently for five years. In that essay I expressed the hope that someone would come forth with a solution to the problem that would enable me to stop returning to it. It is not surprising that no one did, yet it is surprising (to me) that the mere act of publishing Newcomb's problem, and sending my thoughts on it into the world, rid me of it. That is, I was rid of it until the problem was presented in *Scientific American* and I was invited to read more than 650 pages of letters about it. Unfortunately the letters do not, in my opinion, lay the problem to rest. And they have started me thinking about it again! You can't win.

Answers to some of the questions raised last month follow:

If a rook's maximum move in a game of quadraphage is n cells, it can be trapped by two quads per move on a board of side $2n + 2$. The strategy is to consider the unobstructed paths from the rook to sides of the board and place the quads adjacent to the rook to block its movement to the two nearest sides. (If two borders are the same distance away, choose either one.) The illustration on page 106 shows the strategy on a go board, the rook limited to maximum moves of eight cells. The rook clearly can never reach the edge. Eventually it must head toward a quad. When this happens, quads on each side confine it to a segment of the path and entrapment quickly follows.

The same procedure will trap a bishop within the borders of a sawtooth board, $2n + 2$ on a side, where n is the bishop's longest move. The illustration below shows how the strategy operates when the bishop can move no more than eight cells.

I conjecture that a similar strategy, using four quads per move, will trap the



How to trap a bishop with two quads per move

© 1974 SCIENTIFIC AMERICAN, INC

queen. For n greater than 2 a board of at least $4n + 2$ on a side seems necessary. (Thus a queen with maximum move 8 can be trapped on a 34-sided field.) The first move is to put quads on the four corner cells. Thereafter use the "nearest sides" strategy. When there is a choice between blocking equal paths on an orthogonal and a diagonal, block the orthogonal.

Some of the questions raised about "spirolaterals," in last November's column, have been answered by readers. The main question—How can you determine from a spirolateral's formula whether it closes and, if it does, in how many repetitions?—was answered by James Thomas, William Laubenheimer, Steven Wolfson and E. Lawrence McMahon. It turns out (as other readers also reported) that a spirolateral results whenever the angle of turn is a rational number. If the angle is irrational, the spirolateral remains within a bounded region but never closes.

Thomas gave the following procedure: First determine the angle's supplement (its difference from 180 degrees). Multiply this by the difference between the number of left turns and number of right turns. (The difference is equal to the spirolateral's order minus twice the number of left turns.)

From the above result, keep subtracting 360 until the remainder is between -180 and 180. Take the absolute value and call it x. This represents the net angular change after each cycle. If x equals 0, there is either no closure (and the spirolateral is infinite) or it closes after the first cycle.

If x is not zero, divide it into the lowest multiple of 360 that it will go into evenly. The result is the number of cycles required to close the spirolateral.

To express this procedure by a compact formula, McMahon proposed letting n equal the spirolateral's order, k equal the number of left turns and m equal 360 divided by the rational angle. Write the following fraction,

$$\frac{(m-2)(n-2k)}{2m},$$

and reduce to lowest terms. If the result is an integer, the spirolateral either does not close or closes after one cycle. If the result is an integral fraction, a/b, the figure closes after b cycles.

Correction: In January's column it was inadvertently stated that six coins are used in consulting the *I Ching*. As the description of the tossing procedure made clear, only three coins are used.

ANTIQUITY

A Quarterly Review of Archaeology
edited by Glyn Daniel

Founded in 1927 by O. G. S. Crawford, ANTIQUITY has appeared regularly ever since and won acclaim the world over as the most authoritative journal in its field. While written by specialists, the articles, notes and reviews are popular in character and indispensable to all interested in the development of man and his past.

Dr Glyn Daniel of the Faculty of Archaeology and Anthropology in the University of Cambridge, and Fellow of St John's College, has been Editor since 1956.

The annual subscription, postage included, is $12. Subscription forms and bankers' orders are available on request from

ANTIQUITY PUBLICATIONS LIMITED
Heffers Printers Ltd, King's Hedges Road,
Cambridge, England CB4 2PQ

AN INVITATION TO JOIN THE . . .
AMERICAN LITTORAL SOCIETY

- To learn about what lives in coastal waters, marshes, and estuaries, and why.
- To support the Littoral Society's coastal zone conservation projects.
- To receive Society publications — a quarterly journal, newsletters, conservation alerts.

Annual dues are $7.50 ($5 for students) and help support the Littoral Society's conservation projects.

AMERICAN LITTORAL SOCIETY
HIGHLANDS, NEW JERSEY 07732

To join the Littoral Society, please fill in the coupon below:

Please enroll me as a member of the American Littoral Society.
Enclosed is my check for $ ______ for first-year dues.

NAME
ADDRESS
CITY
STATE ZIP

Dues are tax deductible.

FREE catalog

over 1500 unique tools, handy kits, precision instruments, technical supplies.

Our 22nd year of service to the World's finest craftsmen and technicians.

National Camera
2000 West Union Ave. Dept. JAC
Englewood, Colorado, 80110

Send a FREE copy of the NC Flasher
name
address
city
state zip

National Camera
2000 West Union Ave. Dept. JAC
Englewood, Colorado, 80110

© 1974 SCIENTIFIC AMERICAN, INC

THE AMATEUR SCIENTIST

A new kind of spectrohelioscope for observing solar prominences

Conducted by C. L. Strong

Few natural spectacles compare in splendor with the glowing prominences that rise from the surface of the sun. Apart from astronomers few people ever see the display, which is usually lost in the sun's white glare. Those fortunate enough to be in the path of a total eclipse can observe the prominences as scarlet plumes that stand out in vivid contrast to the pearly background of the glowing corona.

Gene F. Frazier of 2705 Gaither Street S.E., Washington, D.C. 20031, views the spectacle routinely with a homemade instrument that in effect blocks from the eye light of all colors except the one emitted with maximum brilliance by the prominences. The hue approximates the darkest red of the setting sun. The emission is radiated by glowing hydrogen at a wavelength of 6,562.8 angstroms. In certain respects Frazier's apparatus resembles the spectrohelioscope previously described in these columns [see "The Amateur Scientist," SCIENTIFIC AMERICAN, April, 1958]. His instrument has an additional diffraction grating but requires no solar telescope or motor-driven optical parts. He describes the principles, construction and operation of the apparatus as follows.

"The effect that my instrument is based on (and that led to the development of the spectrohelioscope) was first described by the French astronomer Pierre Jules César Janssen following his observation in 1868 of the total solar eclipse in India. When Janssen focused the edge of the sun's image on the slit of his spectroscope, he was astonished by the brilliance of the spectral line at 6,562.8 angstroms. It was so bright that on the following day Janssen looked for the color in full sunlight. By opening the slit of the spectroscope he discovered that he could observe a portion of the prominence. A few days later the British astronomer Joseph Norman Lockyer hit on the same technique without the benefit of an eclipse.

"News of the discovery fascinated amateurs of the day, primarily because the brightness of the sun made observations possible with instruments of small aperture and proportionately low cost. It turned out, however, that home-built spectroscopes scattered too much light for satisfactory results. In addition clockwork-driven structures capable of keeping an image of the sun's edge focused exactly on the thin slit of the spectroscope called for a higher level of craftsmanship than most amateurs could attain.

"The design of my instrument sidesteps these requirements. Essentially the instrument employs an external diffraction grating to disperse and reflect sunlight to a concave mirror [*see illustration on opposite page*]. The mirror projects the rays through an adjustable plate of flat glass to a focus in the plane of the entrance slit of a conventional spectroscope.

"A photograph that could be made by putting a photosensitive plate in the position occupied by the slit of the spectroscope would not show the dark absorption lines that normally characterize the solar spectrum. In my system the image of the sun functions as the slit. Hence a photograph is composed not of the series of narrow absorption lines but of overlapping images of the solar disk separated by distances corresponding to the wavelength of the absorbed light.

"The adjustable plate of flat glass that admits incoming light to the slit acts as a vernier for displacing the rays laterally with respect to the slit. Rays that enter the plate at an angle to its perpendicular are refracted and emerge at the identical angle. The amount of deviation is approximately proportional to the angle between the plate and the entering beam. By rotating the plate the observer can shift the spectrum any small distance with respect to the slit. The plate functions as a precision tuner that enables the experimenter to admit any narrow portion of the spectrum to the slit.

"The selected rays, which may span a range of color only 10 angstroms wide, emerge from the slit as a diverging beam. The diverging rays fall on a concave mirror from which they are reflected as a bundle of parallel rays to the internal diffraction grating of the spectroscope [*see illustration on pages 112 and 113*]. The internal grating disperses the colors still more. The angle at which the internal grating is set can be adjusted to reflect rays of essentially monochromatic light to the second concave mirror of the spectroscope. The second mirror reflects the rays to focus in the plane of the eyepiece.

"The details of the filtering action can be demonstrated by replacing the external grating with a flat mirror and letting sunlight fall on the mirror. After adjustment an instrument so modified would display at the eyepiece the normal solar spectrum crossed by dark absorption lines. Assume that the geometry of the diffraction grating of the spectroscope is such that each angstrom of wavelength of the solar spectrum is dispersed through a distance of one millimeter in the focal plane of the eyepiece. This was essentially the case with Janssen's spectroscope. The chromosphere of his solar image was less than one millimeter wide. Therefore he could partly isolate the emission of the prominences from background light by carefully focusing this narrow feature of the image on the slit of his spectroscope.

"Now assume that the flat mirror is replaced by the external diffraction grating of my instrument and that the angle of the grating is carefully adjusted to reflect a narrow band of light on the slit that spans 10 angstroms (from, say, 6,558 to 6,568 angstroms). The spectrum is noncoherent. For this reason the light that reaches the slit consists of many monochromatic images of the sun's disk that overlap on each side of the hydrogen line at 6,562.8 angstroms. If

© 1974 SCIENTIFIC AMERICAN, INC

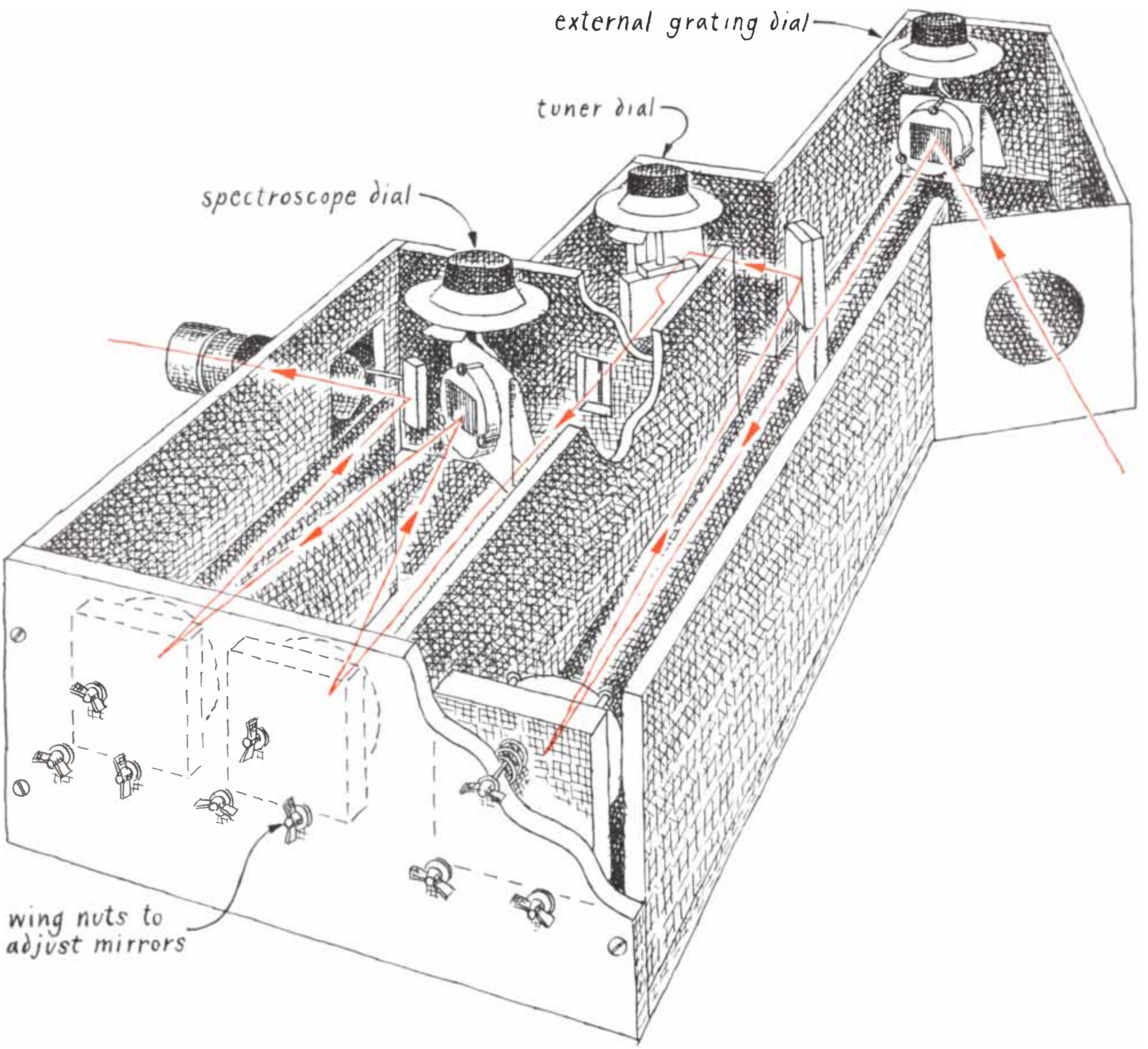
the dispersion of the gratings is assumed to be one angstrom per millimeter, the centers of each of the images of the sun's disk would be separated by one millimeter. At any setting 10 solar disks would overlap.

"This means that a band of color only 10 angstroms wide can enter the slit and that the scattering of light is significantly reduced. When the instrument is adjusted for observing prominences at 6,562.8 angstroms, unwanted light is reduced by more than 95 percent! Indeed, on a clear day it is not unusual for the field to appear completely dark five angstroms from the image. The full solar image appears in the field of view, which helps the observer to keep the edge of the image centered on the 6,562.8-angstrom line. With the aid of the tuner I have easily observed prominences continuously for intervals of more than two minutes.

"The construction requires no special skills, but the quality of the gratings is crucial. They must be mounted with care. The gratings should be ruled with at least 1,200 lines per millimeter for adequate resolution and high dispersion. The ruled area of the gratings in my instrument measures two inches square. The lines are blazed for 6,600 angstroms in the third order, which is to say that the surface of the rulings is cut at an angle that reflects light of maximum intensity at the 6,600-angstrom wavelength in the same direction as the grating disperses these rays in the third order.

"The gratings can be mounted in simple structures bent in the form of a *V* from sheet steel or brass. The gratings can be attached lightly to these mountings with machine screws and insulated from the metal with felt lining. The metal *V*'s are supported by soldering the rear side to a quarter-inch copper rod that fits a radio dial of the vernier type [*see top illustration on page 114*]. The copper rod must be bent to an angle such that the projected axis of the vernier dial bisects the plane of the grating. When the rod is so mounted, the angle of the gratings with respect to the impinging rays can be altered without displacing the spectral orders at the eyepiece.

"I made adjustable cells of plywood for supporting the mirrors. The cells are supported at three equidistant points by



Double-grating filter of Gene F. Frazier's spectroheliosope

© 1974 SCIENTIFIC AMERICAN, INC

machine screws fitted with compression springs and wing nuts. The mirrors can be lightly fastened to the plywood with wood screws insulated by rubber tubing and fiber washers. Incidentally, adjustable cells of cast aluminum are now available commercially at a reasonable price for mirrors of three-inch diameter or more. The cells also accept two-inch mirrors mounted in three-inch washers of plywood.

"The diameter of the mirrors is not critical, but it must be at least as large as the ruled area of the gratings to prevent vignetting (obscuration at the edges of the image) and the scattering of stray light into the image. In addition the focal length of the objective mirror should be an integral multiple of the focal length of the spectroscope mirrors, which, in turn, should be equal to within a tolerance of about 2 percent. The quality of the final image can be optimized by mounting the two mirrors of the spectroscope as close together as possible in order to minimize off-axis aberrations.

"The tuning plate can be made of any optically flat glass about 50 millimeters wide and six to 10 millimeters thick. The piece can be circular or rectangular. Plates of this size that were originally intended for use as optical windows are now available inexpensively from dealers in surplus optical supplies. The plate can be mounted by a frame of sheet metal and adjusted by a supporting shaft and a vernier dial.

"The construction of an adjustable slit of adequate quality has through the years remained the most difficult problem that confronts amateurs who make spectroscopes. The best slits by far are the ones that can be bought from distributors of optical supplies, but they are currently priced from $100 up. The slit must remain rigidly centered when its width is increased from zero to two or three millimeters, which means that both jaws must move equal distances in opposite directions when the device is adjusted.

"A mechanical linkage in the form of a parallelogram can satisfy this condition [*see illustration at bottom left on page 114*]. The system of links can be assembled with snugly fitting machine screws. Excess play in the system can be eliminated by inserting a pair of helical springs to maintain a few grams of tension between the side links that support the jaws of the slit. The jaws can be made of single-edge safety-razor blades, which can be fastened to the supporting links with epoxy cement.

"I prefer jaws made of sections cut from a hacksaw blade. I first grind off the saw teeth with a carborundum wheel. The opposite edges are polished to remove surface irregularities. The procedure is not difficult. I grind two four-inch slabs of plate glass together with a thin slurry of No. 120 carborundum grit in water for a period of six minutes, making elliptical strokes about an inch long and turning the 'sandwich' over every minute. The edges of the blade are ground against the frosted side of either of the glass pieces for two minutes, again with elliptical strokes. I examine the edges for pits and hills and, if necessary, continue grinding until they are straight and smooth.

"One of the completed jaws is soldered or cemented with epoxy to its supporting linkage. After the jaw has been fastened the linkage is moved to the position where the separation of the side links is at a minimum. The ground edge of the companion jaw is placed in full contact with the ground edge of the jaw previously installed and similarly attached to its supporting link.

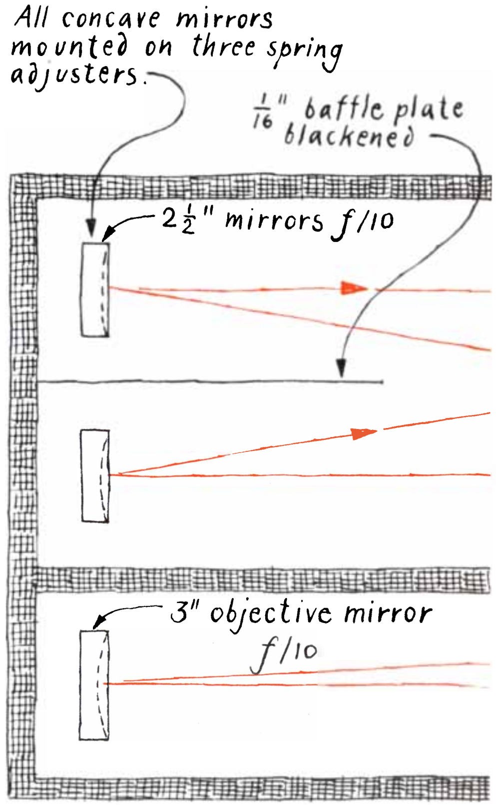
"All optical elements of the instrument must be installed in a lightproof housing. I improvised one of plywood. The spectroscope was made as a separate unit that could be bolted to the housing that supports the external grating and the objective mirror. This arrangement enables me to employ the instrument as a conventional laboratory spectroscope.

"The housings can take the form of simple boxes with removable lids to which the vernier dials are fixed. I minimized the overall dimensions of the apparatus by inserting a pair of plane front-surface mirrors between the objective mirror and the tuner to fold the incoming rays. The plane mirrors are set at an angle of approximately 90 degrees with respect to each other. They must be supported by mountings that can be adjusted a few degrees to align the optical path. Another plane mirror similarly mounted reflects converging rays from the spectroscope to the focal plane of the eyepiece. The interior of the spectroscope housing, including particularly the barrier that separates the concave mirrors, should be painted flat black to minimize the reflection of scattered light.

"All optical parts should be tested before they are installed. Testing the focal length of the mirrors is particularly important. A simple measurement of the focal length can be made by standing the mirror on edge, directing a flashlight toward the metallized surface and catching the reflected image of the lamp filament on a sheet of white paper placed next to the flashlight. Vary the distance of the flashlight and the paper from the mirror until the sharpest possible image of the lamp filament appears on the paper. The focal length of the mirror is equal to exactly half the distance between the image of the filament and the surface of the mirror. The focal length can be measured with greater accuracy by setting up the knife-edge test used for checking telescope mirrors [see *Amateur Telescope Making: Book One*, edited by Albert G. Ingalls, Scientific American, Inc.].

"An instrument similar to mine can be built for less than $250. The cost will increase exponentially with the diameter of the optical parts. The two-inch gratings and mirrors have enabled me to make most of the observations I had in mind when I undertook the construction and also to do a variety of laboratory experiments.

"The instrument has a maximum dispersion of about two angstroms per millimeter, which is equivalent to displaying the rainbow as an image more than six feet wide at the focal plane of the eyepiece. The zone in which the promi-



© 1974 SCIENTIFIC AMERICAN, INC

nences can be viewed at the edge of the sun is less than one millimeter wide, but even so it is sufficient to enable the observer to isolate the 6,562.8-angstrom spectral line of hydrogen.

"When the apparatus has been attached to an equatorial mounting that includes a clock drive to keep the ruled surface of the grating pointed approximately toward the sun, set the tuning plate at a right angle to the optical path and start the clock drive. Cover the objective mirror with a disk of white paper and adjust the external grating to the angle at which reddish light from the sun falls on the paper. Transfer the paper to the position of the slit of the spectroscope and adjust the flat mirrors to angles such that the reddish image falls on the paper at the position of the slit.

"Remove the paper. Adjust the first mirror of the spectroscope (the collimating mirror) to the angle at which the now parallel rays of the reddish light flood the diffraction grating of the spectroscope. If the light is difficult to see, cover the grating with a small sheet of white paper. Adjust the second mirror of the spectroscope to project converging rays to the flat mirror adjacent to the eyepiece. A sheet of paper inserted in the focal plane of the eyepiece should now display an image of the sun.

"While viewing through the eyepiece adjust the angle of both gratings to center the dark 6,562.8-angstrom spectral line on the slit of the spectroscope. The line is the darkest and broadest one in this region of the spectrum. The observer should now see in the eyepiece the complete scarlet image of the sun.

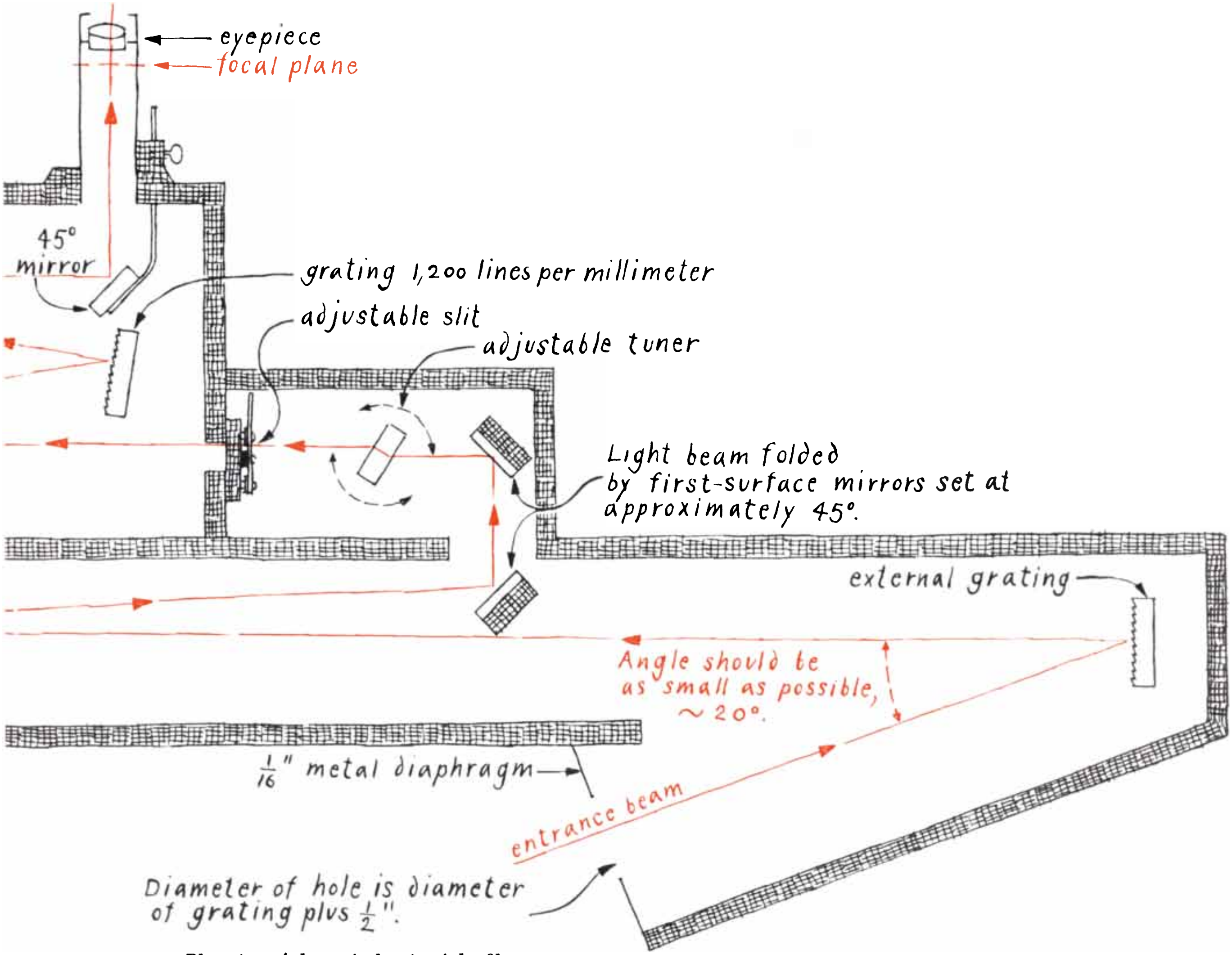
"To observe the limb, rotate the tuner so that the image appears to shift along the absorption spectrum to the point at which the edge of the sun is centered on the line at 6,562.8 angstroms. That is the adjustment at which Janssen made his initial observation. If a prominence happens to be located at this point, the observer will see the absorption line fade and be replaced by a bright area.

"Only a small area of the sun's edge is being observed. For this reason the edge will appear to curve away from the straight slit. To see prominences along a substantial portion of the edge the straight slit must be replaced by one that matches the curvature of the image.

"Curved slits of two kinds are relatively easy to make. Of the two I prefer one that requires the use of an engine lathe to cut a disk of metal equal in diameter to the diameter of the solar image at the focal plane of the objective mirror. The dimension is very nearly equal to the focal length of the objective divided by 114.

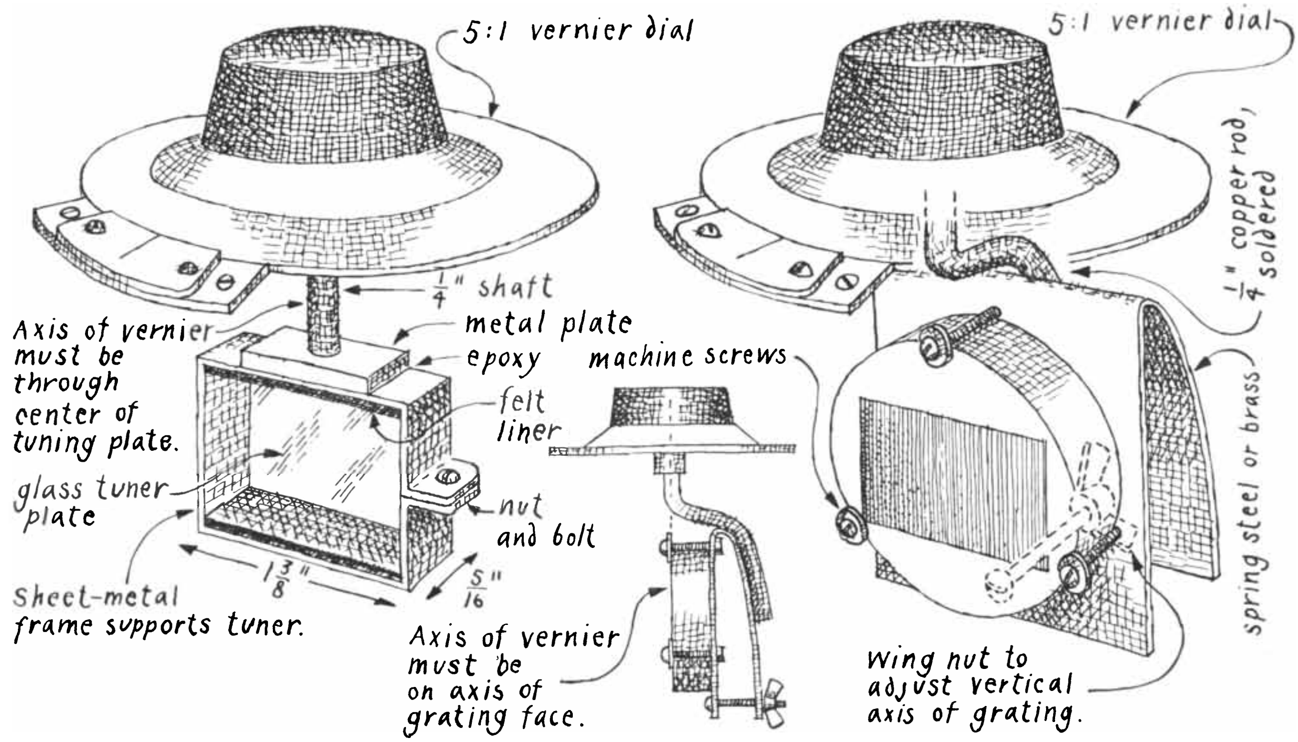
"In the case of an objective mirror with a focal length of 30 inches the radius of the curved slit is approximately .131 inch. A hole .05 inch larger in radius is drilled in a metal sheet. The device is assembled by centering the disk in the hole and tacking it in place with a dab of solder or epoxy cement to leave a clear arc .05 inch wide and extending about 180 degrees.

"Another technique for making the curved slit is easier. Drill a hole slightly larger in diameter than the solar image and place it over a mirror that has been



Plan view of the optical train of the filter

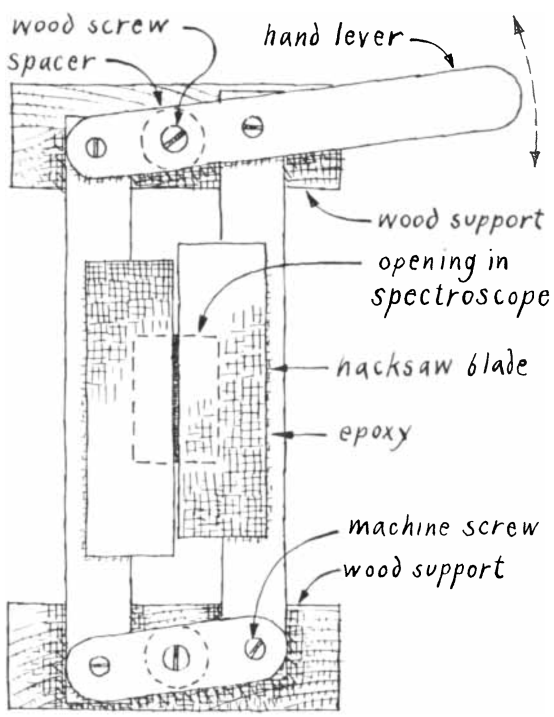
© 1974 SCIENTIFIC AMERICAN, INC



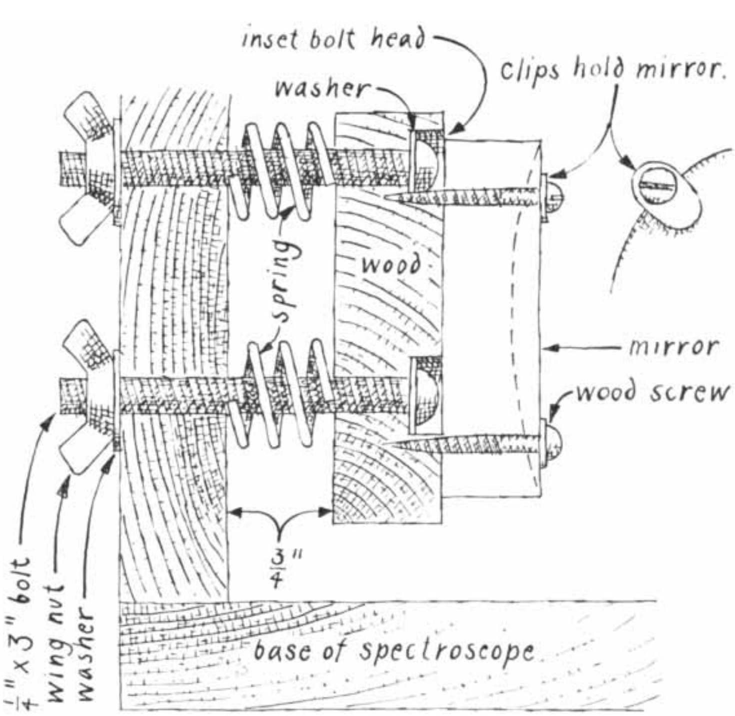
Vernier mounting of the tuning plate (left) *and the diffraction gratings* (right)

aluminized or silvered. Dip the sharpened tip of a wood toothpick in dilute nitric acid. Shake excess acid from the wood. Insert the sharpened tip through the hole in the metal and, with the metal as a guide, trace a semicircle on the coated glass. If the operation is performed with care, the acid will etch a usable slit in the metallic film. To observe solar prominences substitute the curved slit for the straight one.

"In a sense the double-grating filter is analogous to the tuning system of a radio set. It enables the investigator to select for observation a narrow band of light waves much as the dial of a radio set tunes in a narrow interval of the radio spectrum. The filter can serve in a variety of experiments other than the observation of solar prominences. For example, my interests include the use of



Details of the slit mechanism



Adjustable cells for concave mirrors

© 1974 SCIENTIFIC AMERICAN, INC

polarimetry for investigating the characteristics of minerals. At various orientations the surface of a rock can be seen in different colors, which depend on the crystal structure of the specimen. By examining the rock with the tuned filter, with both gratings adjusted to an appropriate angle, distinct colors appear in various areas of the surface that characterize the specimen.

"Mineralogists also routinely dissolve bits of unknown rock in a bead of incandescent borax to identify its constituents by the characteristic colors of the resulting flame. Each chemical element radiates a unique set of wavelengths. With the double-grating filter the experimenter can observe and even photograph the distribution of elements in the glowing gases.

"The rulings of a grating are cut at an angle that optimizes the efficiency of the device as a reflector of light of specified wavelength at a specified angle with respect to the plane of the rulings. As I have mentioned, the angle at which the rulings are cut is called the blaze. The angles at which gratings reflect bundles of rays dispersed in the form of spectra are known as the spectral orders.

"As I have mentioned, the gratings of my instrument are blazed to reflect most of the incident light at 6,600 angstroms in the third order. In general dispersion increases with the spectral order at the cost of brightness. My gratings were selected primarily for viewing solar prominences. Hence they were blazed for maximum brightness of the deep red in the spectral order that resulted in a dispersion of two angstroms per millimeter at the focal plane of the eyepiece. People who design the instrument for experiments of other kinds such as flame spectroscopy would doubtless select gratings blazed for other colors in other spectral orders.

"All optical parts from which the instrument is made are available from the Edmund Scientific Co. (300 Edscorp Building, Barrington, N.J. 08007). Diffraction gratings of the exact kind in my instrument are distributed by Jarrell-Ash Division of the Fisher Scientific Co. (590 Lincoln Street, Waltham, Mass. 02154) and by Bausch & Lomb Inc. (526–68 Lomb Park, Rochester, N.Y. 14602).

"Two people have helped me with this study. Timothy O'Hover of the chemistry department of the University of Maryland provided a laboratory and Gary A. Frazier supplied inspiration and electronic testing equipment for the original studies. I am deeply grateful to them."

Introducing the WDDS-12 Pritchard Loudspeaker System

More confusion?

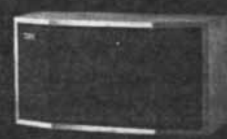
There are literally hundreds of speakers on the market. There are 2-way systems, 3-way systems, omnidirectional, rearward sound projection, et al. All, according to their makers, are outstanding. And, frankly, several are.

Small wonder the public is confused. And good reason why serious listeners rely on expert opinions, as well as recommendations from knowledgeable equipment owners.

Our new Pritchard system will surely add to the confusion. And our claims will fall, unfortunately, on many deaf ears.

But among those people who, over the years, have learned to separate fact from fiction...who trust their own ears...and who respect "inside" opinions, this new speaker will be a revelation.

If you're a serious listener, unfamiliar with ADC speakers, we urge you to talk to people familiar with our products. They know that, among the relatively few outstanding speaker systems on the market, ADC's line ranks among the best. And, now, with the introduction of the Pritchard System, ADC's probably number one!



For most, the confusion about speakers will continue. For some, the WDDS-12 Pritchard System will shout "hello."

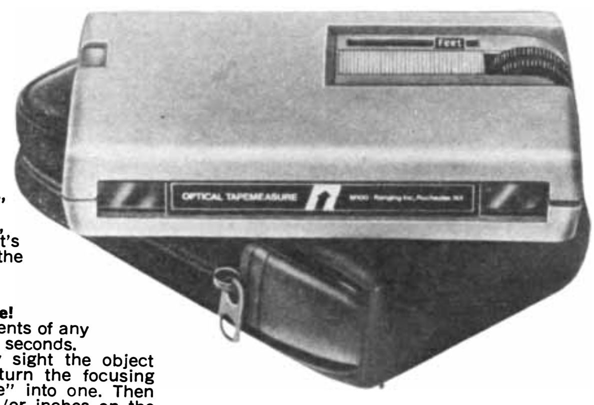
ADC speakers -the insider's choice.

AUDIO DYNAMICS CORPORATION
Pickett District Road, New Milford, Conn. 06776

Write for free WDDS-12 Pritchard Loudspeaker technical bulletin.

NOW...LET YOUR EYES DO THE MEASURING!

NEW *OPTICAL* TAPEMEASURE



NOW MEASURE DISTANCE WITHOUT MOVING AN INCH

... without help from another person ... without climbing or awkward maneuvering ... without worrying about ladders, obstacles, building materials, mud, snow ... without ribbons, tapes, chalk, strings or cords. It's done *optically*—if you can see the target you can *measure* it!

There's never before been an instrument like this—anywhere! Gives you precision measurements of any distance from 6 to 100 feet—in seconds. Measures height, too. Simply sight the object through the viewfinder and turn the focusing dial until two images "merge" into one. Then read the distance in feet and/or inches on the tape indicator. It's fast . . . dependable . . . accurate (99.2% accurate at 25 feet, 97.8% at 100 feet).

Ideal for contractors . . . painting, roofing, siding and wiring estimators . . . surveyors, road builders, real estate agents, construction firms, architects. Great gift for home handymen, gardeners, etc. Pocket-size, Optical Tapemeasure is 5½" x 3¼" x 1" and weighs less than 8 ounces.

THE *EVERYTHING* GUARANTEE
The Gallery guarantees *EVERYTHING:* the quality, accuracy of description, availability, prompt delivery. If not delighted, return Optical Tapemeasure within 3 weeks after you receive it (not the usual 10 days) for an instant refund or cancellation of charges.

Rugged, Cycolac® body; welded steel inner construction; shock-absorbing optical mounts; dust-proof interior. Warranteed for 5 years by the famous American optics manufacturer. Zippered carry case in brown doeskin vinyl. Gift-boxed.

Order today. Send check or money order, or charge your Credit Card. **Give Acct. No., Bank No. (if any), Expiration Date, Signature.** N.Y. residents add tax.

Order No. 41450: Optical Tapemeasure, $19.95 plus $1.50 shipping and handling.

Free Gift Catalog On Request

THE FINER THINGS BY MAIL GALLERY
Dept. 1379, Amsterdam, N.Y. 12010

© 1974 SCIENTIFIC AMERICAN, INC

Why wait?
All the great science books you want to read and own are yours now in

The Library of Science

(retail prices shown)

38011. CITIES AND GEOLOGY. *Robert F. Legget.* The first geological work on the special problems of urban areas. A must for scientists, engineers, planners, and architects. **$15.50**

50110. GENETICS OF THE EVOLUTIONARY PROCESS. *Theodosius Dobzhansky.* The definitive work on the modern synthetic theory of evolution by the world's foremost geneticist. **$12.50**

62801. MODERN COSMOLOGY. *D. W. Sciama.* "This book is a masterpiece . . . . It is almost unbelievably up to date."—*Physics Today* **$9.95**

52080. A GUIDE TO THE MEDICINAL PLANTS OF THE UNITED STATES. *Arnold and Connie Krochmal.* The definitive work on medicinal plants . . how to identify, collect, and even raise them. **$12.50**

79530. A SOURCE BOOK IN PHYSICS. *Edited by William Francis Magie.* Original papers in mechanics, sound, heat, magnetism, and electricity. **$15.00**

69621. POWER PLAY: Oil in the Middle East. *Leonard Mosley.* Behind-the-scenes look at the explosive oil situation in the Middle East, and crackling good history. **$10.00**

34560. ASTROPHYSICAL CONCEPTS. *Martin Harwit.* Challenging new presentation of the science that bridges chemistry, physics, biology, mathematics, astronomy, meteorology, and the earth sciences. **$14.95**

67661. THE PENGUIN DICTIONARY OF SCIENCE. *E. B. Uvarov and D. R. Chapman. Revised and Edited by Alan Isaacs.* Comprehensive reference tool for anyone involved in the biological or physical sciences. Many Illustrations. **$11.95**

63061. MOONS AND PLANETS. *William K. Hartmann.* Outstanding introduction to current knowledge of the solar system. A masterpiece of factual information. **$12.95**

56371. INTRODUCTION TO QUEUEING THEORY. *Robert B. Cooper.* Probably the most successful queueing theory text yet published. **$10.95**

84451. TOWARD A GENERAL SCIENCE OF VIABLE SYSTEMS. *Arthur S. Iberall.* Introduces a science that attempts to unify our understanding of all autonomous systems — natural, organic or man-made. An exceptional book. **$14.50**

Take any 3
(values to $50.95)
all for only $3.95
(if you will join now and agree to take only 3 more books at handsome discounts in the next 12 months)

73270. RACE. *John R. Baker.* A major scientific work on an explosive subject. **$15.00**

63910. THE NATURE AND FUNCTION OF SCIENTIFIC THEORIES. *Edited by Robert G. Colodny.* Six scholars analyze the contemporary status of scientific theories. A striking and original contribution to the philosophy of science. **$12.95**

87540. WORKING WITH NATURE. *John W. Brainerd.* Offers perceptive instructions for modifying the environment anywhere. **$15.00**

79470. A SOURCE BOOK IN MATHEMATICS. *D. J. Struik.* Big, 472-page volume of original selections from the mathematical thought and writings of authors in the Latin world between the 13th and the 19th centuries. **$11.95**

45990. ESSAYS ON CELLULAR AUTOMATA. *Edited by Arthur W. Burks.* Eight experts in the field present a fascinating overview of one of the fastest-growing and most intriguing branches of computer science. **$12.50**

56530. ISLAND AT THE CENTER OF THE WORLD. *Father Sebastian Englert.* Easter Island's mysterious giant statues, the islanders' puzzling past, their lost language. Lavishly illustrated in full color. **$15.00**

44350. ELEMENTS OF CARTOGRAPHY: Third Edition. *Arthur H. Robinson and Randall D. Sale.* Includes newest techniques of air-mapping, and the most recent experiments in design and reproduction. **$12.75**

73451. THE RAW AND THE COOKED / FROM HONEY TO ASHES. *By Claude Lévi-Strauss.* Brilliant anthropologist's trail-breaking two-volume exploration of the basis and interrelations of myths in primitive man. *Counts as 2 of your 3 books.* **$26.00**

61210. MATHEMATICAL THOUGHT FROM ANCIENT TO MODERN TIMES. *Morris Kline.* An enormous (over 1,200 pages) and hugely entertaining history, from ancient beginnings to contemporary concepts. *Counts as 2 of your 3 books.* **$35.00**

64910. NUMBER WORDS AND NUMBER SYMBOLS. *Karl Menninger.* Oversized, handsomely illustrated, 480-page volume filled with unusual lore and amazing insights. **$14.95**

49350. FUNDAMENTALS OF PHYSICS. *David Halliday and Robert Resnick.* A complete updating and revision of this enormously successful modern classic. **$15.95**

54140. THE HUMAN SENSES. *Frank A. Geldard.* Updated second edition. Combines the most accurate experimental studies with examples taken from everyday human experience. **$12.95**

34630. ASIMOV'S GUIDE TO SCIENCE. *Isaac Asimov.* An encyclopedic work of 945 pages and pure pleasure to read. **$15.00**

82640. TECHNOLOGY IN THE ANCIENT WORLD. *Henry Hodges.* A much-needed technical survey of the tools and processes of early civilizations. Delightful and useful. **$10.00**

78190. SEXUAL SELECTION AND THE DESCENT OF MAN. *Edited by Bernard Campbell.* Eleven experts in the fields of biology, genetics and anthropology examine the evidence for Darwin's theory of sexual selection. **$14.75**

45471. ENVIRONMENTAL GEOSCIENCE. *Arthur N. Strahler and Alan H. Strahler.* Masterful exploration of the physical and biological processes involved in interactions between Man and the environment of energy and matter which surrounds him. **$12.95**

If card is missing, write:
The Library of Science
Dept. 2-35A, Riverside, N.J. 08075
and indicate 3 books you want.

© 1974 SCIENTIFIC AMERICAN, INC

BOOKS

Military construction as a theme in modern architecture, and the computer as Kon-Tiki

by Philip Morrison

THE ARCHITECTURE OF WAR, by Keith Mallory and Arvid Ottar. Pantheon Books ($15; paperbound, $6.95). The Great Wall of China, the hilltop villages of Umbria, the city walls of Cracow and Bologna are all monuments of architecture erected out of military need. Our century has seen more war than any, particularly across the continent of Europe; Mars has been served by 20th-century architecture on a scale of gross effort and headlong intensity to which even Shih Huang Ti could not have condescended. This book documents that history in text, plan and photograph. Beauty and grandeur remain somewhere off page for us; the waste and the sorrow glow too hot to be concealed yet, even under the 5.5-meter reinforced-concrete slab over the submarine pens at Doenitz' headquarters port of Brest. (The R.A.F. hit those pens with thousands of tons of high explosive but did "surprisingly little damage.")

Permanent fortifications shelter towns and cities; field fortifications conceal and house armies in the battlefield; the "siege train," which once boasted heavy ladders and towers, counters the strongpoints of the enemy. All three classical divisions of military architecture remain in our times. Around the ancient fortified town of Verdun-sur-Meuse, France and Germany waged a legendary struggle in 1915 and 1916 that consumed the lives of 700,000 young men. Those forts were isolated underground strongpoints of reinforced concrete, fitted with steel cupolas for the turning guns. They had been moved out to ring the city at a distance as cannon range increased; they had to be completely enclosed, not merely walled, because now shells could be given high trajectories; they were made of hardened steel and meters-thick sandwiches of concrete and earth because armor-piercing and explosive shells can penetrate anything less. Thus the elegant urban polygonal masonry bastions of the Sun King had been transmuted into grassy suburban mounds, here and there frowning with turrets and deep-cut portals. "By the end of 1915 Verdun was drained of men and armaments"; Marshal Joffre did not trust his fortresses once he had seen Belgian cupolas crumble under the big German shells. The Germans seized one of the strongest Verdun forts without firing a shot. Now the capture and the defense of Verdun had become a great issue of politics. Entire armies died there, but the garrisoned French forts held. Were they not inscribed "Better to be buried beneath the ruins of this fort than to surrender"?

German-occupied Fort Douaumont withstood 120,000 shells, mostly French, with little damage. "The underground cover remained intact, not one field gun turret was destroyed, and hardly any of the casemates [were] rendered unoccupied." On the mistaken notion that it was such fortresses and not the bloodied infantry that had saved the honor of France at Verdun, a great fortress line was built, resembling a buried armored fleet deployed along 100 miles of border. It was ignominiously outflanked in one month's campaign in 1940 by the panzers rolling through the "impenetrable forest" of Ardennes (which was not fortified). The man whose name is given to this remarkable construction (still in place, overgrown and desolate) was Minister of War André Maginot, who had himself been wounded at Verdun.

Total war meant a new scale of fortification. The Atlantic Wall, the Maginot Line of Albert Speer and Adolf Hitler, stretched 1,400 kilometers along the coast, a bastion for Fortress Europe, its formal architectural military construction built by half a million enslaved men over two years. There were 9,300 fortified works by June of 1944; they were of course not dense everywhere but rather were "so many knots in a piece of string." These towers, pillboxes and casemates screened all the ports of Atlantic Europe from attack by sea. They were bypassed, Speer has written, within two weeks by "a single brilliant technical idea."

That idea was a modern siege train on a prodigious scale. The scaling ladders and fighting towers of the past became two artificial harbors built in Britain and floated to the beaches of Normandy. They were complex constructions, with concrete caissons sunk for solid breakwaters, flexible breakwaters, floating piers, legged pierheads, an entire technology to create square miles of smooth water along that storm-beaten low coast. The worst Channel storm in 40 years wrecked the American-manned harbor entirely. The British port worked. Yet across the open beaches of the American landing place improvisation and energy moved past the wreckage of the big piers and breakwaters twice the tonnage that flowed through the working artificial harbor. Planning and reality sit uneasily together in war; that lesson ought to be written very clearly in the margins of current expert schemes for deterrence, such as ABM, MIRV and Trident.

In the estuary of the Thames the British built advanced radar and antiaircraft emplacements. These sea forts were dual. The Red Sands Fort, closer to shore, was built and manned by the army. Here it is: six big two-story steel houses, each on a high tripod above the sea, each open deck roof bearing a gun battery or a radar dome. It is a little fortified village on stilts, the houses joined by spidery walkways. The navy went farther out to sea. The naval fort was a fixed ship. A concrete barge was flooded and set on the bottom. From it rose two concrete cylinders, largely under water. There the crew lived, with the magazines and storerooms below. The cylinders were covered with the fighting deck of a warship: guns, fighting top, officers' quarters—and whaleboats! Here is a lesson in the pervasive nature of design tradition.

The total war meant mass housing and dehousing. The temporary structures are here too, for example the British Nissen hut. That ubiquitous semi-

© 1974 SCIENTIFIC AMERICAN, INC

cylinder was born in 1916, designed by a Canadian engineer officer. It represented the first "real mass production of complete buildings as opposed to the mass production of components." A *Daily Mail* reporter wrote in 1917: The "creature... began to appear in the conquered areas...; by twos and threes, and tens and hundreds, its fellow-monsters would appear, so that in a week or two you would find a valley covered with them that had been nothing but pulverised earth before." Huge colonies of the same beasts, now known as Quonset huts, occupied America and the Pacific during World War II.

Light prefabs and massive concrete with unfinished surfaces are the architectural legacy of World War II. The authors, two architect graduates of the University of Bath, find that the civil "bunker style" is not easily attributed to the similar shapes and textures of the German Atlantic Wall. It was the explicit contribution of Le Corbusier, who was tending toward it even before the war. It is in scale, materials and speed of construction that the longest-lasting influences are to be found. Nowadays, in a time of unlimited explosive power, it is light domes, buildings almost of air, and their opposites, underground silos and construction within mountains, that characterize military design. This is a volume eloquent in implication and meticulous in statement. It deserves and repays thoughtful reading and looking.

THE SETTLEMENT OF POLYNESIA: A COMPUTER SIMULATION, by Michael Levison, R. Gerard Ward and John W. Webb, with the assistance of Trevor I. Fenner and W. Alan Sentance. University of Minnesota Press ($10.75). Captain James Cook recognized that the Polynesian islands were occupied by what was essentially one nation, and he came to believe that "the inhabited islands in this Sea have been at first peopled" by small groups of castaways set adrift by accident and carried to a new island landfall. During the 19th century the legends of the islanders were carefully collected. These tell a different tale: the glorious ancestors had set out from home, discovered new lands and returned. At some later time their kinfolk followed clear sailing directions to colonize these new places. In 1956 Andrew Sharp, in a most persuasive book, returned to Cook's view: the old settlements were made by one-way voyagers, either by people blown off course or, less frequently, by forced exiles making a desperate voyage into the unknown that was occasionally happily successful. The argument continues hotly, and everyone knows that Thor Heyerdahl on his balsa raft *Kon-Tiki* tried to demonstrate that settlement moved not from the Asian shore but from the American.

In this volume we encounter five Heyerdahls of the computer. They voyage not across the heaving Pacific on a floating haystack but along magnetic tapes, fed into the Atlas computer at the University of London and plotted on ocean maps by a Stromberg-Carlson offline microfilm plotter. (The clouds of points that mark these 101,016 simulated drifts and 8,052 guided voyages are to a large extent reproduced here on 42 pages of maps with a plastic island overlay.) Their work presents no salty verisimilitude like that of the *Kon-Tiki,* but it provides what the real voyagers cannot: an appropriately large and varied sample. The investigators began by determining the winds and currents, five-degree square by five-degree square, across the wide Pacific, relying for these on the summaries of a century of observations by British ships plying the sea. For each square they listed the distribution of wind and current speeds and directions. They start a "canoe" on a random day of the year from a given island or zone and follow it day by day, letting the winds and the water move it until the voyage ends.

The successful voyage terminates with a landfall, for which are assigned sighting radii of 10 miles for low islands, 20 miles for high islands and 30 miles for coastlines. Most voyages are ended by exhaustion of water or food, however, according to an adopted table of risks (based on experience) that allows the longest voyage to be 26 weeks and the half-life of crews about 10 weeks. Sometimes a canoe is swamped at sea; the program gives an even chance of survival through each day in which the wind blows with Force 9 or greater! The wind and current data tables hold some 800,000 entries; these must be stored in tape memories, not the central one, and this fact dominates the program design. The programs are listed here, in a dialect of ALGOL.

The results of the study are fascinating. They are presented modestly, in a considerable and well-informed context that helps to explain the model and the entire history of the problem. It seems plausible that drifting vessels could have carried people long ago from the islands of Micronesia into westernmost Polynesia. Of a batch of about 800 random drifts out of the southern Gilberts nearly 22 percent ended in Polynesia. From the Americas, from Mexico south to Peru, it is most unlikely that drifts can reach Polynesia (unless they begin hundreds of miles off the coast of Peru). The whirl of the Equatorial Current caught every computer-run canoe; of 4,392 drifts originating off the coast, none reached Polynesia. That is surprising, since six real rafts have made the voyage since the pioneer effort of *Kon-Tiki* and one reached at least the inhospitable Galápagos. The reconciliation appears to be that the crews of these rafts made "as much westing as possible" to avoid being thrown right back onto the coast they had left. These voyages were in fact made by navigators trying to sail westward.

Even after island-hopping and some drifting had settled the west, the winds allowed no easy drifts into the core of eastern Polynesia. It seems possible that a single colonizing voyage carried the islanders to the still unknown Marquesas. They would have followed a long tradition of travelers in island-studded seas, searching for new land and confident in the belief that, as usual, "islands would rise over the horizon" to meet them; such navigators sail by the image that their canoe is a little stationary world past which islands drift.

The Marquesas once occupied, drift could people the rest of the tropical islands; the chances grow large. A voyage of intent, sailing due east from Samoa, able to hold its track only within 90 degrees of the wind, carries an eastbound canoe to a successful landfall in eastern Polynesia with a probability of some 20 to 30 percent (counting various possible terminal islands, not all of them suitable). The more remote islands, from Hawaii to Easter Island to New Zealand, are hard to reach by drift, but a southwest course held within 90 degrees of the varying wind reaches New Zealand 60 percent of the time from Raratonga, 800 miles northeast. It appears that some element of purpose was needed for man to reach the outer groups, but no unreasonable gambles. "There are good chances of successfully crossing all the major ocean stretches... with a very limited degree of navigational skill, within a reasonable survival period, and in craft which have poor capabilities of sailing to windward." Holding a course by the stars will suffice for navigation; it is not a specific landfall but any land that can support life that such pioneers seek.

What is missing is any estimate of what a workable probability really is. Perhaps a small chance of survival would

© 1974 SCIENTIFIC AMERICAN, INC

do, if many castaways were driven by sudden storms out of well-populated canoe routes among close islands. On the other hand, even an unlikely voyage might have happened just once, by foolhardy intent or through great luck and skill. *Kon-Tiki* was designed to prove possibilities; it certainly could establish no probabilities at all.

The debate will continue, the authors wisely say. They have given us a real piece of insight and a method, even if it is not final. The "accepted dogma" that presents the Maori myth of settlement begins with brave Kupe, who was chasing a huge octopus when he encountered the uninhabited coast of New Zealand 1,000 years ago. It continues with a mix of castaways, purposeful seamen, chance discoveries, navigated expeditions of settlement by "the Fleet" and even two-way voyages to fetch the sweet potato, news of which had come by castaway. The computer seems to agree at least in this: the true story is complex, and no single view will explain the whole of that marvelous dispersal of human beings over 10 million square miles of the ocean deeps.

THE DETECTION OF FISH, by David Cushing. Pergamon Press ($16.50). All anglers know that fish are mobile and migratory and congregate seasonally in certain favored spots. These may be spawning areas, feeding areas or temperature and current zones where the conditions induce concentration of fish. Particularly in zones where upwelling cold water from the depths, enriched in nutrients, can support a luxurious growth of plankton do the fish abound, and where fish abound there fishermen follow. Nearly 40 years ago C. H. Townsend published his study of the logbooks of the old Nantucket whalers. They had recorded their catches of 36,908 sperm whales between 1729 and 1919. The seasonal world maps of these catches mark out with astonishing completeness the main upwelling centers of the world ocean. The yellowfin-tuna fishery on the Pacific Equator, worked only recently by American and Japanese long-liners, is plainly disclosed in the Nantucketers' logbooks.

You could see the big whale blow, "and sparm at that." Fish, however, are seen only rarely, sometimes from the air, sometimes by a night glow, perhaps by satellite photographs of sun glint marking very calm water "to which fish may aggregate." Fishermen's signs have always been oil or turbid water, plants and plankton in the sea and, more gen-

To preserve your copies of SCIENTIFIC AMERICAN

¶ A choice of handsome and durable library files—or binders—for your copies of SCIENTIFIC AMERICAN.

¶ Both styles bound in dark green library fabric stamped in gold leaf.

¶ Files—and binders—prepacked for sale in pairs only.

Index for entire year in December issue. *

FILES *shown at left*
Hold 6 issues in each.

Single copies easily accessible.

Price: $4.00 for each pair (U.S.A. only).

Address your order, enclosing check or money order, to: *Department 6F*

BINDERS *shown at right*
Hold 6 issues in each.

Copies open flat.

Price: $4.50 for each pair (U.S.A. only).

Address your order, enclosing check or money order, to: *Department 6B*

New York City residents please add 7% sales tax
Other NYS residents please add 4% state sales tax plus local tax

*Supply of back copies of SCIENTIFIC AMERICAN is limited. A list of back issues still in stock is available upon request.

SCIENTIFIC AMERICAN, 415 Madison Ave., New York, N.Y. 10017

© 1974 SCIENTIFIC AMERICAN, INC

erally, the presence or absence of predators. To this day the pelicans usefully point out the anchoveta off Peru (when there are any) and the Japanese tuna men keep a sharp eye for birds, which can indicate tuna to a boat three miles off. Competing trawlermen are the sign most followed by trawlermen in the North Atlantic; their radio chatter and then careful direction-finding are cues to the canny skipper. In the calmness of the Java Sea the fishermen stick their heads underwater and listen for fish.

The recording echo sounder was first used for fishing about 40 years ago. This volume by an English fishery biologist is a clear and knowing summary of the echo detection of fish for three groups of specialists: the physicists and engineers who design the equipment and study sound in the sea, the fishermen who use the equipment and the fishery biologists who need to count and size fish quantitatively. The book tells the engineer about fisheries in general and tells the biologist about acoustics and methods of sampling, and it presents many echo traces and their interpretation.

The usual sonar pulse is a fairly narrow-band pip of sound at a frequency between some 10 and 200 kilocycles, lasting from tenths of milliseconds to a couple of milliseconds and emitted repeatedly in a 10- to 20-degree cone pointed toward the sea bottom. A nickel transducer at the lower frequencies or a piezoelectric ceramic transducer at the higher ones generates sound from an electrical pulse with a few kilowatts of peak power. The echo returns to be displayed on an oscilloscope screen and usually to make a paper record as well. A single fish will give a small distinctive hyperbolic trace; a shoal makes a mountain of a record. The shape is determined by the range change as the target crosses the cone of the beam. At high frequencies the sea noise is small but the absorption is high; the resolution is better, of course, and the sound cone is more easily made narrow. Low-frequency, high-powered sounders seem most suited to deep water and to big fishes; the other option works for smaller fishes in shallower depths. At 500 kilocycles the sea becomes so absorbing that a practical limit is reached.

Can fish be counted and measured reliably by their echoes? The task is complex. The air within the swim bladder causes a reflection much louder than the flesh of the fish, which is acoustically not very different from seawater; bone is intermediate. Half of the signal may come from the swim bladder, which has only a twentieth the volume of the fish. Most tunas, sharks and flatfishes have no swim bladders, so that they show up somewhat weakly. Resonance in that air bubble is usually at low frequency; it seems hard to exploit.

New methods are in use. There are side-scan sonars that show an entire seascape. The fishermen have not used this mode anywhere nearly as much as the geologists ("one of the mysteries of present-day fisheries research"). There are simple transducers that can keep a 3-D eye on the opening of the trawl itself. Frequency-modulation sonars, which are continuous rather than pulsed, have been developed from mine-hunting gear. They scan and average rapidly, and they can exploit a Doppler signal to measure speed. Now the sector scanner, in which many receiving channels analyze the echo to provide directional detail, bids to become the most important new tool. Here you can see the shape of shoals of fish and a fish's path into the trawl net itself; a fish count becomes direct. The traces reproduced here more or less resemble high-contrast underwater photographs, although with plenty of noise. There is work ahead for fishermen and biologists alike.

A fascinating chapter is devoted to unraveling the famous mystery of the deep-scattering layer. During World War II American and Japanese sonar experts accumulated extensive records of an echo-producing layer observed in all the oceans. It was about 75 meters thick and lay at a depth of between 200 and 500 meters but migrated upward at night; it could extend over hundreds of miles. By the 1970's it was clear that the layer consisted of targets whose cross section was like that of small fish rather thinly distributed. Frequency responses indicated a variation in cross section with depth, and the change could be fitted to the changing size of swim bladders with depth.

The biologists have supported the acoustic work. The depth of the layer has been correlated with light intensity, although bioluminescence complicates the study. Capturing the animals has not been easy. Animals of the plankton are what you catch when your net is in the layer, but the signals come from the rarer predators that feed on them. The layer is almost surely made up mostly of finger-length lantern fish, with some larger but fragile pelagic jellyfish. There is no great protein reserve in this worldwide false bottom; a very large trawl towed for one hour in the layer by a research vessel caught thousands of the fishes, all of which could be "poured into two buckets." The layer interposes one more link in the food chain between the plankton and the delectable tuna.

AFRICA COUNTS: NUMBER AND PATTERN IN AFRICAN CULTURE, by Claudia Zaslavsky. Prindle, Weber & Schmidt, Inc., Boston, Mass. ($12.50). The trading centers of the Yoruba in southwestern Nigeria have been vigorous cities since medieval times, and in them reading and writing were long the property of scholars (in Arabic) even more exclusively than learning once was (in Latin) in Europe. Arithmetic, on the other hand, spreads a very wide net among the farming Yoruba, since they are a market people. Their number system is individual and illuminating. It is based on 20, although with much use of doubling of five and 10. The first 10 digits have their own names and the teens use 10 to form a compound, just as we do. Fifteen to 20, however, are reckoned as 20 less five, four, three, two and one. Thirty has the next distinct name, but each multiple of 20 is given only a compound name, as in "six 20's." All the intermediate decades—50, 70, 90 and so on up to 190—are named by subtraction; for example, $(20 \times 6) - 10 - 4$ is 106. Two hundred has a name of its own; beyond that, up to 2,000, the general pattern is algebraic, albeit irregular; for example, 525 is $(200 \times 3) - (20 \times 4) + 5$. An exception is 20×20; "the elephant of figures," *Irinwo,* it is the highest number with a name. The rest are made by compounds and multiples: 40,000 is described as ten 2,000's twice, a million as a thousand 1,000's, and the Yoruba reckoner still has plenty of numbers left on his way to infinity.

This strongly subtractive system, reminding the reader of such Roman numerals as IX, may take its origin from mechanical tasks of reckoning—as the Roman system did. The Romans used an abacus or a counting board with tokens, and it is handier to take away one counter than to add nine, or even four. Cowries were the ancient coinage of the Yoruba; inflated by trade with the maritime Europeans, who had access to Zanzibar's tonnage sources of these small tropical seashells, cowries became numerous. Two bags of 20,000 shells each bought a man in the Lagos slave trade 150 years ago. There are careful witnesses of professional cowrie-counting, in which experts swiftly and repeatedly drew groups of five from a big heap. Again it is easier to subtract from a higher-multiple pile of 20 and of

© 1974 SCIENTIFIC AMERICAN, INC

200 in order to minimize hand motions. All this arithmetic is current mental practice (cowries are, to be sure, no longer the common currency), although decimalization is spreading and elision, contraction and euphony modify the systematic principles—all phenomena that are characteristic of the fast, witty style of the Yoruba tongue.

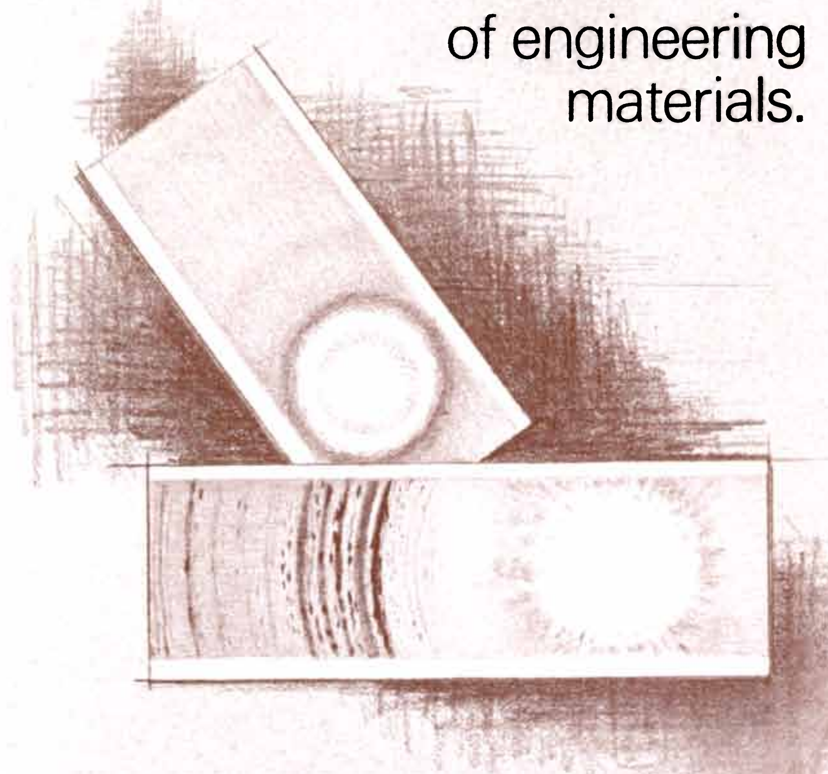
The famous university at Timbuktu on the Niger reached a considerable level of scholarship over the centuries, but mathematics was always subordinate there to theology, law, grammar and history. The Hausa city of Katsina became a center of learning in the 18th century and sheltered at least one numerologist, Muhammad ibn Muhammad. His 1732 manuscript on magic squares is excerpted here in a lively chapter full of schemes for the devotee, who can reconstruct the old magical arrays up to seven-by-seven cells.

This unusual volume, typographically most attractive and illustrated with many photographs, drawings and diagrams, is the work of a teacher and applied mathematician who has spent years at the task. She has used the older literature with suitable care and has interviewed Africans on the scene and in the U.S. Her book is a "preliminary survey" of mathematics in traditional Africa, citing something of the mathematical culture of more than 50 sub-Saharan peoples. It cuts a wide swath: counting games, everyday counting, number magic and taboo, weights and measures, number gestures and signs, timekeeping, even the group theory of artistic pattern. All 24 possible plane and strip patterns—the one- and two-dimensional crystal classes—appear in the art of Islam. The designs found on the famous raffia pile cloth and also in other art objects of the Bakuba, in the Congo, display 19 of these possibilities. (One group never occurs, although it is a type of design used by their near neighbors, the Bayomba people.)

The general reader will find this book smooth going, and yet it is valuable as a working reference. It gives plenty of detail and careful citations, although it is not for the most part a primary ethnographic source. On the other hand, the professional Africanists have not provided any such display of mathematical interest since the 1915 Vienna monograph of Marianne Schmidl, which is one major source. For a long time, moreover, the thoughtless and the self-serving tried to place African peoples in a patronizing light. Mathematical culture, like any other, reflects mainly social and

MATERIALS
RESEARCH
CENTER
REPORTS . . .

On a new class of engineering materials.



A Materials Research Center (MRC) group working in the field of structural materials has discovered alloys and developed processes for the practical production of a new class of high-strength metallic glasses. They are the forerunners of an advance in metallurgical technology of considerable promise.

METGLAS™ alloys consist of such base metals as Fe and Ni combined with various metalloids (B, C, Si, P, etc.). The basic fabrication process, although varying in detail with the base-metalloid combination, consists of extremely rapid quenching from the liquid state. In this way the random atomic structure of the liquid phase is retained in the solid state at ordinary temperatures. Thus, unlike conventional metallic alloys, METGLAS materials are not aggregates of small crystals and therefore do not have the "grain boundaries" that tend to weaken conventional alloys and make them susceptible to chemical attack. On the other hand, and unlike ordinary glasses, the new materials are not brittle and they are good electrical conductors.

The unique atomic structure of METGLAS alloys gives them interesting properties and combinations of properties. Thus, they exhibit exceptional strength combined with good ductility, superior corrosion resistance, high hardness, low acoustic attenuation. Also, a variety of electrical and magnetic properties can be readily obtained to meet electromagnetic requirements.

The metallurgy of metallic glasses has been extended by the MRC group through the development of economical, high-speed methods for spinning wires, ribbons and strips.

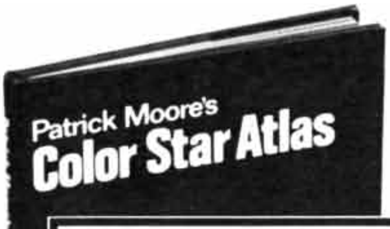
Work on both compositions and processing is continuing.

Allied Chemical Corporation/Materials Research Center
P. O. Box 1021R • Morristown, New Jersey 07960.

™Trademark of Allied Chemical Corporation.

Allied Chemical
MATERIALS
RESEARCH
CENTER

© 1974 SCIENTIFIC AMERICAN, INC



A beautifully illustrated "Open Sesame" to the heavens

■ The new wonders of astronomy in the space age are put into terms anyone can understand in this observer's guide to the major stars and constellations—fully illustrated with more than 130 diagrams and photos in full color and black and white, 16 pages of six-color star maps covering the entire universe, and 16 pages of constellation charts printed in colors that remain clear by day or night. Size 7" x 10½".

$7.95, now at your bookstore, or send check or money order to CROWN PUBLISHERS, 419 Park Ave. South, N.Y., N.Y. 10016

Crown

26 top specialists show all the ways science is giving nature another chance.

NIGEL CALDER, Editor

NATURE IN THE ROUND

A GUIDE TO ENVIRONMENTAL SCIENCE

A collection of original essays about our current ecological knowledge and the role of traditional sciences in keeping our planet green and alive.
$8.95

THE VIKING PRESS
625 Madison Avenue • New York, N.Y 10022

economic realities, not aberrant structures of the mind. Francis Galton told of the nomads who failed at the game of trading four sticks of tobacco for two sheep (the rate having been set at two sticks for one sheep) until they could carry out the trade twice, one sheep at a time. Yet they knew their flocks with precision, recognizing every animal individually.

Africa counts, everywhere, with the style and the intensity suited to its needs. Even the Bushman's "one, two, many" sequence, a long way from the West African marketplace, does not end Bushman counting: "They assign a specific meaning to the word 'many' by means of appropriate gestures."

Teachers and students of school mathematics are the explicit audience this book aims at. They should find it both entertaining and instructive, but it would have been a pity to leave it to them alone.

STEREOCHEMISTRY, by G. Natta and M. Farina. Translated by A. Dempster. Harper & Row, Publishers ($6.95). The Italian edition of this interesting introduction bore the subtitle: "Molecules in 3-D." Most scientific readers, even those far from chemistry, know the broad story of how chemists long ago found, from the logic of multiple compounds with the same atomic composition, that organic substances had to be regarded as small structures in space, well represented by the ball-and-stick models so widely familiar. Even the complex helixes of life, proteins and nucleic acids, appear to us as such carefully fitted tinker toys. In August Kekulé's daydream the six carbon atoms first joined into a ring, like a snake swallowing its tail, and the chemistry of rings was born a century ago.

The world is not so static as such models. Classical stereochemistry described the relative direction of bonds, the handedness of structures that differ from their mirror image (it all began with Pasteur's separation of mirror-image tartrates), the places of attachment of the several side chains and other details represented well enough by nearly rigid models. It is clear that molecules often do not lie meekly in the plane of a paper or blackboard diagram. Double electron bonds may, as in the benzene ring, have no fixed localization but require description by a more probabilistic scheme reflecting their quantum origin. Of course, molecules move on their erratic paths as they collide and spin. More than that, they can rotate internally, not freely but with motions "like a ball rolling over uneven ground."

Consider the molecule ethane, "the keystone of modern stereochemistry." It consists of two carbon atoms bonded to each other and to three hydrogens each: two tetrahedrons fused. These two tetrahedrons rotate slowly around the carbon-carbon bond that links them. In a short-exposure photograph—say 10 nanoseconds—an end view of the molecule would show the hydrogens usually staggered, but from time to time they turn into the lined-up position—eclipsed, as it is called. The staggered state is more stable but only by a little; at normal temperatures the molecule spends some time in all positions.

These several forms are called conformations; conformational analysis is the completion of the old program for mapping molecules in space. Ethane is only the simplest example; multiple conformations, with varying rates of transition among them, have for a decade or two much increased the detail with which we need to study molecules. X-ray diffraction, infrared spectroscopy and nuclear magnetic resonance are all now essential tools for the organic chemist, but reaction logic and delicate analyses of his products (by gas chromatography, for example) are still his master keys. By now it is clear, for example, that beta-glucose, the building block of cellulose and "the organic compound most commonly found on earth," is a chairlike puckered ring of six carbon atoms, whose substituents all lie roughly in its equatorial plane. It is therefore in three-space the most stable of all the sugars. The apparent disorder seen in its old plane structural formula was only a "crude representation of a more profound rationality."

Giulio Natta, Nobel laureate, is the director of the famous industrial-chemistry laboratory at the Milan Polytechnic, and Mario Farina is his long-time collaborator. Their own work is a high point of this genuinely interesting, understandable book for students at the college level who are specialists not in organic chemistry but in biology or physics. The Milan laboratory first prepared, about a decade ago, synthetic polymers with highly specific conformations. They were of crystalline regularity. The front cover of the book shows an end-on view of a model of one of these compounds, polypropylene, whose long strings of monomers are laid down so precisely that all the methyl groups lie on the outside, carefully lined up in three complete lines, the scheme repeating every three

© 1974 SCIENTIFIC AMERICAN, INC

units around the helix. The subtle stabilizing effect of the growing chain itself on the conformation next attached may be the main cause of this regularity, with the stable states sensitive to conditions of solvent or temperature. There is also an effect of the insoluble catalytic surface that is usually employed, which may allow only the right spatial orientation for each successively added group. These effects may coexist; we cannot yet assess their relative importance. A specificity greater than 99 percent sometimes results, higher than in the great majority of simple organic reactions, reminiscent of genetic specificity.

WATER ATLAS OF THE UNITED STATES, by James J. Geraghty, David W. Miller, Frits van der Leeden and Fred L. Troise. Water Information Center, Inc., Port Washington, N.Y. ($35). A novel atlas: here are 86 maps of the coterminous U.S., each done on the 7.5-by-12-inch outline base map of the Coast and Geodetic Survey, with a text for each, plus 36 maps of smaller size showing Hawaii and Alaska. All the maps are about water. Rain, thunderstorms, hail, fog, snow and glaciers are mapped to start; then such processes and variations as frost, evaporation and air and water temperature. The trunk of the blue annual-runoff tree is the Mississippi; the Columbia and the St. Lawrence are its only heavy branches. (The Connecticut and the Rio Grande are not even listed, being down by a factor of more than 25.) Waterways, ground water, flood losses, dams and deep-well disposal laws are mapped, taking into account human intervention. Saline waters and desalting plants, hard water and soft, natural fluorides and 30-odd maps on pollution: might these be called nonwater maps?

There is a familiar map of blue dots for population and one of brown dots for fish kills. (The worst fish kill listed was near Plant City, Fla., in 1969. The dumping of food products reduced the dissolved oxygen in Lake Thonotosassa and 26 million fish died in nine days.) "Hail Alley" is where Colorado, Wyoming and Nebraska meet; an average year there brings hail on eight or nine days. Irrigation in California is the largest single use of water by state and purpose. Many of the maps are rather simple, merely categorizing entire states or even regions, and would be just as clear on a much smaller scale, but the atlas makes much lore accessible and is full of interest, although it is high-priced. The two senior authors are consulting specialists in ground-water studies.

INDEX OF ADVERTISERS

MARCH 1974

© 1974 SCIENTIFIC AMERICAN, INC

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE GASIFICATION OF COAL

CHEMISTRY OF COAL UTILIZATION: SUPPLEMENTARY VOLUME. Edited by H. H. Lowry. John Wiley & Sons, Inc., 1963.

EVALUATION OF COAL GASIFICATION TECHNOLOGY: PART 1—PIPELINE QUALITY GAS. R & D Report No. 74, Interim Report No. 1. Office of Coal Research of the U.S. Department of the Interior, April, 1973.

A DYNAMIC MODEL OF CELL MEMBRANES

THE FLUID MOSAIC MODEL OF THE STRUCTURE OF CELL MEMBRANES. S. Jonathan Singer and Garth Nicolson in *Science,* Vol. 175, No. 4023, pages 720–731; February 18, 1972.

THE LOW POLARITY OF MANY MEMBRANE PROTEINS. Roderick A. Capaldi and Garret Vanderkooi in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 69, No. 4, pages 930–932; April, 1972.

EVIDENCE FOR BOUNDARY LIPID IN MEMBRANES. Patricia C. Jost, O. Hayes Griffith, Roderick A. Capaldi and Garret Vanderkooi in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 70, No. 2, pages 480–484; February, 1973.

MEMBRANE STRUCTURE: SOME GENERAL PRINCIPLES. Mark S. Bretscher in *Science,* Vol. 181, No. 4100, pages 622–629; August 17, 1973.

THE STRUCTURE OF MITOCHONDRIAL MEMBRANES. Roderick A. Capaldi in *Mammalian Cell Membranes: Vol. II,* edited by Graham Jamieson and David Robinson. Butterworth's Publishers, London, 1974.

THE NEURAL BASIS OF VISUALLY GUIDED BEHAVIOR

WHAT THE FROG'S EYE TELLS THE FROG'S BRAIN. J. Y. Lettvin, H. R. Maturana, W. S. McCulloch and W. H. Pitts in *Proceedings of the Institute of Radio Engineers,* Vol. 47, No. 11, pages 1940–1951; November, 1959.

A QUANTITATIVE ANALYSIS OF MOVEMENT DETECTING NEURONS IN THE FROG'S RETINA. O.-J. Grüsser, U. Grüsser-Cornehls, D. Finkelstein, V. Henn, M. Patutschnik and E. Butenandt in *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere,* Vol. 293, pages 100–106; 1967.

NEUROPHYSIOLOGIE DES BEWEGUNGSSEHENS. O.-J. Grüsser and U. Grüsser-Cornehls in *Ergebnisse der Physiologie,* Vol. 61, pages 178–265; 1969.

NEURAL MECHANISMS OF PREY-CATCHING AND AVOIDANCE BEHAVIOR IN THE TOAD (BUFO BUFO L.). J.-P. Ewert in *Subcortical Visual Systems,* edited by D. Ingle. S. Karger, 1970.

LOKALISATION UND IDENTIFIKATION IM VISUELLEN SYSTEM DER WIRBELTIERE. J.-P. Ewert in *Fortschritte der Zoologie,* Vol. 21, pages 307–333; 1973.

THE CHEMISTRY OF THE SOLAR SYSTEM

AN INTRODUCTION TO PLANETARY PHYSICS: THE TERRESTRIAL PLANETS. W. M. Kaula. John Wiley & Sons, Inc., 1968.

MOONS AND PLANETS: AN INTRODUCTION TO PLANETARY SCIENCE. William K. Hartmann. Bogden & Quigley, Inc., Publishers, 1972.

SPACE SCIENCE REVIEWS, Vol. 14; 1973.

INORGANIC POLYMERS

INORGANIC POLYMERS. Edited by Francis G. Stone and W. A. Graham. Academic Press, 1962.

PHOSPHONITRILIC COMPOUNDS: VI, HIGH MOLECULAR WEIGHT POLY(ALKOXY- AND ARYLOXYPHOSPHAZENES). H. R. Allcock, R. L. Kugel and K. J. Valan in *Inorganic Chemistry,* Vol. 5, No. 10, pages 1709–1715; October, 1966.

RING-CHAIN EQUILIBRIA. H. R. Allcock in *Journal of Macromolecular Science: Reviews in Macromolecular Chemistry,* Vol. C4, Part 2, pages 149–189; 1970.

INORGANIC RINGS AND CAGES. D. A. Armitage. Edward Arnold (Publishers) Limited, 1972.

PHOSPHORUS-NITROGEN COMPOUNDS: CYCLIC, LINEAR, AND HIGH POLYMERIC SYSTEMS. H. R. Allcock. Academic Press, 1972.

VORTEXES IN AIRCRAFT WAKES

AIRCRAFT WAKE TURBULENCE AND ITS DETECTION. Edited by John H. Olsen, Arnold Goldburg and Milton Rogers. Plenum Press, 1971.

EXPLORATORY FLIGHT INVESTIGATION OF AIRCRAFT RESPONSE TO THE WING VORTEX WAKE GENERATED BY JET TRANSPORT AIRCRAFT. William H. Andrews, Glenn H. Robinson and Richard R. Larson. NASA TN D-6655, March, 1972.

FLIGHT IN TURBULENCE. AGARD Conference Proceedings No. 140. Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organization, 1973.

THE CRY OF THE HUMAN INFANT

SENSORY-VOCAL STUDIES OF A TWIN PAIR WITH CEPHALIC UNION. Philip Peltzman, Peter Ostwald, Charles Yeager and David Manchester in *Neuropaediatrie,* Vol. 2, pages 79–97; 1970.

THE SOUNDS OF INFANCY. Peter F. Ostwald in *Developmental Medicine and Child Neurology,* Vol. 14, pages 350–361; 1972.

THE SEMIOTICS OF HUMAN SOUND. Peter F. Ostwald. Mouton, 1973.

FERDINAND BRAUN AND THE CATHODE RAY TUBE

FERDINAND BRAUN—INVENTOR OF THE CATHODE-RAY TUBE. G. Lewis and F. J. Mann in *Electrical Engineering,* Vol. 25, No. 4, pages 319–327; December, 1948.

ELECTRICAL OSCILLATIONS AND WIRELESS TELEGRAPHY. Karl F. Braun in *Nobel Lectures in Physics: Volume I—1901–1921.* American Elsevier Publishing Co., Inc., 1964.

FERDINAND BRAUN: LEBEN UND WIRKEN DES ERFINDERS DER BRAUNSHEN RÖHRE. Friedrich Kurylo. Heinz Moos Verlag München, 1965.

MATHEMATICAL GAMES

NEWCOMB'S PROBLEM AND TWO PRINCIPLES OF CHOICE. Robert Nozick in *Essays in Honor of Carl G. Hempel,* edited by Nicholas Rescher. Humanities Press, 1970.

NEWCOMB'S PARADOX REVISITED. Maya Bar-Hillel and Avishai Margalit in *The British Journal for the Philosophy of Science,* Vol. 23, No. 4, pages 295–304; November, 1972.

THE AMATEUR SCIENTIST

FUNDAMENTALS OF OPTICS. Francis A. Jenkins and Harvey E. White. McGraw-Hill Book Company, Inc., 1957.

TOOLS OF THE ASTRONOMER. Gerhard R. Miczaika and William M. Sinton. Harvard University Press, 1961.

© 1974 SCIENTIFIC AMERICAN, INC

Satellites have taken up where Noah left off.

The Ark may be obsolete, but there's a new way to help preserve the species against flood—and a host of other perils.

It's the watchful eye of the satellite, keeping vigil over our planet and its ecology.

These orbiting sensors see into the ocean depths. They monitor our waterways not only for flood control but for pollution. They check on the health and composition of our forests. They track storms to help curb the threat to life and property.

RCA built the first U.S. weather satellite—TIROS—in 1960. Since then, we've created 78 other successful satellites for a variety of missions. They've logged 60 billion miles in orbit.

And they've grown increasingly sophisticated. Our latest satellites not only stand watch on the global environment, they even take earthly temperatures from hundreds of miles in space. Fittingly enough (because they're named for our customer, the National Oceanic and Atmospheric Administration) the new series is called NOAA.

Electronics is creating new ways to make life better. And RCA, which helped create the technology, is still innovating the electronic way.

The electronic way

RCA

© 1974 SCIENTIFIC AMERICAN, INC

© 1974 SCIENTIFIC AMERICAN, INC