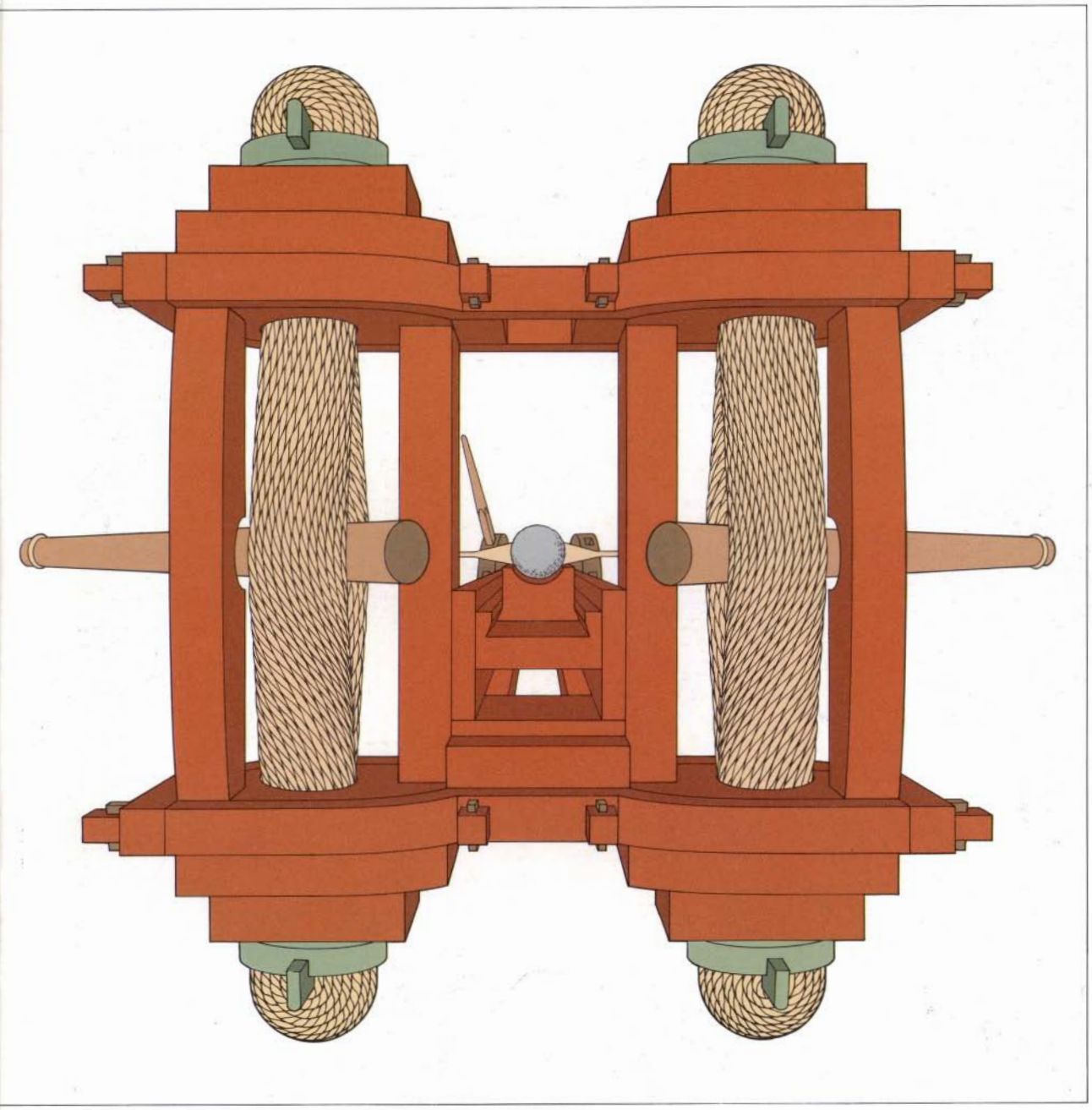
SCIENTIFIC AMERICAN



ANCIENT CATAPULTS

$1.50

March 1979

18 EPA* EST. MPG **28** EST. HWY

*Use the estimated mile-per-gallon number for comparison purposes. Your mileage may differ depending upon speed, weather and trip length. Actual highway mileage will probably be lower than the highway estimate. California mileage lower and automatic transmission is required.

LeBARON TOWN & COUNTRY. THE MOST DISTINCTIVE WAGON IN AMERICA TODAY.

No other wagon stands out on the road quite like a LeBaron Town & Country. With the classic Town & Country look, reminiscent of handcrafted mountain ash and mahogany woodwork.

A distinctive size. Versatile. Ready to handle either a trip to the antique shop or the country club in style.

Town & Country is a wagon rich in Chrysler detail. With the finest Chrysler options available. Like genuine leather seating, made specially soft to the touch. A 4-speaker digital stereo system. And most every other power-assisted feature you can imagine.

The classic look, the modern convenience, the traditional Chrysler luxury. It all makes LeBaron Town & Country the most distinctive wagon in America today. And no other wagon even comes close. *Add a little life to your style*

CHRYSLER LeBARON

Discover more.

Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.

Most premium Scotch is blended and depends on several whiskies for taste and smoothness.

Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.

Try the taste beyond premium Scotch.

The Glenlivet
Unchanged since 1824.

THE TASTE BEYOND PREMIUM SCOTCH

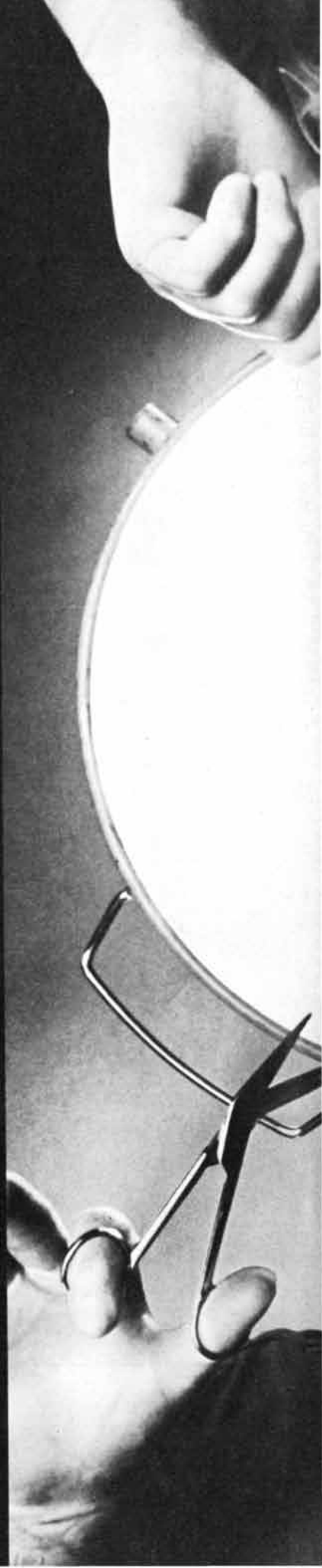
Established 1845 **SCIENTIFIC AMERICAN** March 1979 Volume 240 Number 3

ARTICLES

DEPARTMENTS

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

You may become susceptible to a major surgical complication.



Needless hospitalization.

These days, the hospital bill can be more of a shock than the operation. Surgical fees alone have doubled in the last five years.

That's a problem for you. And us.

Through healthy competition, private health insurance right now covers 177,000,000 Americans. 147,000,000 of these are protected against even catastrophic expenses.

But, though the system is basically sound, no one, including us, has done enough about containing health care costs. So we'd like to recommend something that could be done today: reduce in-hospital time for surgery.

In many cases, minor surgery and pre-surgery testing could be handled on an ambulatory basis. We also support—and an increasing number of insurance companies pay for—second, and even third, opinions for elective surgery.

We have a booklet called *Reducing Surgical Costs*. You can have a free copy if you'll write us at the Health Insurance Institute, Dept. 13, 1850 K Street, N.W., Washington, D.C. 20006.

Exercising your rights as an involved consumer could do a lot to get health costs in trim.

THE HEALTH INSURANCE COMPANIES IN AMERICA

Let's keep health care healthy

SAAB

Introducing the SAAB 900 series.

THE NEW LOOK OF PERFORMANCE.

The new Saab 900 series. Advanced technology fused to stirring design creates new and superior performance automobiles.

The totally new Saab 900 series. The longest, sleekest Saab cars ever. But increasing length wasn't for looks alone. Their longer wheelbase is integrated with a new steering system and new suspension geometry for outstanding roadability.

The totally new Saab 900 series. Introduces a new and innovative asymmetrical instrument panel. Controls are practically equidistant from the driver's hands. Gone forever is searching and fumbling.

The totally new Saab 900 series. Even the air inside the car hasn't been overlooked. Saab engineers have designed a unique fresh air filter that eliminates all airborne pollen. You won't find it on any other car in the world.

To these striking new automotive advances the new 900 series adds fuel injection, front wheel drive, and rack and pinion steering. The result? Truly astonishing performance. Even for the world acclaimed Saab.

The 900 Turbos

The high point of the thrilling, new 900 series. Probably the most exciting cars you can buy. The remarkable Saab turbo engine delivers the kind of energy formerly reserved for racing, but puts it at your command in everyday driving situations. With the Turbo you can have your exhilarating performance in either 3-door or for the very first time, 5-door configuration.

The 900 EMS

This is the Saab for the sports-minded. An extraordinary sports sedan. It's fast, responsive and well-equipped. Automatic transmission with power-assisted steering is now available for the first time. Bilstein gas shock absorbers and sports steering wheel are standard. Sliding steel sun roof (as shown) is optional.

The 900 GLE

One of the few cars in the world that can truly be called a high-performance luxury sports sedan. The interior is sheer elegance. Plush, soft velour, color coordinated door panels, seamless deep-pile carpeting, electrically heated front seats, and stereo speakers are standard. Automatic transmission, power steering and sun roof are available as options.

These are just a few of the examples of how the 900 series has become the new look of performance. To appreciate the complete excitement test drive one of the new 900's at your Saab dealer. The new look of performance will be a driving revelation.

THE SAAB 900 SERIES
THE COMMAND PERFORMANCE CARS

THE COVER

The design on the cover shows the front end of a large stone-throwing catapult, drawn according to specifications laid down by the Roman military engineer Vitruvius in about 25 B.C. This particular weapon was based on the torsion principle: tightly stretched bundles of elastic fibers, consisting of animal sinew, horsehair or human hair, were further strained by a rigid bow limb as the catapult was brought to a full draw (see "Ancient Catapults," by Werner Soedel and Vernard Foley, page 150). The *H*-shaped structure seen in cross section at lower center is the stock, a compound wood beam that formed the main axis of the device. The view here is straight back along the trajectory of the stone, which with the weapon cocked rested at the rear end of a long, shallow trough cut into the top surface of the slider, a movable beam that in turn fit into a dovetail groove in the top of the stock. The stone was loaded in a pouch at the midpoint of the bowstring, which was grasped and released by a claw-and-trigger arrangement attached to the top rear surface of the slider (out of sight in this view). A winch with a ratchet wheel was used to force the slider to the rear, taking the bowstring with it. Catapults of this type, which were capable of throwing stones weighing as much as 78 kilograms, could also be adapted to fire large arrows. A side view of a similar machine appears on page 152.

THE ILLUSTRATIONS

Cover illustration by Dan Todd

Page	Source
21–27	Ilil Arbel
36	Lotte Jacobi
47	*The Cosmographical Glass: Renaissance Diagrams of the Universe.* © 1977, The Huntington Library
48–50	Bernarda Bryson
51–53	Ilil Arbel
55	Richard F. McCrosky and Cheng-Yuan Shao, Harvard College Observatory (*top*); Eleanor F. Helin, California Institute of Technology (*bottom*)
56	National Aeronautics and Space Administration
57	Andrew Tomko
58–59	Dan Todd
60	Albert Miller
61	Earth Physics Branch of the Department of Energy, Mines and Resources, Ottawa
62–65	Albert Miller
67	Garth L. Nicolson, University of California at Irvine
68–76	Carol Donner
94–95	Gerhard Herzberg (*top*), Jon Brenneis (*bottom*)
96–109	Gabor Kiss
113	Alan L. Gilchrist, State University of New York at Stony Brook
114	Alan L. Gilchrist, State University of New York at Stony Brook (*top*); Ilil Arbel (*bottom*)
116–124	Ilil Arbel
127	Alexander von Muralt, University of Bern
128–135	Jerome Kuhl
137	Photo Researchers, Inc. (*top*), Rapho-Photo Researchers Inc. (*bottom*)
138–143	Tom Prentiss
144	Bernd Würsig, University of California at Santa Cruz
145–146	Tom Prentiss
148	Photo Researchers, Inc. (*top left and bottom*); Bernd Würsig, University of California at Santa Cruz (*top right*)
151–160	Dan Todd
162–172	Michael Goodman

HP measurement and computer advances

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

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

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

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

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

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

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

A working partnership with HP.

This informative management booklet outlines HP's approach to doing business. It summarizes the expertise, resources, support, and computer products we bring to customers. For a free copy write to A.P. Oliverio, Vice President, Marketing, Hewlett-Packard Co., 1501 Page Mill Rd., Palo Alto, CA 94304.

extend your possibilities.

HP's new fiber optic link: signal advantages in digital data transmission.

Hewlett-Packard's HFBR series fiber optic system can simplify error-free data communications between computers or within systems, especially in environments where electromagnetic interference is generated.

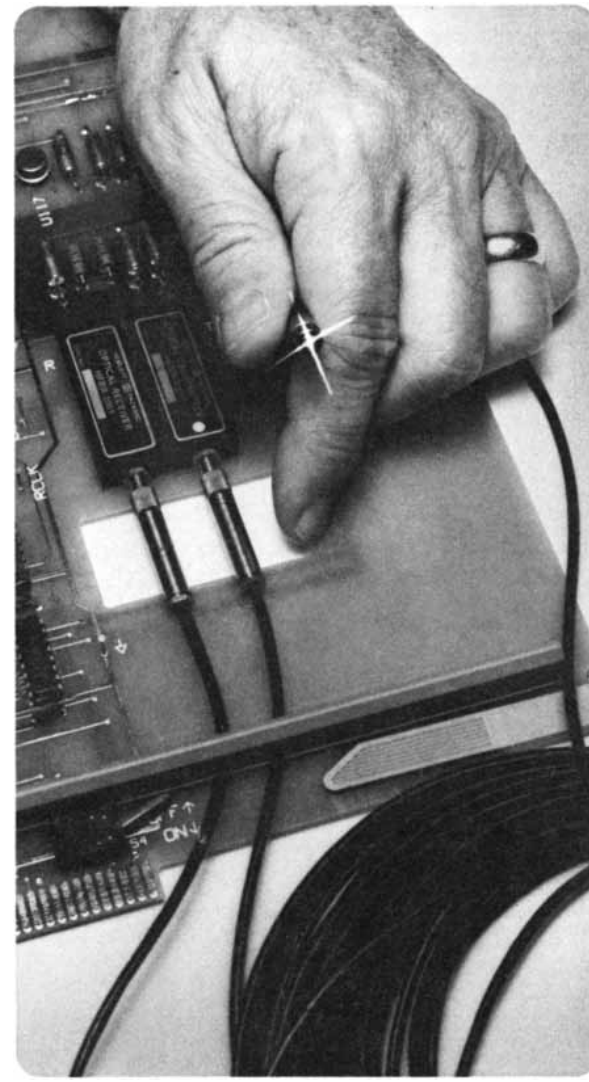
In environments where high-voltage equipment, heavy electrical machinery, transformers, or other devices generate electromagnetic interference, reliable data transmission between computers or within computer-instrument systems can be a problem. The usual data links of twisted wire or coaxial cable may necessitate the use of expensive shielding, conduit, isolation transformers, or data error checking and retransmission circuitry.

However, by converting the electrical signals to light pulses with no electrical charge, data can be transmitted over a single, hair-thin optical fiber without error or disruption. It is precisely such a system that HP has developed. Using standard TTL electrical interfaces and requiring only a single 5-volt power supply, small optical transmitters and receivers translate the coded electrical transmission into light pulses at one end of the fiber, and back to electrical signals at the other. The fiber, reinforced and jacketed in polyurethane, can span distances of up to 100 metres.

The system offers other advantages compared to electrical connections: broad bandwidth over long distances; light-weight, small-diameter, flexible cables; electrical isolation between connected units; and no externally radiated signal.

Typical applications include distributed system networks, process control systems, secure communications, aircraft or shipboard data links, high-voltage or electromagnetic field research, remote instrumentation systems, and factory data collection.

Prices start at $570* for a 10-metre system. The HFBR-1001 Transmitter and HFBR-2001 Receiver are available separately at $225* each, and various fixed-length connector-cable assemblies are available.



Fiber optic transmitter and receiver are connected to modulation/demodulation circuitry on a PC board for insertion in a computer.

HEWLETT hp PACKARD

1503 Page Mill Road, Palo Alto, California 94304

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

Mail to: Hewlett-Packard, 1503 Page Mill Road, Palo Alto, CA 94304
Please send me further information on

( ) HP 300 computer
( ) HFBR series fiber optic system

Name _____
Company _____
Address _____
City _____ State _____ Zip _____

*Domestic U.S. prices only

00940

LETTERS

Sirs:

In Charles F. Westoff's article "Marriage and Fertility in the Developed Countries" [SCIENTIFIC AMERICAN, December, 1978] he notes that the birthrate in the U.S. has been declining over the past two centuries except for the brief "baby boom" after World War II. He states that "the real question, and the more perplexing one, is what caused that exception." I submit that the real question is what caused the protracted slow decline. Insight into why the birthrate of the U.S. and that of most of the developed nations have reached such low levels that their population growth is approaching zero would be of greater significance in dealing with the current population crisis in the developing world than would causation of the baby boom. It is much easier to make plausible assumptions about what caused the baby boom than to account for the historic decline in our birthrate.

After hedgingly pointing out that the record of population projections is not a happy one and that it is foolhardy even to speculate about future trends in fertility and population growth Westoff takes a long look ahead and sees the strong possibility that declining fertility in the developed nations may continue beyond zero population growth even to the point of threatening them with self-extinction. With that startling outlook the news media had a field day. One leading newspaper warned that several European countries are "struggling to cope with population declines." Struggling with what—the good life? With world population continuing to expand rapidly it seems premature to be stirring apprehension over low birthrates in a small segment of the global population, regardless of stage of development.

In prognosticating that the U.S. population (220 million) may reach only 253 million by the year 2015 Westoff disregards the census undercount and the annual influx of illegal immigrants, now at flood stage. Most estimates of their numbers approximate or exceed our current net gain from births over deaths. These immigrants should not be discounted simply because they cannot be counted. I expect on the basis of less rose-tinted spectacles and the fickleness of futurology that the U.S. will have 253 million inhabitants by the year 2000. The producers are already at hand, modern contraceptive technology is in serious need of improvement and the urge to reproduce remains a deep-seated attribute of human biology. Moreover, life expectancy keeps increasing and the administrative stance on illegal immigrants is leniency. All the same, for the welfare of oncoming generations, I sincerely hope that Dr. Westoff is right.

ROY O. GREEP

Harvard Medical School
Boston, Mass.

Sirs:

Dr. Greep asserts that the more important question for the developing world today is why the birthrate declined historically in the developed countries rather than why the baby boom occurred. I completely agree with that observation. The serious population problems of the developing countries was not, however, the point of my article, which was addressed exclusively to the emerging reactions to zero or negative population growth in the developed countries. Although world population growth is still much higher than is desirable, governments, including the United Nations, do not react to the world as a whole. Different nations have different population problems and will react accordingly. For example, the government of East Germany is already trying to stimulate fertility in response to negative population growth, and this domestic concern is not altered by the fact that Bangladesh and two-thirds of the rest of the world have a different kind of population problem.

Dr. Greep also complains in effect that my population prognosis for the U.S. ignores the "influx of illegal immigrants" (the other factors of the census undercount and increasing life expectancy that he points to are not really important in terms of the magnitude of future population growth) and that I should not have discounted its impact simply because the volume of illegal migrants cannot be estimated accurately. His characterization of this volume as now "at flood stage" suggests that he is susceptible to the earlier Department of Justice concerns and to the probable exaggerations of the press (which, he rightly deplores, sensationalized my article). Serious students of the subject of undocumented aliens are not at all convinced that the numbers have really increased much from earlier times or that much of the movement is not of a back-and-forth nature.

I share Dr. Greep's concern with the problems of rapid population growth and with the need for improvement of contraceptive technology. There are, however, many well-meaning advocates of zero population growth who evidently find it difficult to tolerate any discussion of situations where the perceived need is to arrest declining populations. There are countries that might be quite satisfied to stay at zero population growth or even to return to it.

CHARLES F. WESTOFF

Princeton University
Princeton, N.J.

Sirs:

Over the years it has been mentioned a number of times in the pages of *Scientific American* that the supernova of July 4, 1054, was not observed in Europe. The matter is dealt with most recently in your note "A Star Is Reborn" ["Science and the Citizen," SCIENTIFIC AMERICAN, October, 1978].

There is probably good reason why a report of this extraordinary celestial event cannot be found in Christian chronicles: it was highly embarrassing to the clergy of both the East and the West. In July, 1054, the papal legate laid on the altar of Hagia Sophia a bull excommunicating the patriarch of Constantinople, thereby rending asunder the Christian world for the next 900 years in the Great Eastern Schism. The belief prevailing at the time, that significant events on the earth are attended by portents in the sky, made it appear that the supernova was lending unwanted publicity to the internal squabble. Then, as now, people celebrate their marriages and are glad to have them broadcast, but they tend to hush up their divorces. It was churchmen who wrote the chronicles, and they could hardly have been eager to highlight the breach.

C. L. THOMAS

Andelfingen, Switzerland

Scientific American, March, 1979; Vol. 240, 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, vice-president and treasurer; C. John Kirby, vice-president; Arlene Wright, assistant treasurer.

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

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

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

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

Name

New Address

Old Address

The Jameson Irish Sweepstakes

This could be your ticket to Ireland

Win one of 15 round trips for 2 to Ireland on Aer Lingus

Jameson Irish Whiskey will be giving 15 lucky winners two round trip tickets to Ireland on Aer Lingus, Ireland's quality International Airline... Plus, $1,000 expense money to help you enjoy your stay.

Plus 300 other prizes

300 lucky runners-up will win an elegant Irish lead crystal heart-shaped pendant with a sterling silver chain.

Simply answer the questions on the coupon correctly. The answers to all the questions appear on the Jameson label. Not the one in this ad, however. That would be too easy.

Just take a short trip to your liquor store, favorite bar or restaurant, and look at a bottle of Jameson. It's one short trip that could turn into one beautiful trip to Ireland. You'll also find the answer to "Why Jameson instead of Scotch?" (Because it's lighter and more delicate tasting than fine Scotch.)

Jameson...World's largest selling Irish Whiskey.

OFFICIAL RULES

1. On entry form, or a 3″ x 5″ piece of paper, print your name, address and zip code. Then, answer the 3 questions with information found on the front label of any bottle of Jameson Irish Whiskey. If you don't own a bottle visit your favorite restaurant or tavern, or go to any participating liquor store and look at a bottle of Jameson.
2. Enter as often as you wish, but each entry must be properly completed, addressed and mailed in a separate envelope and received by June 30, 1979, to be eligible. Prize winners will be determined in a witnessed random drawing of correct entries by Frederick Siebel Associates, an independent judging organization whose decisions are final. No purchase required.
3. Each grand prize winner will receive 2 round trip economy tickets to Ireland on Aer Lingus plus $1,000 cash. It's incumbent upon winners to make their own arrangements for trip. Winners shall depart from the normal Aer Lingus departure points within the Continental U.S.A. Any additional expenses incurred on the trip will be borne by the winners. Trips must be taken by Dec. 31, 1980. Any necessary taxes must be paid by winners. The 300 second prize winners will receive an elegant Irish cut glass lead crystal heart-shaped pendant with a sterling silver chain. Prizes are non-transferrable and non-redeemable.
4. Only one prize per family or household. The odds of winning will be determined by the number of correct entries received. All prizes will be awarded.
5. Sweepstakes open to residents of the Continental U.S.A., Alaska and Hawaii. Employees and their families of Calvert Distillers Company and Aer Lingus, their affiliated and subsidiary companies, liquor wholesalers and retailers, their advertising agencies and judging organization, are not eligible. Sweepstakes void where prohibited or restricted by law. All federal, state and local laws apply.
6. Entrants must be of legal drinking age under the laws of their home state.
7. A list of all major winners can be acquired at the conclusion of the sweepstakes by sending a stamped, self-addressed envelope to Jameson Irish Sweepstakes, P.O. Box 8260, St. Paul, Minnesota 55182. NO PURCHASE REQUIRED.

JAMESON IRISH SWEEPSTAKES

I have read the contest rules listed on this page, and would like to enter the Jameson Irish Sweepstakes. My answers are written below (correct answers appear on the front label of each Jameson bottle).

1. What year appears on the Jameson label? _____
2. What Dublin street name appears on the Jameson label? _____
3. What is the motto under the Jameson coat of arms? _____

Name _____
Address _____
City _____ State _____ Zip _____

Mail entries to: Jameson Irish Sweepstakes,
P.O. Box 8209,
St. Paul, Minnesota 55182

For more information on travel to Ireland, call Aer Lingus toll free (see local directory).

Jameson. World's largest-selling Irish Whiskey. 80 PROOF • CALVERT DIST. CO. N.Y.C.

You're looking at the actual size of three small sonic wonders. The Micro Series by Technics.

Their size, design and elegance are simply amazing. Yet even more amazing is how Technics combined micro technology with our own brand of audio know-how to prove components no longer have to be big and bulky to sound big and beautiful.

Take our SE-CO1 DC power amp. Instead of gigantic capacitors and a massive transformer, you'll find a high-speed switching power supply. Its filter capacitors recharge 40,000 times every second instead of the usual 120, making them smaller and more efficient at providing pure operating DC. This innovative pulse power supply not only enabled us to use a micro power transformer, it also added to the SE-CO1's sonic excellence.

So does direct coupling. Because when coupling capacitors are eliminated, nice things happen. Like a deep, tight bass. All the way down to DC (0 Hz). Bass can't get any deeper.

With an amplifier this accurate, you want power meters that measure up to it. 12 LED's per channel provide true peak-power indication and extremely fast attack time.

Another big surprise is the Micro Series SU-CO1 preamp. It's one preamp, but it works like two. Because it has a built-in pre-preamp for moving coil cartridges. Most preamps don't. Most preamps don't give you gold-alloy-plated connections, either. We do, to maximize signal transfer.

The SU-CO1 delivers a phono S/N ratio that goes far beyond the word quiet, 94 dB (MM input). And when it comes to 0.005% distortion (3V output), the word inaudible takes on a special meaning. So do tonal accuracy and

SE-CO1

Continuous power per channel into 8 ohms, 20 Hz – 20 kHz	Total Harmonic Distortion	Signal to Noise Ratio
40 watts	0.03%	110 dB

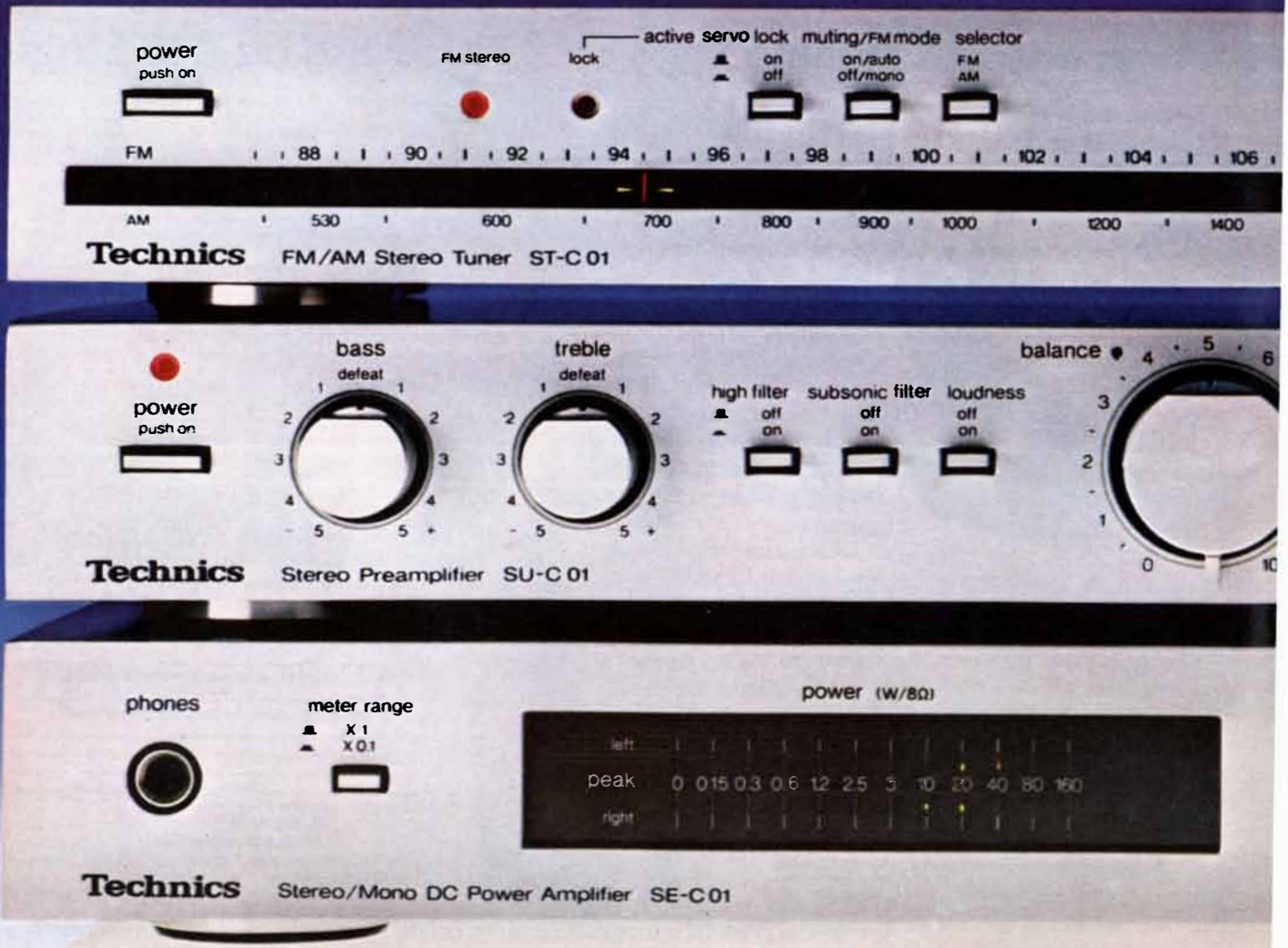
SU-CO1

Total Harmonic Distortion	Phono Signal to Noise Ratio	Frequency Response
0.005% (phono) 0.003% (aux)	94 dB (MM, 5mV) 78 dB (MC, 250 μV)	3Hz–100kHz, –1 dB

ST-CO1

FM Sensitivity 50 dB (stereo)	FM Selectivity	Stereo Separation (1 kHz/10 kHz)	Total Harmonic Distortion (stereo)
38.3 dBf	75 dB	45/35 dB	0.15%

Introducing the Micro Series by Technics. With performance this big, the last thing you expect is components this small.



dynamic range, especially with tight RIAA equalization of ± 0.2 dB. And enough headroom to handle 200 mV signals without overload distortion.

To add the finishing touch to the Micro Series, there's the ST-C01 tuner. Not only does it give you great FM specs, it gives you great FM sound. And that's a big achievement, considering its small size.

Two RF stages with 4-pole, dual-gate junction FETs boost sensitivity while Technics-developed flat group delay filters provide high selectivity and low distortion. And for a beautifully defined stereo image, there's a phase-locked-loop IC in the MPX section.

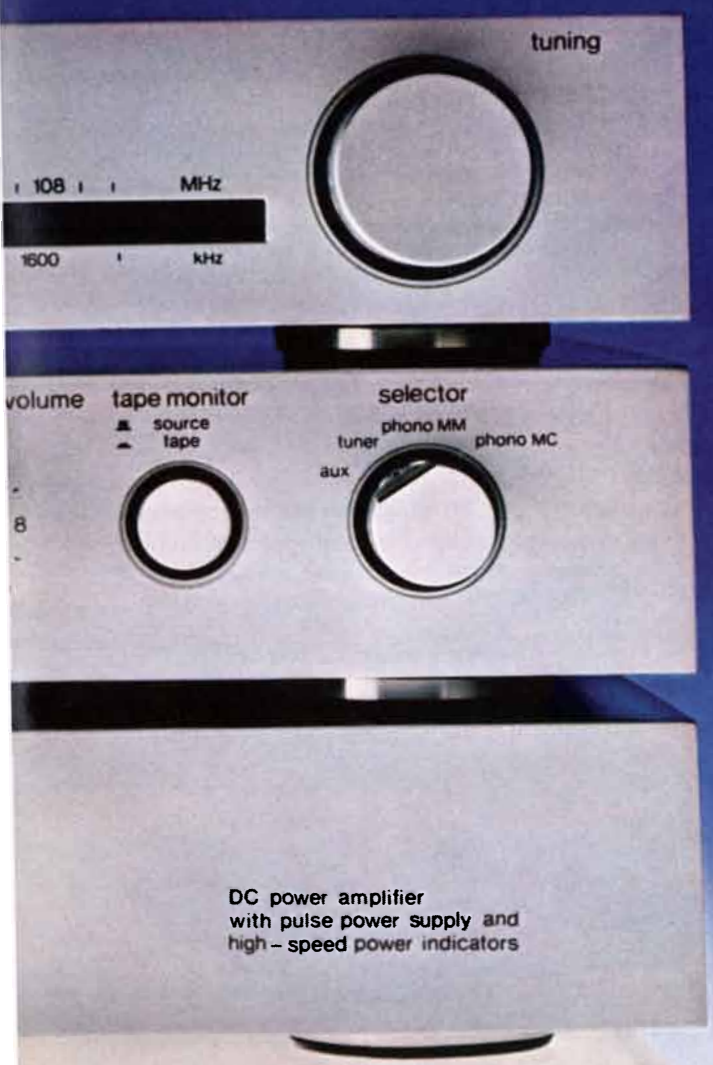
The ST-C01 is a breeze to tune. Instead of conventional separate tuning meters, the center-of-channel indicator is on the tuning dial where it's easy to see. Two LED arrows point you in the right direction for fine tuning. When both arrows disappear, you're perfectly tuned.

To keep you perfectly tuned, Active Servo Lock (ASL) follows the FM signal, so even if the signal drifts, reception won't.

Experience the Micro Series. Once you do, you'll agree: The big thing about them is definitely not their small size.

Technics
micro series

Actual Size



There is now a Mercedes-Benz station wagon

The new 300 TD. Before it could be a Mercedes-Benz station wagon, it had to behave like a Mercedes-Benz car.

Mercedes-Benz has engineered a station wagon that handles like – and is as solidly built as – a Mercedes-Benz car.

It does not wallow about on the road like a hippopotamus.

And it has been designed to spare its owner the aggravation of a chorus of rattles and creaks.

"First and foremost," reports *Car and Driver* magazine, "is that when you drive it empty, you very quickly forget you're driving a wagon. It handles very much like a sedan, with none of the ponderous, tail-happy lethargy we've come to expect..."

But what happens to the handling when the 300 TD is heavily loaded?

There is a sensor buried deep within every Mercedes-Benz station wagon. Should a passenger step aboard, or luggage be added, this sensor orders a special pump to send an extra supply of hydraulic fluid to the rear shock absorbers.

This raises the rear of the vehicle to its normal, most effective level.

And this, in turn, maintains the correct front-wheel geometry so that the station wagon can be steered precisely even under heavy load, instead of wallowing clumsily from side to side.

It sounds like a Mercedes-Benz car

Creaks, rattles, thrumming noises – are they the inescapable bugaboos of a station wagon?

Not so, said the Mercedes-Benz engineers.

Car and Driver agrees: "The back end doesn't sound like a crate of empty muffin tins whenever you hit a bump, which is something you can't say for your average wagon. Mercedes-Benz designers have found ways of securely buttoning down all the movable panels that usually cause station-wagon rattles..."

Good brakes are not enough

The handling ability of a car is considered of paramount importance at Mercedes-Benz. The reason is simple: if at all possible, it is always preferable to *avoid* an accident.

The new Mercedes-Benz 300 TD.
Only 2,500 will be available in the U.S. this year.

Good 4-wheel disc brakes alone are often not enough. The car must also be capable of violent evasive action without loss of control.

And so the new Mercedes-Benz 300 TD station wagon has been fitted with the sophisticated steering and suspension systems of a Mercedes-Benz car.

Steering that "talks" to you

There is zero-offset steering built into the front suspension of the 300 TD to provide more precise steering control in emergency situations.

The power steering itself has "feel." It tells your hands exactly what the front wheels are doing at every moment.

And the suspension is independent on all four wheels. Should one wheel hit a bump in the middle of a corner, that wheel alone is affected, leaving the other wheels independent to maintain their grip on the road. There is none of that alarming, sideways hop-skip-and-lurch that one expects from a less-expensive solid rear axle; the Mercedes-Benz will simply shift a few inches to one side, then smoothly resume its course.

It uses Diesel fuel. Sparingly.

The Mercedes-Benz station wagon is no gas-guzzling leviathan. For one thing, it measures a trim and nimble 15 feet 11 inches from stem to stern; it is the length of a Mercedes-Benz car. For another, it doesn't use gas at all.

The powerplant is the remarkable 5-cylinder Mercedes-Benz *Diesel.* This 3-liter engine is coupled to a 4-speed automatic transmission to achieve a balance of performance *and* economy. It delivers an estimated 23 mpg.

This estimate is from the U.S. Environmental Protection Agency. *Remember:* Compare this estimate to the "estimated mpg" of other station wagons. You may get different mileage, depending on how fast you drive, weather conditions, and trip length.

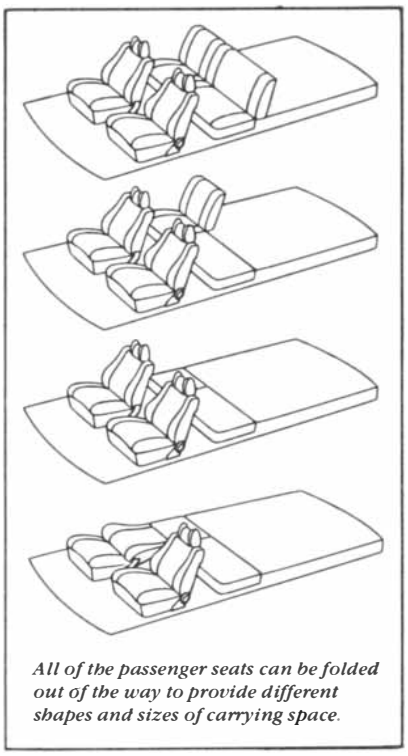
A lesson in the use of space

The interior design of the 300 TD station wagon is an object lesson in the ingenious use of space.

Raise the tailgate and look for its hinges and damper units. You'll find them tucked discreetly into the roof. It costs more to do it this way, but the Mercedes-Benz engineers wanted an absolutely smooth inner-roof surface so that bulky packages wouldn't snag on outjutting hinges.

There are seats for five adults. All of the passenger seats can be folded out of the way (*see illustration, right*) in a multitude of sizes and shapes of carrying space.

"If the object is *really* long," noted *Car and Driver,* "say a carton of record-length spaghetti or a python with rigor mortis, that's no problem either. Just unlatch and remove the right-side cushion of the second seat. Fold that side of the seatback forward. Then fully recline the passenger-side front bucket. That leaves an uninterrupted stretch from the tailgate in back to the glove-box door in front. What more could you ask?"



All of the passenger seats can be folded out of the way to provide different shapes and sizes of carrying space.

More standard equipment

In typical Mercedes-Benz fashion, the list of standard equipment is quite staggering in its length.

To name a few items: Climate control air-conditioning. Cruise control. A 4-speed automatic transmission.

Power-assisted steering and brakes. Electric window lifts.

Also, an AM/FM stereo receiver with electric antenna. A quartz-crystal chronometer. Wiper, washer, and defroster for the tailgate window. Halogen fog lamps. A pair of outside mirrors, both adjustable from inside the car.

There is even central locking: a turn of the key in the driver's door locks all four doors, the tailgate, and the fuel filler port. Simultaneously.

Engineered like no other car in the world

The Mercedes-Benz aim is doggedly single-minded. It is to build safe, comfortable, practical cars with as few imperfections as possible.

This philosophy puts engineering ahead of petty economies and precludes the mass production of inexpensive cars.

A Mercedes-Benz is engineered like no other car in the world.

Load this space, or add passengers, and a sensor orders the suspension to raise the rear of the station wagon to its normal, most effective level.

The alloy wheels shown are at extra cost.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1929: "An event unique in the annals of the United States takes place on the fourth of March: the first engineer to hold the office of Chief Executive of the nation will be inaugurated. Other presidents have had some engineering experience, but none has had such complete, worldwide and successful experience as Mr. Hoover. He takes office with such a well-rounded knowledge of human, national and international affairs that it is believed the United States will assume a higher place as a world power under his leadership."

"Dr. F. A. Vening Meinesz, the great geodesist of Holland, who is the only man in the world to design an apparatus for determining gravity at sea with a satisfactory accuracy, has just completed an expedition into the Atlantic, the Caribbean and the Gulf of Mexico on an American submarine, the *S-21,* assigned by the Secretary of the Navy. Gravity observations were made over the Nares Deep to the north of the island of Porto Rico, the Bartlett Deep, lying to the south of Cuba, and the Sigsbee Deep in the Gulf of Mexico. There are many deeps, or troughs, under the ocean waters, but what caused them and whether they are permanent features of the ocean bottom are problems unsolved. In or near them some of the greatest earthquakes occur—why, we do not know. The values of gravity obtained by Dr. Meinesz and his colleagues may throw much light on these strange depressions in the earth's surface."

"A quarter of a century ago the Wright brothers, Wilbur and Orville, made the first successful flight in an airplane at Kitty Hawk, N.C. At the dedication of Wright Field in Dayton, Ohio, several months ago, Orville watched many airplanes roar by in a terrific race. A direct question put by someone nearby seemed to carry his thoughts back to the days of the early experiments made by his brother and himself, and he is said to have sighed and answered slowly: 'No, neither of us thought it would ever come to this.' In its 25 years aviation has made undreamed-of strides in spite of the fact that the first five of those years were required to convince the world that man *could* fly. After little more than another five years a war showed the great military value of airplanes and, because of the extensive use of them under trying circumstances, gave a hint of their potentialities during peace time. Following that conflict the field of their usefulness was greatly expanded: mail and passenger air lines were started, oceans were crossed and the North Pole was reached, and just the other day five Army men stayed aloft in a three-motored airplane for more than 150 hours."

"The tangible objects with which we are familiar are constituted of molecules. These in turn are composed of atoms, and these of the positively charged and massive protons and the negatively charged and mobile electrons. The light that makes plants grow and that gives us warmth has the double characteristics of waves and particles and is found to consist ultimately of photons. Having carried the analysis of the universe as far as we are able, there thus remain the proton, the electron and the photon—only these three. We sometimes think of standardization as being the distinctive keynote of modern industry. But even a Ford car has hundreds of parts that differ from one another. What then shall we say of the Workman who by using only three different parts, protons, electrons and photons, has made a universe with its infinite variety of beauty and life?"

"Dr. Tuve, Dr. Gaviola and Mr. Hafstad, scientists at the Department of Terrestrial Magnetism of the Carnegie Institution of Washington, have developed 5,000,000-volt equipment for the study of atomic structure. With it they have produced the highest voltage ever obtained: 5,200,000."

SCIENTIFIC AMERICAN

MARCH, 1879: "*The Engineer* has published accurate copies of the three principal drawings of Edison's electric-light apparatus lodged with the French specification, and we have no doubt that they are in all respects identical with those that accompany the applications Mr. Edison has made in other countries for a patent. In all of this it can be seen that we have not one word concerning any new or extraordinary contrivances for the electric light. Mr. Edison has many other patents in progress, but the one with which we have just dealt is, no doubt, *the* patent, the one that has attracted more attention, and the publication of the contents of which has been looked for with more avidity than perhaps any other patent ever applied for. It may not be improper to say that Mr. Edison's complete specification has not yet been filed in this country and cannot be seen at the Great Seal Patent Office, or anywhere else save Paris."

"In Kant's exposition of the theory of the development of the solar system he supposes the comets to be formed from the matter of the condensing solar nebula. Laplace, on the other hand, in his exposition of the nebula hypothesis, took the ground that comets were formed from the matter scattered through the stellar spaces, and that in their origin they have no relation to the solar nebula. Have we, in the accumulation of facts since the days of Kant and Laplace, learned anything that may help us to decide between these theories? Such is the inquiry proposed by Prof. Hubert A. Newton, who considers first what peculiarities each of the theories requires in the shape and distribution of the cometic orbits, and second compares with the theories the facts that have been observed with regard to the paths of 247 comets. The cometic paths are represented by Prof. Newton in two graphic curves, and he arrives at the conclusion that the curve corresponding to the actual cometic paths is in good agreement with the theoretical curve deduced from Laplace's hypothesis, whereas it does not agree so well with that deduced from Kant's. It would seem, then, that the origin of comets must be placed in interstellar space."

"After an interruption of five and a half years, not idly spent, Dr. Schliemann has resumed his search for the buried Troy. The work of opening trenches on Mount Hissarlik began on September 30 of last year. Dr. Schliemann attacked the western and northern front of the circuit wall, the construction of which Homer ascribes to Poseidon and Apollo, as well as all that remains of the prehistoric mansion immediately to the northwest and west of the great gate, which he attributes to the ancient town's chief or king, because it is the largest and best-built house in the city. The walls of this mansion are at a depth of 26 to 28 feet below the surface of the hill and are precisely like the Cyclopean walls of the Royal Palace at Mycenae and the Cyclopean houses at Tiryns."

"In spite of midwinter cold, cases of yellow fever are reported at New Orleans and Vicksburg. In some cases returned refugees entering houses that have been closed since summer are said to have taken the disease; in other instances the disease is attributed to the active disinterment of victims hastily buried when the plague was at its height. In Mississippi and Tennessee physicians insist that if this dangerous work is not stopped, a return of the epidemic may be expected in the spring, and in some cases the health commissioners have been compelled to forbid the removal of last summer's victims. Wealthy families no doubt find it painful to leave their dead in potter's fields, but personal feelings should not be suffered to endanger the entire community."

THE AUTHORS

ELI GINZBERG ("The Professionalization of the U.S. Labor Force") is A. Barton Hepburn Professor of Economics at the Columbia University Graduate School of Business. A native New Yorker, he received his bachelor's, master's and doctor's degrees from Columbia, the last in 1934. He has served for many years as an adviser on manpower to the White House and is the author or coauthor of more than 50 books on a variety of economic and labor issues. Ginzberg wishes to express his appreciation to Anna B. Dutka for her help in preparing the data used in his article.

GEORGE W. WETHERILL ("Apollo Objects") is director of the Department of Terrestrial Magnetism of the Carnegie Institution of Washington. He was educated at the University of Chicago, from which he obtained his Ph.D. in physics in 1953. He then joined the technical staff of the Department of Terrestrial Magnetism. In 1960 he was appointed professor of geophysics and geology at the University of California at Los Angeles, and he served from 1968 to 1972 as chairman of the U.C.L.A. department of planetary and space science. He returned to the Department of Terrestrial Magnetism in 1975.

GARTH L. NICOLSON ("Cancer Metastasis") is professor of cell biology and associate director of the oncology program at the University of California at Irvine. A graduate of the University of California at Los Angeles, he obtained his master's degree in biophysics at the University of Hawaii and his Ph.D. in cell biology from the University of California at San Diego. As a graduate student in the laboratory of S. J. Singer at San Diego, he helped to formulate the fluid-mosaic model of biological-membrane structure. After working for several years at the Salk Institute for Biological Studies he moved to Irvine in 1976. His current research is on the structure of the cell surface in normal and tumor cells. Nicolson has also worked as a professional diver and diving instructor.

THEODOR W. HÄNSCH, ARTHUR L. SCHAWLOW and GEORGE W. SERIES ("The Spectrum of Atomic Hydrogen") have utilized lasers for high-resolution optical studies of simple atoms. Hänsch is professor of physics at Stanford University. A native of West Germany, he was educated at the University of Heidelberg, which gave him his Ph.D. in physics summa cum laude in 1969. The following year he went to Stanford as a NATO Fellow, and in 1972 he joined the faculty. From 1973 to 1975 Hänsch was an Alfred P. Sloan Research Fellow. Schawlow is J. G. Jackson–C. J. Wood Professor of Physics at Stanford. Born in upstate New York, he received his Ph.D. from the University of Toronto in 1949. After two years as a postdoctoral fellow and research associate at Columbia University he became a research physicist at Bell Laboratories. He has been on the Stanford faculty since 1961. He was coauthor with Charles H. Townes of the first paper describing the laser. In 1977 Schawlow was awarded the third Marconi International Fellowship, which was presented in Stockholm by the king of Sweden. Series is professor of physics at the University of Reading. He studied at the University of Oxford, where he obtained his Ph.D. in atomic spectroscopy. He then became lecturer in physics at Oxford, moving to Reading in 1969. Series is a Fellow of the Royal Society.

ALAN L. GILCHRIST ("The Perception of Surface Blacks and Whites") is assistant professor of psychology at the State University of New York at Stony Brook. He grew up in the Pacific Northwest and received his bachelor's degree from Portland State College. He then came east to do graduate work at the Institute for Cognitive Studies at Rutgers University, obtaining his Ph.D. there in 1975. He writes: "My real interest lies in problems concerning the nature of the thinking process; I chose to study perception because I believe it currently represents the area of research closest to consciousness in which substantial progress can be made." Gilchrist wishes to acknowledge the National Science Foundation's support of his work.

RICHARD D. KEYNES ("Ion Channels in the Nerve-Cell Membrane") is professor of physiology at the University of Cambridge. He was born in London in 1919 and was educated at Trinity College, Cambridge. After spending World War II as a civilian scientist working on sonar and radar for the Royal Navy he returned to Cambridge in October, 1945, and took his B.A. in physiology the following June. Four years later he received his Ph.D.; the subject of his dissertation was the measurement of ion movements during nerve activity. Until 1960 he remained at Cambridge, first as a research fellow of Trinity College and later as a teaching fellow of Peterhouse. He then moved to the Agricultural Research Council Institute of Animal Physiology in Babraham, serving as its director from 1965 until 1973, when he was appointed professor of physiology at Cambridge. Keynes writes: "Since 1968 I have devoted a good deal of time to a book titled *The Beagle Record,* soon to be published by the Cambridge University Press. It contains a selection from the original pictorial and written records of the voyage of the *Beagle,* the aim being to produce something as close to an eyewitness account as is possible. As a person as well as a scientist, Darwin was quite outstanding, and his letters are superb. I have enjoyed getting to know my great-grandfather in this way."

BERND WÜRSIG ("Dolphins") is a postdoctoral fellow at the Center for Coastal Marine Studies of the University of California at Santa Cruz. He writes: "I am one of three sons of parents who are refugees from East Germany. We settled in Ohio in 1956, and although I did not get to touch seawater until 1969 (in Livorno, Italy, of all places), long before that time I developed a lively interest in marine biology and particularly in the whales, great and small. Later, at Ohio State University, I gave up a promising start in developmental biology simply because I knew I would never be quite happy in the laboratory." Würsig did his graduate work in cetacean behavior at the State University of New York at Stony Brook, obtaining his Ph.D. in 1978.

WERNER SOEDEL and VERNARD FOLEY ("Ancient Catapults") are on the faculty of Purdue University. Soedel is professor of mechanical engineering. He was born in Prague and belonged to the Sudeten German minority that was evicted to Germany in 1945. After attending school in a small village in southwestern Bavaria, he studied mechanical engineering at the State Engineering School in Frankfurt and received his first degree in 1957. He then worked as an automotive and development engineer with Opel, a General Motors subsidiary, until 1963. During this time he spent two years in the U.S. at the General Motors Technical Center. In 1963 he immigrated to the U.S. and attended Purdue, from which he obtained his Ph.D. degree in 1967 and where he subsequently joined the faculty. He teaches and does research in the area of theoretical and experimental mechanics. Soedel has been interested in the physics of catapults since his undergraduate days, when he saw shooting demonstrations of some of the reconstructed catapults at the Saalburg, a Roman fortification converted into a museum in the Taunus hills north of Frankfurt. Foley is associate professor of history at Purdue. He received his bachelor's degree at McPherson College, his master's degree at the University of Kansas and his Ph.D. from the University of California at Berkeley, finishing in 1970. Both graduate degrees were in European history with emphasis on the history of science and technology.

There may still be places where Grand Marnier isn't offered after dinner.

For free recipe booklet, write Carillon Importers, Ltd., 745 Fifth Avenue, New York, New York 10022. Product of France. Made with fine cognac brandy. 80 proof.

Jeep Wagoneer *Limited*

The ultimate in 4-wheel drive

Only one vehicle can boast that it has both the ultimate in 4-wheel drive performance and the ultimate in comfort and luxury. The Jeep Wagoneer Limited.

Not only does it handle magnificently off-road, the Limited travels on pavement effortlessly in all kinds of weather. And it does so beautifully, with such luxurious, standard features as real leather seats and extra thick carpeting for a quieter ride. Woodgrain trim accents both the inside and outside. To add to your convenience, an AM/FM radio with your choice of CB or stereo tape deck.

To add to your performance capabilities, Quadra-Trac,® Jeep's exclusive, automatic 4-wheel drive. And of course, many of the most wanted features are also standard.

Indulge yourself in the ultimate in 4-wheel drive—the Jeep Wagoneer Limited.

Jeep®

We wrote the book on 4-wheel drive

Fleet and leasing plans available. See your Jeep dealer for details. • Jeep Corporation, a subsidiary of American Motors Corp.

MATHEMATICAL GAMES

On altering the past, delaying the future and other ways of tampering with time

by Martin Gardner

"It is impossible to meditate on time and the mystery of the creative passage of nature without an overwhelming emotion at the limitations of human intelligence."

—ALFRED NORTH WHITEHEAD,
The Concept of Nature

There has been a great deal of interest among physicists of late in whether or not there are events on the elementary-particle level that cannot be time-reversed, that is, events for which imagining a reversal in the direction of motion of all the particles involved is imagining an event that cannot happen in nature. Richard Feynman has suggested an approach to quantum mechanics in which antiparticles are viewed as particles momentarily traveling backward in time. Cosmologists have speculated about two universes for which all the events in one are reversed relative to the direction of time in the other: in each universe intelligent organisms would live normally from past to future, but if the organisms in one universe could in some way observe events in the other (which many physicists consider an impossibility), they would find those events going in the opposite direction. It has even been conjectured that if our universe stops expanding and starts to contract, there will be a time reversal, but it is far from clear what that would mean. Most of the speculations of this kind are quite recent, and interested readers will find many of them examined in four chapters added to my *Ambidextrous Universe,* which is now back in print in a revised edition published by Charles Scribner's Sons.

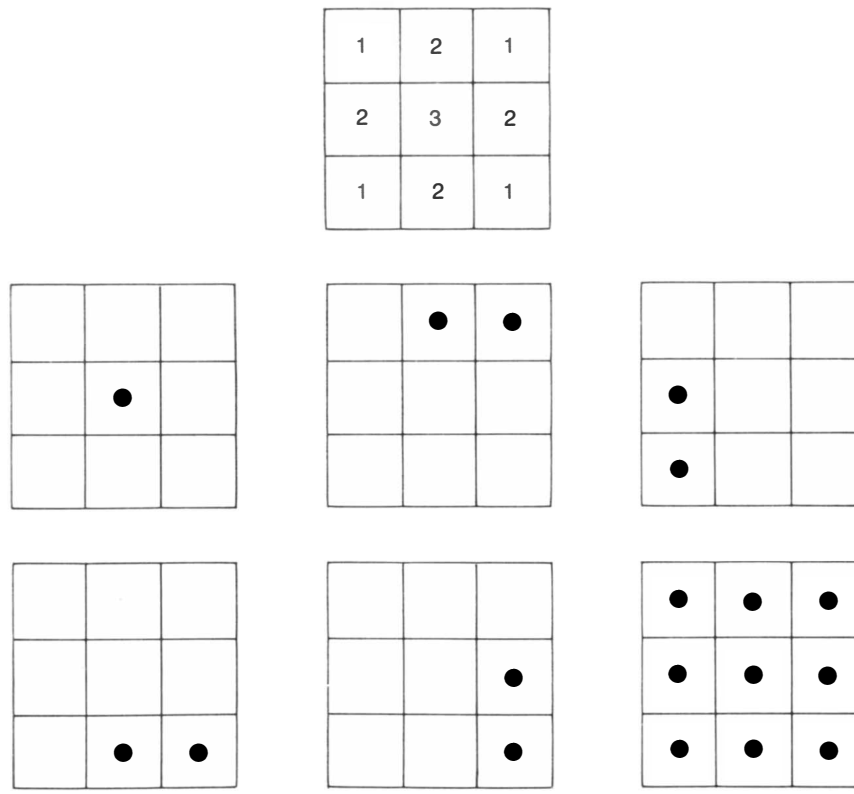
This month I shall consider two bizarre questions about time that are not discussed in the book. Indeed, these questions are of so little concern to scientists that only philosophers and writers of fantasy and science fiction have had much to say about them: Is it meaningful to speak of time stopping? Is it meaningful to speak of altering the past?

Neither question should be confused with the familiar subject of time's relativity. Newton believed the universe was pervaded by a single absolute time that could be symbolized by an imaginary clock off somewhere in space (perhaps outside the cosmos). By means of this clock the rates of all the events in the universe could be measured. The notion works well within a single inertial frame of reference such as the surface of the earth, but it does not work for inertial systems moving in relation to each other at high speeds. According to the theory of relativity, if a spaceship were to travel from our solar system to another solar system with a velocity close to that of light, events would proceed much slower on the spaceship than they would on the earth. In a sense, then, such a spaceship is traveling through time into the future. Passengers on the spaceship might experience a round-trip voyage as taking only a few years, but they would return to find that centuries of earth-years had elapsed.

The notion that different parts of the universe can change at different rates of time is much older than the theory of relativity. In the Scholastic theology of the Middle Ages angels were considered to be nonmaterial intelligences living by a time different from that of earthly creatures; God himself was thought to be entirely outside of time. In the first act of Lord Byron's play *Cain, A Mystery,* the fallen angel Lucifer says:

With us acts are exempt from time,
and we
Can crowd eternity into an hour
Or stretch an hour into eternity·
We breathe not by a mortal
measurement—
But that's a mystery.

In the 20th century hundreds of science-fiction stories have played with the relativity of time in different inertial systems, but the view that time can speed up or slow down in different parts of our universe is central to many older tales. A popular medieval legend tells of a



Winning strategies for three-by-three turnablock

monk who is entranced for a minute or two by the song of a magical bird. When the bird stops singing, the monk discovers that several hundred years have passed. In a Moslem legend Mohammed is carried by a mare into the seventh heaven. After a long visit the prophet returns to the earth just in time to catch a jar of water the horse had kicked over before starting its ascent.

Washington Irving's "Rip Van Winkle" is this country's best-known story about someone who sleeps for what seems to him to be a normal time while two decades of earth-years rush by. King Arthur's daughter Gyneth slept for 500 years under a spell cast by Merlin. Every culture has similar sleeper legends. H. G. Wells used the device in *When the Sleeper Wakes,* and it is a common practice in science fiction to put astronauts into a cryogenic sleep so that they can survive interstellar voyages that are longer than their normal life span. In Wells's short story "The New Accelerator" a scientist discovers a way to speed up a person's biological time so that the world seems to come almost to a halt. This device too is frequently encountered in later science fiction.

1	2	1	4	1	2	1	8
2	3	2	8	2	3	2	12
1	2	1	4	1	2	1	8
4	8	4	6	4	8	4	11
1	2	1	4	1	2	1	8
2	3	2	8	2	3	2	12
1	2	1	4	1	2	1	8
8	12	8	11	8	12	8	13

Strategy chart for turnablock

PERSISTENCE	NUMBER
1	10
2	25
3	39
4	77
5	679
6	6,788
7	68,889
8	2,677, 889
9	26,888,999
10	3,778,888,999
11	277,777,788,888,899

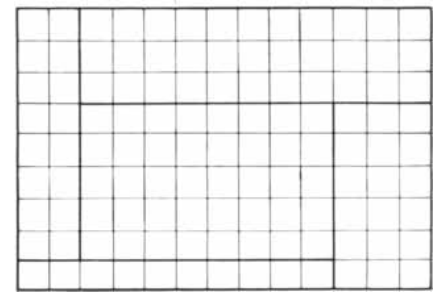
Smallest numbers with persistence 11 or less

The issue under consideration here, however, is not how time can vary but whether time can be said to stop entirely. It is clearly meaningful to speak of all motion ceasing in one part of the universe, whether or not such a part exists. In the theory of relativity the speed of light is an unattainable limit for any object with mass. If a spaceship could attain the speed of light (which the theory of relativity rules out because the mass of the ship would increase to infinity), then time on the spaceship would stop in the sense that all change on it would cease. In earth time it might take 100 years for the spaceship to reach a destination, but to astronauts on the spaceship the destination would be reached instantaneously. One can also imagine a piece of matter or even a human being reduced to such a low temperature (by some as yet unknown means) that even all subatomic motions would be halted. For that piece of matter, then, one could say that time had stopped. Actually it is hard to understand why the piece of matter would not vanish.

The idea of time stopping creates no problems for writers of fantasy, who are not constrained by the real world. For example, in L. Frank Baum's "The Capture of Father Time," one of the stories in his *American Fairy Tales* (which is now back in print in a Dover facsimile edition), a small boy lassoes Time, and for a while everything except the movements of the boy and Father Time stops completely. In Chapter 22 of James Branch Cabell's *Jurgen: A Comedy of Justice,* outside time sleeps while Jurgen enjoys a pleasurable stay in Cocaigne with Queen Anaïtis. Later in the novel Jurgen stares into the eyes of the God of his grandmother and is absolutely motionless for 37 days. In Jorge Luis Borges' story "The Secret Miracle" a writer is executed by a firing squad. Between the command to fire and the writer's death God stops all time outside the writer's brain, giving him a year to complete his masterpiece.

Many similar examples from legend and literature show that the notion of time stopping in some part of the universe is not logically inconsistent. But what about the idea of time stopping throughout the universe? Does the notion that everything stops moving for a while and then starts again have any meaning?

If it is assumed that there is an outside observer—perhaps a god—watching the universe from a region of hypertime, then of course the notion of time stopping does have meaning, just as imagining a god in hyperspace gives meaning to the notion of everything in the universe turning upside down. The history of our universe may be like a three-dimensional motion picture a god is enjoying. When the god turns off the projector to do something else, a few millenniums may go by before he comes back and turns it on again. (After all, what are a few millenniums to a god?) For all we can know a billion centuries of hypertime may have elapsed between my typing the first and the second word of this sentence.



Rectangled rectangle of smallest area: 117

Suppose, however, all outside observers are ruled out and "universe" is taken to mean "everything there is." Is there still a way to give a meaning to the idea of time stopping for a while? Although most philosophers and scientists would say there is not, a few have argued for the other side. For example, in "Time without Change" (*The Journal of Philosophy,* Vol. 66, No. 12, pages 363–381; June 19, 1969) Sydney S. Shoemaker, now a philosopher at Cornell University, makes an unusual argument in support of the possibility of time stopping.

Shoemaker is concerned not with the real world but with possible worlds designed to prove that the notion of time stopping everywhere can be given a reasonable meaning. He proposes several worlds of this kind, all of them based on the same idea. I shall describe only one such world here, in a slightly dramatized form.

Imagine a universe divided into regions *A, B* and *C.* In normal times inhabitants of each region can observe the inhabitants of the other two and communicate with them. Every now and then, however, a mysterious purple glow permeates one of the regions. The glow always lasts for a week and is invariably followed by a year in which all change in the region ceases. In other words, for one year absolutely nothing happens there. Shoemaker calls the phenomenon a local freeze. Since no events take place, light cannot leave the region, and so the region seems to vanish for a year. When it returns to view, its inhabitants are unaware of any passage of time, but they learn from their neighbors that a year, as measured by clocks in the other two regions, has elapsed. To the inhabitants of the region that experienced the local freeze it seems that instantaneous changes have taken place in the other two regions. As Shoemaker puts it: "People and objects will appear to have

TWINCH is here. Keep it to yourself.

You don't have to share JVC's TWINCH — the two-inch screen TV with FM/AM radio — with anyone. Unless, of course, you want to.

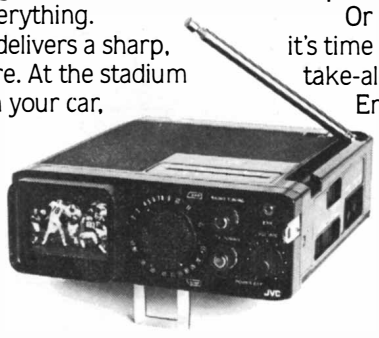
Take "The Big Game." You show up with your new TWINCH, which is about the size and weight of a good book. You pull up the antenna, flick the instant battery power on and bingo — you're watching the game at the game. Instant replays and everything.

The rugged TWINCH delivers a sharp, steady black and white picture. At the stadium or a beach, out camping or in your car, absolutely wherever you are.

But when you're in a sharing mood, turn on the built-in FM/AM radio. Sophisticated microelectronics provide the clean, crisp sound you expect from JVC. (After all, we're known for innovating in high fidelity.)

So find your nearest JVC Home Entertainment Dealer by calling TOLL-FREE 800-221-7502. In New York State, call (212) 476-8300. Then go for the versatile TWINCH and keep it to yourself.

Or better yet, pass the word. Maybe it's time everybody got TWINCH, the take-along TV/Radio. From JVC Home Entertainment Division, Maspeth, NY 11378. TV reception simulated



JVC®

US JVC Corp.

moved in a discontinuous manner or to have vanished into thin air or to have materialized out of thin air; saplings will appear to have grown instantaneously into mature trees, and so on."

In the history of each of the three regions local freezes, invariably preceded by a week of purple light, have happened thousands of times. Now suppose that suddenly, for the first time in history, purple light appears simultaneously in regions *A, B* and *C* and lasts for a week. Would it not be reasonable, Shoemaker asks, for scientists in the three regions to conclude that time had stopped for a year (whatever that means) throughout the entire universe even though no minds were aware of it?

Shoemaker considers several objections to his thesis and counters all of them ingeniously. Interested readers can consult his paper and then read a technical analysis of it in the fifth chapter of G. Schlesinger's *Confirmation and Confirmability* (Oxford University Press, 1974). Schlesinger agrees with Shoemaker that an empirical, logically consistent meaning can be found for the sentence "A period of time *t* has passed during which absolutely nothing happened." Note that similar arguments about possible worlds can provide meanings for such notions as everything in a universe turning upside down, mirror-reversing, doubling in size and so on.

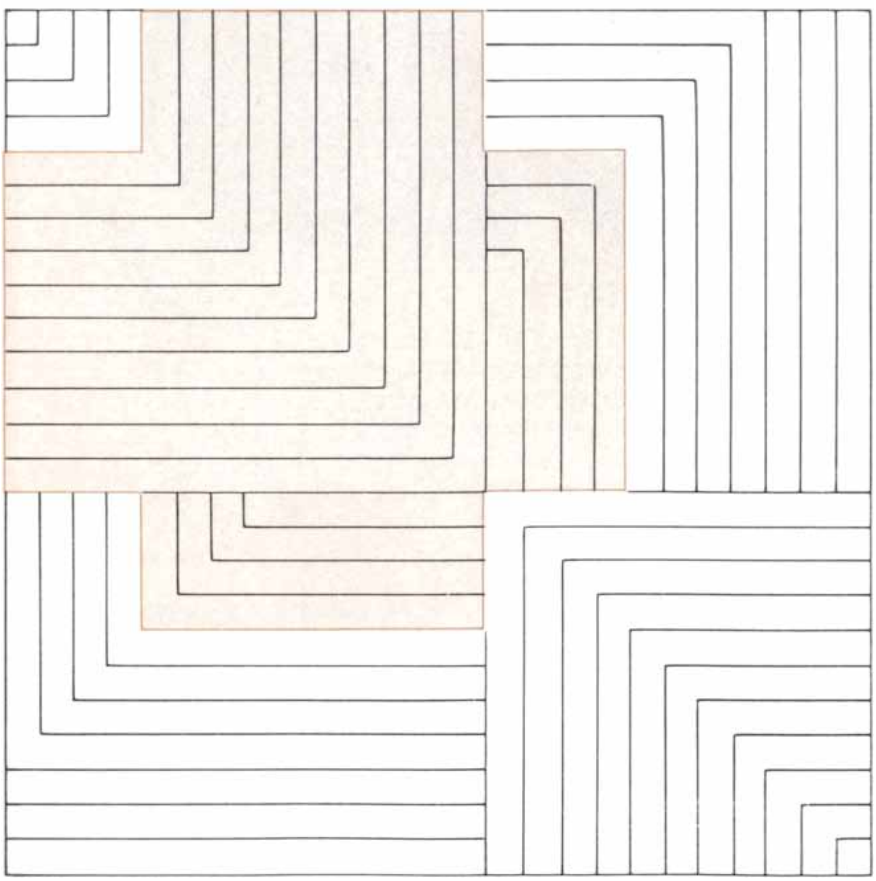
The question of whether the past can be changed is even stranger than that of whether time can stop. Writers have often speculated about what might have happened if the past had taken a different turn. J. B. Priestley's play *Dangerous Corner* dealt with this question, and there have been innumerable "what if" stories in both science fiction and other kinds of literature. In all time-travel stories where someone enters the past the past is necessarily altered. The only way the logical contradictions created by such a premise can be resolved is by positing a universe that splits into separate branches the instant the past is entered. In other words, while time in the old branch "gurgles on" (a phrase from Emily Dickinson) time in the new branch gurgles on in a different way toward a different future. When I speak of altering the past, however, I mean altering it throughout a single universe with no forking time paths. (Pseudoalterations of the past, such as the rewriting of history satirized by George Orwell in *1984,* obviously do not qualify.) Given this context, can an event, once it has happened, ever be made not to have happened?

The question is older than Aristotle, who in his *Ethics* (Book 6) writes: "It is to be noted that nothing that is past is an object of choice, for example, no one chooses to have sacked Troy; for no one *deliberates* about the past, but about what is future and capable of being otherwise, while what is past is not capable of not having taken place; hence Agathon is right in saying: 'For this alone is lacking even to God, to make undone things that have once been done.'"

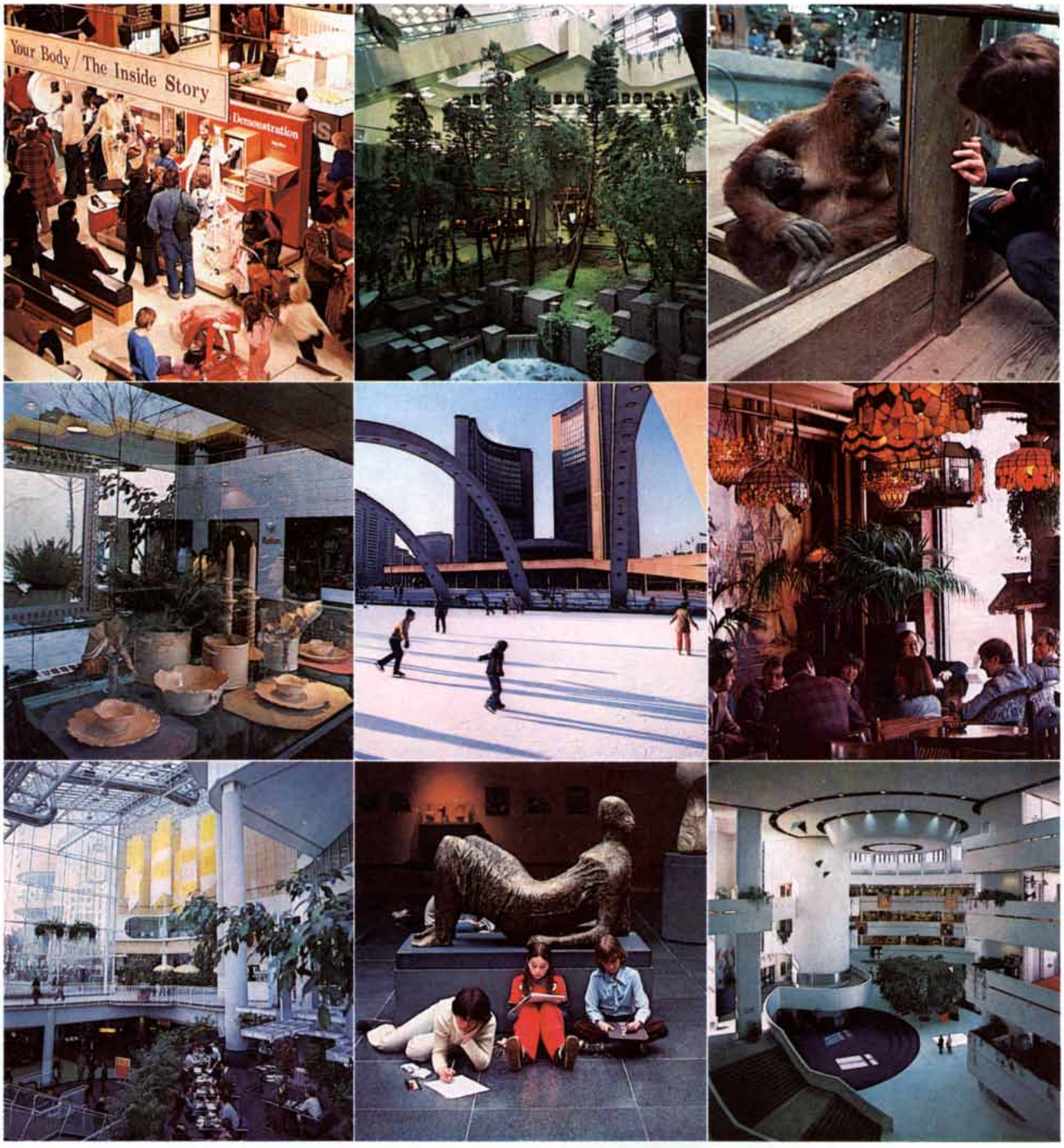
Thomas Aquinas believed God to be outside of time and thus capable of seeing all his creation's past and future in one blinding instant. (Even though human beings have genuine power of choice, God knows how each one will choose; it is in this way that Aquinas sought to harmonize predestination and free will.) For Aquinas it was not possible for God to do absolutely impossible things, namely those that involve logical contradiction. For example, God could not make a creature that was both a human being and a horse (that is, a complete human being and a complete horse, rather than a mythical combination of parts such as a centaur), because that would involve the contradiction of assuming a creature to be simultaneously rational and nonrational. Similarly, God could not alter the past. That would be the same as asserting that the sack of Troy both took place and did not take place. Aquinas agreed with Aristotle that the past must forever be what it was, and it was this view that became the official position of medieval Scholasticism. It is not so much that God's omnipotence is limited by the law of contradiction but rather that the law is part of God's nature. "It is best to say," Aquinas wrote, "that what involves contradiction cannot be done rather than that God cannot do it."

Edwyn Bevan, in a discussion of time in his book *Symbolism and Belief,* finds it odd that Aquinas would deny God the ability to alter the past and at the same time allow God to alter the future. In the 10th question of *Summa Theologica* (Ia. 10, article 5.3), Aquinas wrote: "God can cause an angel not to exist in the future, even if he cannot cause it not to exist while it exists, or not to have existed when it already has." For Aquinas to have suggested that for God the past is unalterable and the future is not unalterable, Bevan reasons, is surely to place God in some kind of time, to contradict the assertion that God is outside of time.

I know of no scientist or secular philosopher who has seriously believed the past could be altered, but a small minority of theologians have maintained that it could be. The greatest of them was Peter Damian, the zealous Italian reformer of the Roman Catholic church in the 11th century. In *On Divine Omnipotence,* his most controversial treatise, Damian argued that God is in no way bound by the law of contradiction, that his omnipotence gives him the power to do all contradictory things including changing the past. Although Damian,



Solution to one of last month's problems



Toronto–the inside story

All the year round, whatever the weather, Toronto extends the warmest of welcomes–indoors. Take a swing through the jungle in our brand new zoo. Discover the push-button wonders of the Science Centre. Stroll our glittering underground city, sleek new malls or chic little shops. Spend a day in the galleries, a night on the town, theatre, cabarets, opera, ballet and all that jazz. Savour cosmopolitan cuisine and the creature comforts of first-class hotels–we treat you royally. This year you'll enjoy a substantial premium on every dollar–and the removal of Ontario's 7% room tax is an added plus. For your Traveller's Encyclopaedia call us COLLECT (416) 965-4008, or write: Ontario Travel, Dept. I.S., Queen's Park, Toronto, M7A 2E5.

Ontario Canada

LOVE LETTERS TO A CAR COMPANY?

You'd probably never dream of writing a love letter to the company that built your car. Yet at Volvo, we get them all the time.

Maybe because 9 out of 10 people who buy new Volvos are happy.

But some people who own Volvos aren't merely happy. They're absolutely ecstatic.

People like Orv Jacobson of Ocean Shores, Washington: "Now that it's at the 200,000 mile mark, I figure it has another 200,000 more miles in it...you just don't trade in members of your family."

Mary Howard of West Trenton, New Jersey: "This is the first time in my life that I still had a good car after it was paid for."

Or, The Reverend Ernest G. Olsen of Westfield, New Jersey who, after 225,000 trouble-free miles, says simply: "Love that Volvo."

If you can't think of any love stories to write about your present car, maybe your next car should be a Volvo.

After all, why buy a car you might regret? When with a Volvo, you'll probably end up never having to say you're sorry.

VOLVO

A car you can believe in.

who started out as a hermit monk, argued his extreme views skillfully, he regarded all reasoning as superfluous, useful only for supporting revealed theology. It appears that he, like Lewis Carroll's White Queen, would have defended everyone's right to believe six impossible things before breakfast. (Damian was also a great promoter of self-flagellation as a form of penance, a practice that became such a fad during his lifetime that some monks flogged themselves to death.)

One of my favorite Lord Dunsany stories is the best example I know of from the literature of fantasy that illustrates Damian's belief in the possibility of altering the past. It is titled "The King That Was Not," and you will find it in Dunsany's early book of wonder tales *Time and the Gods*. It begins as follows: "The land of Runazar hath no King nor ever had one; and this is the law of the land of Runazar that, seeing that it hath never had a King, it shall not have one for ever. Therefore in Runazar the priests hold sway, who tell the people that never in Runazar hath there been a King."

The start of the second paragraph is surprising: "Althazar, King of Runazar...." The story goes on to recount how Althazar ordered his sculptors to carve marble statues of the gods. His command was obeyed, but when the great statues were undraped, their faces were very much like the face of the king. Althazar was pleased and rewarded his sculptors handsomely with gold, but up in Pegāna (Dunsany's Mount Olympus) the gods were outraged. One of them, Mung, leaned forward to make his sign against Althazar, but the other gods stopped him: "Slay him not, for it is not enough that Althazar shall die, who hath made the faces of the gods to be like the faces of men, but he must not even have ever been."

"Spake we of Althazar, a King?" asked one of the gods.

"Nay, we spake not."

"Dreamed we of one Althazar?"

"Nay, we dreamed not."

Below Pegāna, in the royal palace, Althazar suddenly passed out of the memory of the gods and so "became no longer a thing that was or had ever been." When the priests and the people entered the throne room, they found only a robe and a crown. "The gods have cast away the fragment of a garment," said the priests, "and lo! from the fingers of the gods hath slipped one little ring."

Here are the answers to last month's problems:

1. The following five moves will suffice to get all four glasses in the wells of the rotating table turned either all up or all down.

(1) Reach into any diagonal pair of wells. If the glasses are not both turned up, adjust them so that they are. If the bell does not ring:

(2) Spin the table and reach into any adjacent pair of wells. If both glasses are turned up, leave them that way; otherwise invert the glass that is turned down. If the bell fails to ring, you know that now three glasses are turned up and one is turned down.

(3) Spin the table and reach into any diagonal pair of wells. If one of the glasses is turned down, invert it and the bell will ring. If both are turned up, invert one so that the glasses are arranged in the following pattern:

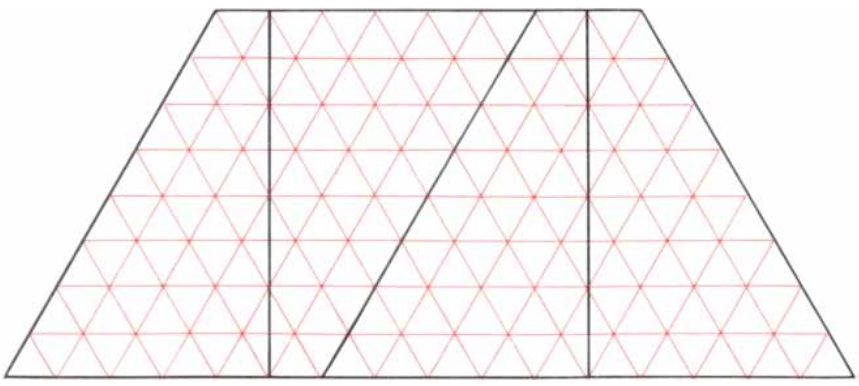
Up	Down
Up	Down

(4) Spin the table and reach into any adjacent pair of wells. Invert both glasses. If they were both turned in the same direction, the bell will ring; otherwise the glasses are now arranged in the following pattern:

Up	Down
Down	Up

(5) Spin the table, reach into any diagonal pair of wells and invert both glasses. The bell will ring.

Ronald L. Graham of Bell Laboratories and Persi Diaconis of Stanford University have together considered two ways of generalizing the problem. One



Triamond divided into four identical regions

is to assume that the player has more than two hands. If the number of glasses n is greater than 4, the problem is solvable with $n - 2$ hands if and only if n is not a prime number. Therefore for five glasses (5 is a prime) the problem is not solvable with $5 - 2$, or 3, hands. It may be that if n is a composite number, then for some values of n the problem can be solved with fewer than $n - 2$ hands. The problem can also be generalized by replacing glasses with objects that have more than two positions. Hence the rotating table leads into deep combinatorial questions that as far as I know have not yet been explored.

2. The first player's strategy for winning John Horton Conway's order-three game of turnablock requires numbering the nine cells as is shown in the illustration on page 21. The player must only make moves that leave black counters on cells whose numbers have a "nim sum" of zero. The nim sum of a collection of numbers is traditionally obtained by writing the numbers in binary form and adding without carrying. If all the digits in the total are zero, then the nim sum is zero. Conway suggests a simpler procedure for finding a nim sum: express each of the numbers to be added as a sum of distinct powers of 2, and then cancel pairs of like powers. If no powers remain, then the nim sum is zero; otherwise the nim sum is simply the sum of the remaining powers. For example, consider $1 + 5 + 12$. Writing each number in this expression as the sum of distinct powers of 2 gives $1 + 4 + 1 + 8 + 4$. The pairs of 1's and 4's cancel, leaving a nim sum of 8. Similarly, consider $1 + 2 + 5 + 6$. This expression becomes $1 + 2 + 4 + 1 + 4 + 2$. All the powers cancel, and the nim sum is zero.

The dots in the illustration on page 21 indicate the blocks that can be turned over for a winning first move. In other words, after each of these moves black counters are left on cells whose numbers have a nim sum of zero. Any second move by the other player will necessarily leave black counters on cells with a nonzero nim sum. The first player will always be able to respond with a zero nim-sum move, and by continuing in this way he is sure to win.

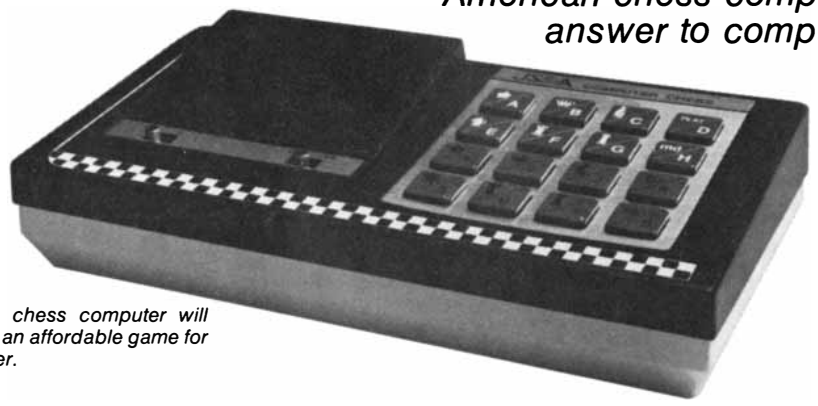
The upper illustration at the left on page 22, supplied by Conway, gives the cell numbering for all rectangular turnablock boards with sides from one through eight. The numbering of an a-by-b board smaller than the order-eight board is given by the a-by-b rectangle in the upper left-hand corner of the matrix. The numbers in the cells are obtained by a process called nim multiplication, which you will find explained in Conway's *On Numbers and Games* (Academic Press, 1976, page 52).

3. The smallest number in base 10 of multiplicative persistence 5 is 679. The chart in the lower illustration at the left

Mr. Karpov and our computer.

Karpov Accepts

Soviet chess champion agrees to endorse American chess computer in surprise answer to computer challenge.



The new $100 JS&A chess computer will make computer chess an affordable game for the American consumer.

Anatoli Karpov is the undisputed World Chess Champion. Last month, JS&A challenged Karpov to play against our chess computer. We offered him $50,000 and a percentage of each unit we sold if he beat our game.

Karpov rejected our offer and quite frankly, we were at a loss to explain why.

Our concept was simple. If Karpov played against our computer, he would focus worldwide attention on our product. This attention would increase its sales and win or lose, we would sell more computers.

We had to sell more computers. We wanted our unit to sell for $100 even though units with similar capabilities were selling for up to $400. But we had to do two things to keep our low price: First, we had to manufacture our unit in Hong Kong where labor costs are very low. Secondly, we had to sell large quantities to keep production costs down.

When we first announced our challenge, the Soviets rejected it. Could Karpov have been afraid to play against it? Or was the circus-like atmosphere that might surround the event not in keeping with the prestige of the coveted world title, even if Karpov won?

Quite honestly, we had no idea. So we asked Karpov for an explanation and got one. Here's what he said:

"I can appreciate your desire to sell the JS&A Chess Computer but you have missed an important point of the entire challenge. It took somebody to program the unit and since I am the world's chess champion, it is highly unlikely that the programmer person could beat me. You Americans have a saying, "Garbage in, garbage out."

Karpov continued, "Your unit is definitely a good product. I played several games at level six and found it to be a challenge even for me. What I like about your unit is that it is priced low enough so that most Americans can afford the unit and this will help promote chess."

The JS&A Chess Computer indeed has six levels of chess. Level one is perfect for beginners. Level six is a real challenge for any Soviet Chess Champion.

LIKE PLAYING KARPOV

The system is the perfect way to sharpen your chess skills. It not only has six different skill levels, but if you are playing against the computer at level two and you are beating it, you can switch the unit to level six. It's like having Karpov as your new opponent–right during mid game.

To play against the computer, you enter your move on the unit's keyboard. You then wait until the computer examines all its options and selects its move. You then move the computer's chess piece to correspond with its request as shown on the display. A board layout is provided to show you where each chess piece should be moved.

The JS&A Chess Computer is programmed for such complex functions as castling, pawn promotion and *enpessant*. It also allows you to start in midgame, setting up any situation you choose. This is perfect for those players who wish to examine particularly intricate problems and allows for an infinite number of game variations.

SHARPEN SKILLS

If you already play chess, the JS&A unit provides a new chess dimension. If you haven't played chess, the system is a good way to learn and sharpen your skills.

The JS&A Chess Computer measures only 2⅛"x 4⅞"x 8⅞" and weighs just a few ounces, so if service is ever required you can slip it in its handy mailer and send it back to our prompt service-by-mail center. Service should never be required, but it is reassuring to know that service is an important consideration in this program.

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

We suggest you order a JS&A Chess Computer and use it for 30 days. Play against it. Raise or lower the level as you play and watch how the computer's personality can change right in mid game–from a tough competitor to a push over.

Test our level six and see if you'd have much of a chance against the Soviet Champion Karpov. Then, after you've really given it a workout, decide if you want to keep it. If not, you may return your unit within our 30 day trial period for a prompt and courteous refund, including your $2.50 postage and handling charge. There is no risk. Each JS&A Chess Computer comes complete with instructions and an AC adapter (no batteries are required).

To order our JS&A Chess Computer, send your check for **$99.95** plus $2.50 for postage and handling (Illinois residents please add 5% sales tax) to the address below or credit card buyers may call our toll-free number below.

Karpov told us (and this is a direct quote) "I have played all the rest of the American chess computers and find that the JS&A unit is the best value for the dollar. I will accept an offer to endorse the game, however, you can forget about your $50,000 offer and spend the money on advertising to spread the word."

Thank you, Mr. Karpov. We appreciate your honesty and your generosity and we promise to spread the word.

And you can help us America. Order a JS&A Chess Computer at no obligation, today.

JS&A® PRODUCTS THAT THINK®

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

on page 22 gives the smallest numbers with multiplicative persistence of 11 or less. It is taken from N. J. A. Sloane's paper "The Persistence of a Number," which appeared in *Journal of Recreational Mathematics* (Vol. 6, No. 2, pages 97–98; Spring, 1973).

The smallest number in base 10 of additive persistence 4 is 19,999,999,999,999,999,999,999. The sum of the digits in this number is 199, the smallest number of additive persistence 3. The sum of the digits in 199 is 19, the smallest number of additive persistence 2. The sum of the digits in 19 is 10, the smallest number of additive persistence 1. More generally, the second step in the sequence for the smallest number of additive persistence k gives the smallest number of additive persistence $k-1$. All such numbers start with 1 and are followed by 9's. Therefore the smallest number of additive persistence 5 is the number consisting of 1 followed by 2,222,222,222,222,222,222,222 9's. For further discussion of the problem see Harvey J. Hinden's "The Additive Persistence of a Number" (*Journal of Recreational Mathematics,* Vol. 7, No. 2, pages 134–135; Spring, 1974).

4. The first parody of Edgar Allan Poe's "The Raven" is by Howard W. Bergerson, and it is called an automynorcagram. The first letters of each word in the parody, taken in the order they appear, spell out the first verse of the parody. For the complete parody see *Word Ways* (Vol. 8, No. 4, pages 219–222; November, 1975).

The second parody of Poe's poem is by A. Ross Eckler, editor and publisher of *Word Ways.* It is homoliteral, that is, each consecutive pair of words have at least one letter in common. The full parody appears in *Word Ways* (Vol. 9, No. 2, pages 96–98; May, 1976).

The third parody, also by Eckler, is found in *Word Ways* (Vol. 9, No. 4, pages 231–233; November, 1976). It is heteroliteral, that is, each pair of consecutive words has no letter in common.

5. The illustration at the right on page 22 displays the rectangle of smallest area (117) that can be cut into smaller rectangles with integral sides so that no two sides of any pair of internal rectangles are alike. The pattern is unique.

6. (1) The solution is shown in the illustration on page 24. The puzzle is adapted from an optical illusion in Charles H. Paraquin's *Eye Teasers: Optical Illusion Puzzles* (Sterling Publishing Co., Inc., 1978).

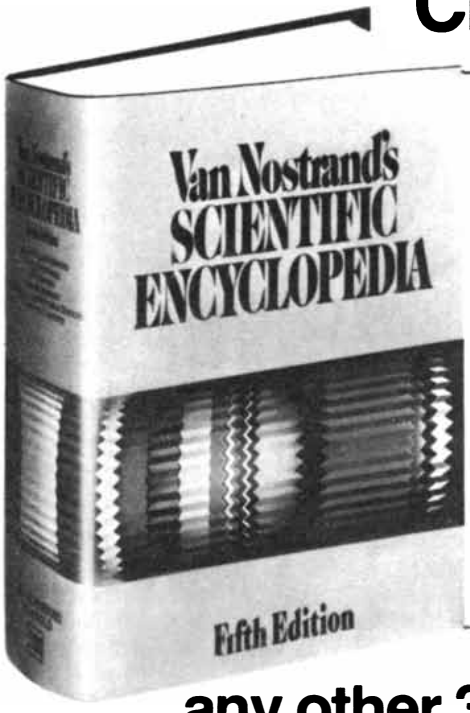
(2) The rule for obtaining the third pattern in each row is to superpose the first two patterns and eliminate any lines they have in common. Hence the pattern to be placed at the end of the third row is simply a square.

(3) The illustration on page 27 displays the only other way the triamond can be cut into four identical regions.

As your introduction to

Natural Science Book Club

Choose either



this $67.50 Classic for only $2.97 A saving of 95%

VAN NOSTRAND'S SCIENTIFIC ENCYCLOPEDIA.

Edited by Douglas M. Considine. Nearly 200 experts have contributed to this thoroughly revised and greatly expanded fifth edition of the most authoritative single-volume source of scientific knowledge ever assembled. Enormous 9¼" x 12" volume contains 2.2 million words, 2382 pages, 2500 photographs, drawings, and charts, and 500 tables. 7200 articles cover everything from the earth and space sciences to energy technology, from the life sciences to mathematics, from information sciences to physics and chemistry. " an amazing book . . for both the general and scientific reader. " —*The New York Times*

OR

any other 3 books for only 99¢ each

if you will join now for a trial period and agree to take 3 more books—at handsome discounts—over the next 12 months

(values to $70.30)

45050. THE ENDS OF THE EARTH. *Isaac Asimov.* Everything about the frozen polar wastes. **$15.00**

58900. LOOKING AT ANIMALS. *Hugh B. Cott.* Explores through word and picture the world's richest wildlife area—East Africa. **$15.00**

55547. INSIDE THE ANIMAL WORLD: An Encyclopedia of Animal Behavior. *Maurice & Robert Burton.* An elegantly illustrated, comprehensive survey of every facet of animal activity. **$17.50**

69420. PLANTCRAFT. *Richard Mabey.* A guide to the many and varied uses of wild plants. **$10.00**

77000. SCIENTIFIC ELITE: Nobel Laureates in the United States. *Harriet Zuckerman.* **$14.95**

61300. MATHEMATICAL MAGIC SHOW. *Martin Gardner.* Delightful collection of brain teasers and inspired dissertations. **$8.95**

38900. A CLOSER LOOK. *Michael A. Godfrey.* A look at the ecology of a typical backyard. **$14.95**

87890-2. WORLD PREHISTORY IN NEW PERSPECTIVE. *Grahame Clark.* Tour of man's pivotal rise from the stone age to the dawn of civilization. *Counts as 2 of your 3 books.* **$29.95**

52850. THE HIGH FRONTIER: Human Colonies in Space. *Gerard K. O'Neill.* **$8.95**

48485. FIELD PHOTOGRAPHY: Beginning and Advanced Techniques. *Alfred H. Blaker.* Tricks-of-the-trade on how to obtain publication-grade nature photos. **$22.50**

32480. AIRBORNE. *William F. Buckley, Jr.* Cross-Atlantic sailing trip made in 1975 by Buckley, his son and crew. **$12.95**

85780-2. VOLCANO. *Maurice and Katia Kraft.* Word-and-picture masterpiece of the many varied moods of volcanoes. *Counts as 2 of your 3 books.* **$35.00**

73939. THE RESTLESS UNIVERSE/WHITNEY STAR FINDER. A highly acclaimed 450-page tour of the ideas, principles, and process of modern astronomy. *Plus* a topnotch field guide to the night sky. *The 2 count as one book.* **$21.90**

82771. TEMPLES, TOMBS AND HIEROGLYPHS/RED LAND, BLACK LAND. Two highly acclaimed works on Egyptian history. *The 2 count as one book.* **$25.90**

54985. THE ILLUSTRATED ENCYCLOPAEDIA OF THE MINERAL KINGDOM. **$20.00**

55000. THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. *Ian Roxburgh.* **$16.95**

84870. TUTANKHAMUN: The Last Journey. *William MacQuitty.* The legendary treasures of the boy king. **$10.95**

36395. BLACK HOLES, QUASARS AND THE UNIVERSE. *Henry L. Shipman.* An absorbing exploration of the frontiers of astronomy. **$15.50**

61610. MAYA. *Charles Gallenkamp.* The rise of Mayan civilization and its march to extinction. **$12.95**

54995. THE ILLUSTRATED ENCYCLOPEDIA OF ARCHAEOLOGY. *Glyn Daniel.* 1000 entries. More than 100 illustrations **$17.95**

44330. THE EMERGENCE OF SOCIETY. *John Pfeiffer.* Offers a fascinating account of man's earliest religions, hierarchies, and cities. **$15.00**

34730. THE ATLAS OF EARLY MAN. *Jacquetta Hawkes.* A panoramic overview of concurrent events in history through charts, graphs, and illustrations, from prehistory to the Roman Empire. **$20.00**

63340-2. MYSTERIES OF THE PAST. Beautifully illustrated volume explores the riddles of past civilizations. *Counts as 2 of your 3 books.* **$34.95**

43085. EARTH'S AURA. *Louise B. Young.* Grand tour of the atmosphere. Fully illustrated. **$12.95**

55125. IN THE DESERTS OF THIS EARTH. *Uwe George.* Lavishly illustrated look at every aspect of deserts. **$14.95**

74590. THE ROOTS OF CIVILIZATION. *Alexander Marshack.* A remarkable re-creation of the life of prehistoric man. Illustrated. **$17.50**

(Publishers' Prices shown)

If the reply card has been removed, please write to
Natural Science Book Club
Dept. 4-A5E Riverside, N.J. 08370
to obtain membership information and an application

BOOKS

Human sexuality and birth, Szilard and Einstein, light-water reactors, biogeography, cosmography

by Philip Morrison

PATTERNS OF SEXUALITY AND REPRODUCTION, by Alan S. Parkes. Oxford University Press ($7.25). LYING-IN: A HISTORY OF CHILDBIRTH IN AMERICA, by Richard W. Wertz and Dorothy C. Wertz. The Free Press, a division of the Macmillan Publishing Co., Inc. ($10). Primates with a difference, our species has a remarkable reproductive biology even if we disregard the social phenomena that so strongly mold reproductive behavior. *Patterns of Sexuality and Reproduction,* an easily read yet authoritative set of essays by a distinguished London specialist in the subject, gives a clear account of the reproductive life cycle. Its focus is not on anatomy or even physiology but rather on the variations and variables revealed by clinical and statistical research, with a few less than compelling analogies from the study of laboratory animals. The nine chapters are rich with data in tables and graphs, a sampling of which will offer a fair account of the contents.

There are the *S* curves that show the age at which the changes of puberty appear in boys and girls. Fifty percent of a sample of American boys acquire pubic hair by the age of 13 and a half (Kinsey data from 30 years ago); 50 percent of English girls have begun to menstruate by 13 (1962 data). The graph gives data for English girls back to 1845, when the same point was reached more than two years later in life. Both the spread in the age of menarche and its mean value are dropping steadily in English samples, generation after generation. The length of the menstrual cycle is plotted for a sample of about 500 in England not long ago; the mean is about 28 days, variability is high and the peak is broad.

A strange result, quite inexplicable at the moment, is shown by comparing the protein content of the milk of American and British women. The spread among individuals is about the same, but British women produce on the average a milk that is richer in protein by almost a third. Some human females yield milk in great excess of a single infant's needs (although "these records are all more than 40 years old"). In spite of Dr. Cook's old claims, there is good evidence that Eskimo women in Greenland do not restrict their conception to the daylight months; yet it is plain that in England, Wales and Bavaria, for example, nowadays one can count on a 15 percent rise in conceptions during the "springtime, the only pretty ringtime," with a secondary peak in December. Strangely, the U.S. shows a reversed trend, as does New Zealand; the birth-rate in Puerto Rico shifted from the European to the American pattern within 20 years after 1941. Surely this is a socially mediated phenomenon, not a response to day length or other simple biology.

The story of twinning, of both the one- and the two-egg kind, is well presented. So is the complex enigma of the sex ratio, which is now about 1.1 or 1.2 males per female at conception and a somewhat smaller excess of males at birth. Once differential infant mortality brought this number down so sharply that there was a considerable excess of females by the early marriageable ages; now equality is reached only in the age group from 30 to 34. "We are heading for a world shortage of marriageable females." Women outnumber men by more than two to one among people over 80. The most remarkable graph is the record (found "buried in a book on the pulse-rate") showing a simultaneous pulse-rate recording for a man and a woman during an hour that included intercourse. Sir Alan does not hide his opinions, which are a little wry; the strength of this book is not in its social interpretation. It is nonetheless full of interest, a real resource.

Reproduction requires, of course, that a woman bear a live infant. That event too is much more than biological. *Lying-In* is a social chronicle of the circumstances of childbirth in the U.S. from the 17th century to the present. Its aim is to trace the perceptions and social climate of the event as a contribution to assaying the place of women in American society. It is clear that colonial rituals of delivery were fully the province of women. A coterie of female neighbors and older relatives attended each birth, which was a social expression of female solidarity, given biological strength by the presence of an experienced birth attendant, the midwife (literally the "with woman"), just as in old England.

One of the greatest dramas of the Metropolitan Opera is the day-to-day struggle of maintaining its standard of excellence.

It's a drama in which you can play a very important role.

We need your support now.

We at Pioneer High Fidelity together with The National Endowment For The Arts will triple the amount of whatever you can give.

Naturally, the more you contribute to the Metropolitan Opera, the more it can contribute to society.

Even a quarter will help.

You see, the way we figure it, you'll not only be supporting one of the world's great cultural institutions, but the very foundation it's built on. Civilization. PIONEER High Fidelity. We all owe a lot to music.

Enclosed is my donation of: $_____ ☐$10 ☐$25 ☐$50 ☐$100
(Checks should be made payable to: The Metropolitan Opera/Pioneer Fund.)

Name _____

Address _____

City _____ State _____ Zip _____

Mail the coupon with your donation to: The Met, c/o U.S. Pioneer Electronics Corp., 85 Oxford Dr., Moonachie, N.J. 07074. SA

PLAY A SUPPORTING ROLE IN THE METROPOLITAN OPERA THIS SEASON.

Birth took place in the home, of course. Today we have not the midwife but the obstetrician and a medical team. Men are the leaders, and the entire procedure is held to be medical and risky to health and life, justifying powerful interventions. Indeed, in articulate reaction to this very change there are now a number of movements for "natural" and "humane" delivery, often at home. The chronicle of change is interestingly and fairly told here, although a somewhat more quantitative approach would have added strength to the social and psychological analysis.

It is clear that childbirth was indeed a risk, although never as serious a risk as the stereotyped image from a New England cemetery: a farmer's tombstone surrounded by those of his several wives, all apparently dead in childbirth. It appears that premedical childbirth must indeed have accounted for a good part of all deaths of women during their childbearing years, roughly one mother dead in 100 live births. The more celebrated midwives whose records we have did somewhat better. Scientific medicine entered the field in gleaming and fearsome mechanical style, coming from France in about 1750. Plenty of intricate clamps and forceps appeared, but no great success. It was Boston's great Oliver Wendell Holmes who warned in 1843 of the danger of childbed fever borne by the physician—just before Ignaz Semmelweis, whose hard data proved the case.

It was really only in about 1935, with teams, antibiotics, transfusions and the rest, that physicians began to cut maternal mortality sharply to its present U.S. level, down fifty- to a hundredfold from the old plateau. We can well afford now to mitigate the emotional costs of all that manipulative and impersonal doctoring. What we cannot afford is the shameful discrimination poverty still enforces. The loss of black mothers remains three times that of white ones even though both rural and urban deliveries now take place in hospitals, attended by strong obstetrical teams. The causes lie largely outside the ward and before the delivery: septic abortion, malnutrition and lack of adequate prenatal care all take their toll.

Prematurity remains the biggest cause of infant early mortality. In municipal hospitals about half of the women get prenatal care either too late or not at all; in private hospitals only 5 percent lack such care from the start. Class and caste impose at birth "a terrible fate for individuals and for the society in which they are born."

LEO SZILARD: HIS VERSION OF THE FACTS, SELECTED RECOLLECTIONS AND CORRESPONDENCE, edited by Spencer R. Weart and Gertrud Weiss Szilard. The MIT Press ($17.50). EINSTEIN PORTFOLIO, photographs by Lotte Jacobi, foreword by Richard M. Bacon. NECCO, Box CC, Hillsboro, N.H. 03244 ($20, postage paid). The century of Albert Einstein's birth is marked this month almost everywhere. Here we look a little obliquely at that luminous figure. It is the prescient and witty mind of Leo Szilard, in "all things counter, original, spare, strange," that is opened to us, from his youth to the end of 1945. Szilard's scientific papers were published some years ago, as was recorded in these columns. This volume contains about 120 items that are not scientific papers. Most of them are letters, a few are long policy documents from critical junctures along his (and our) road into the nuclear age, and there are 56 fascinating pages of carefully edited taped recollections.

Szilard and Einstein are forever linked in history. It is not surprising that the young Szilard, a Ph.D. student with Max von Laue in Berlin in 1921, was working on a problem in relativity that Laue had posed for him. What followed was less likely. "I forced myself to work on it, but it just wouldn't go at all . . . for about six months." Then, in three weeks, an onrush of ideas carried Szilard to his novel view of the second law of thermodynamics, which was entirely unrelated. He was then a member of a student seminar under Einstein. "I told him what I had done. And Einstein said, 'That's impossible. . . .' And I said, 'Well, yes, but I did it.' . . . It didn't take him five or ten minutes to see, and he liked this very much." Laue read it overnight and rang up "early in the morning" to accept it. The paper and its follow-up half a year later were to become cornerstones of modern information theory.

Between 1927 and 1930 seven German patents were issued to the partners Szilard and Einstein (the reader may guess whose was the enterprise) on an ingenious way of pumping liquid metal by means of moving magnetic fields. The German General Electric Company (A.E.G.) wanted to use the scheme for refrigerators, but it turned out in practice to be noisier than conventional pumps with moving parts. Such schemes have since been used in nuclear reactors, the first of them being the Los Alamos mercury-cooled plutonium fast reactor of 1947.

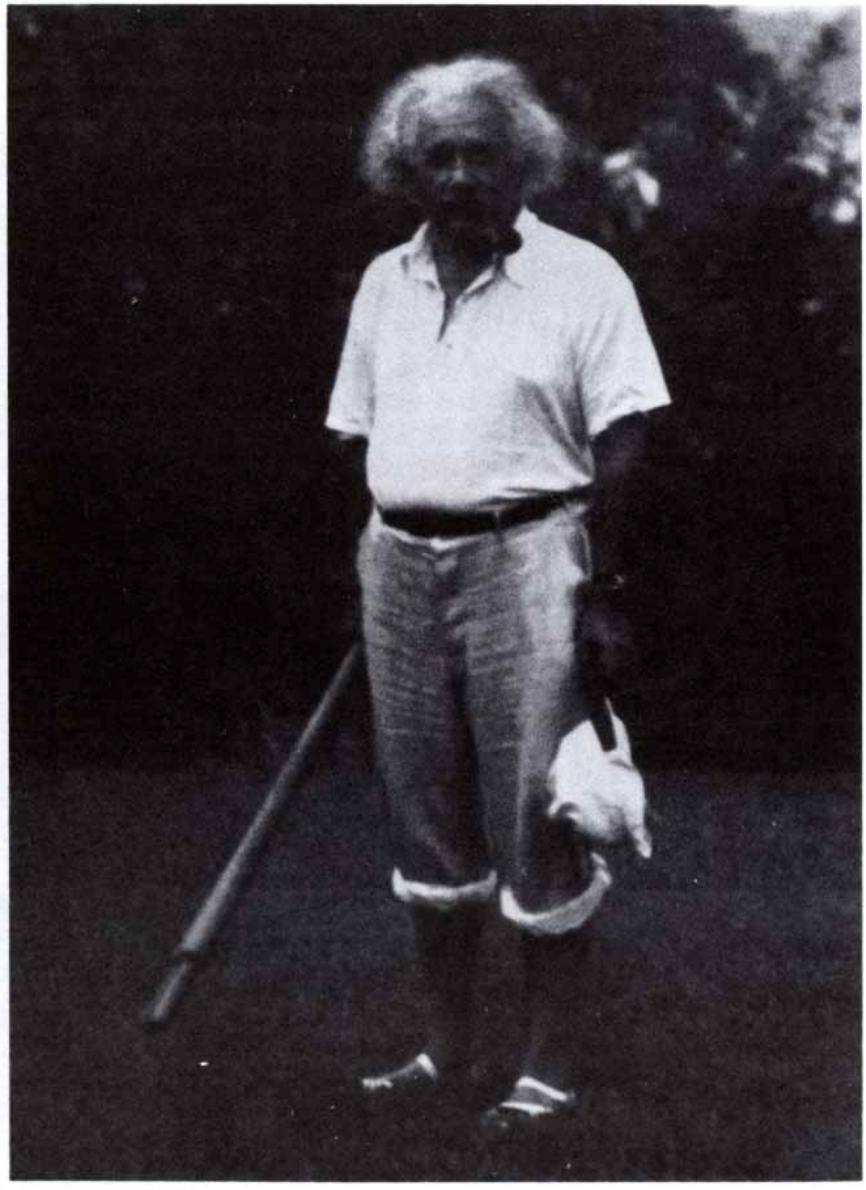
Einstein enters again as the second-act hero of Szilard's tragicomedy of starting the U.S. uranium project. As an émigré in London in 1933 Szilard had been seized by the idea of a neutron chain reaction "as I was waiting for the light to change." The idea became an obsession. In January, 1939, long after Szilard's secret patent on the concept of a neutron chain reaction and weeks after the discovery of fission itself, he wrote that the news of that discovery "revives all the hopes and fears . . . which I have as good as abandoned." Fission "might lead to a large-scale production of energy . . .

SEAGRAM DISTILLERS COMPANY, N.Y.C. 80 PROOF. DISTILLED DRY GIN. DISTILLED FROM GRAIN.

unfortunately also perhaps to atomic bombs." Much experiment, intrigue, counseling and writing of urgent letters ended with the two drafts of July, 1939, of a letter for Einstein's signature in "Dr. Moore's cabin" at Peconic on the Long Island shore. (Nobody knew Moore; everybody knew Einstein was living in his cabin.) The addressee was President Roosevelt; the letter was delivered by hand by the intermediary Alexander Sachs, was acknowledged and was acted on. (A curiously misguided approach Szilard made to get Charles Lindbergh to serve as a spokesman is also documented.)

The war burned on, and in the spring of 1945 Szilard was again seeking to impose order on events: "But now, with the war won, it was not clear what we were working for." Again Einstein acted as a spokesman, this time through Eleanor Roosevelt and this time in an effort to put the genie back into the bottle the earlier letter had unsealed; he spoke of international controls and moral opposition to dropping the bomb. The memo was written in late March, 1945, and on April 12 Franklin Roosevelt lay dead. Events unfolded apace; when the news of Hiroshima arrived, the peaceable Einstein uttered, we know, a cry of despair that is "not conveyed by the translation 'Alas.'"

The volume closes with Szilard the lobbyist active in Washington at the end of 1945. Within a month after Hiroshima more worldly forecasts appeared. Jacob Viner, the University of Chicago economist, saw neither preventive war nor enforceable international agreement, the outcomes most visible to Szilard and the physicists. Instead, Viner said, "Russia also will have the bomb,



IN 1937 EINSTEIN was photographed holding a pair of oars and other gear as he prepared to go sailing on Long Island Sound. The photograph is in a portfolio made by Lotte Jacobi.

LOOK HOW MUCH LUXURY WE CAN FIT IN A MID-SIZE LEMANS.

One look is all you'll need to appreciate what a full measure of luxury this sleek new Grand LeMans delivers.

Luxury you can see in the well-cut contours that make Grand LeMans photogenic from any angle. With a solid look of Body by Fisher.

Luxury you can feel in these available, swank 60/40 cloth seats. And a more expansive choice of interior fabrics than last year's Grand LeMans.

Luxury you can hear in the way this acoustically engineered mid-size surrounds you with a sense of quiet. Or with any one of six available stereo systems to surround you with song. And best of all, luxury you can share as Grand LeMans accommodates six.

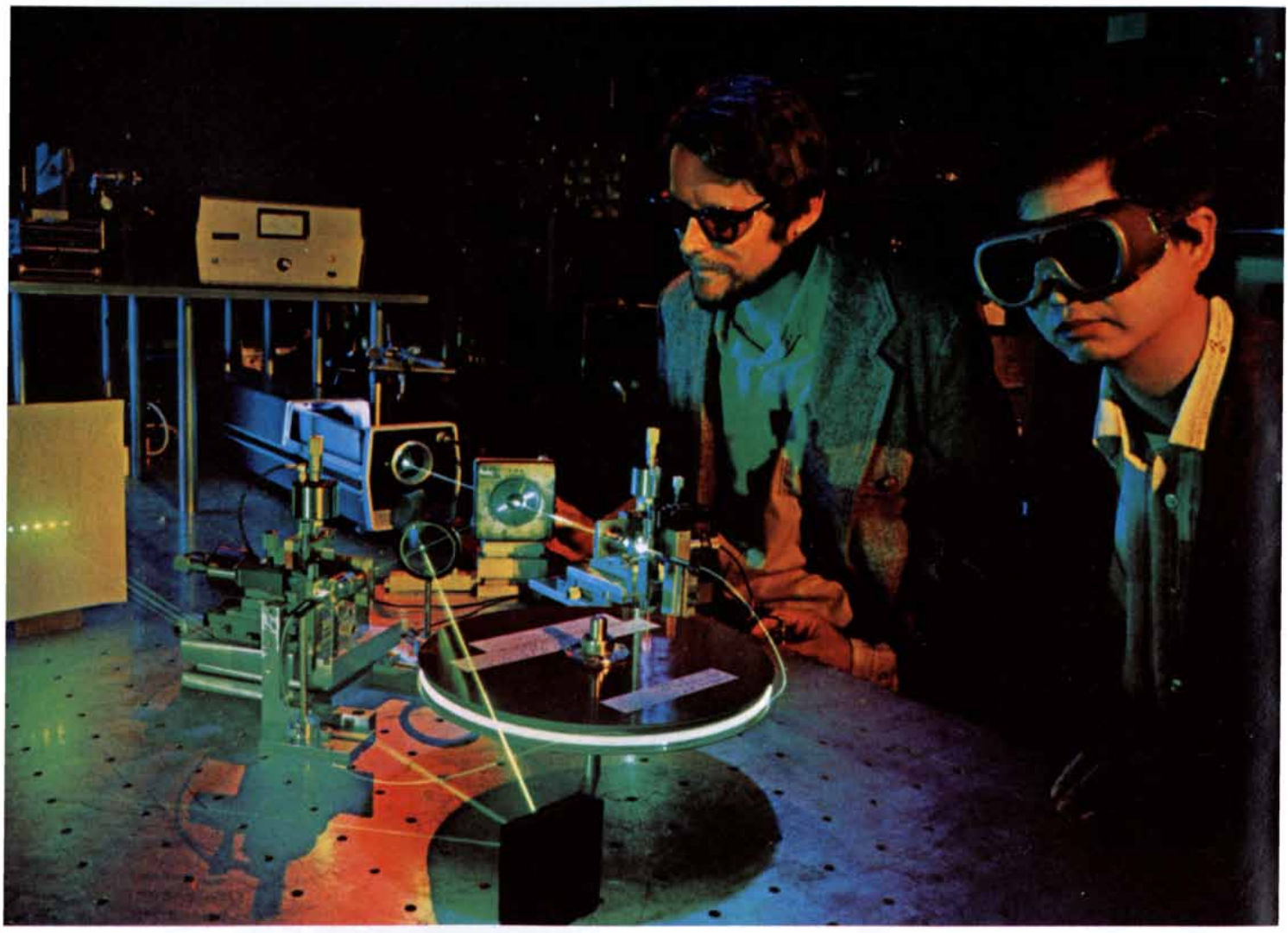
Several items shown are optional at extra cost. But if you're looking for an affordable mid-size loaded with luxury and powered by a new, freer breathing V-6 engine,* buy or lease our Grand LeMans. You'll like the way it fits.

*LeMans is equipped with GM-built engines produced by various divisions. See your Pontiac dealer for details.

GM

THE 1979 PONTIACS OUR BEST GET BETTER

The longest in our long line of laser firsts...



Bell Laboratories
Murray Hill, New Jersey 07974

Bell Labs scientists Roger Stolen and Chinlon Lin work with a fiber Raman laser, one of a new class of light sources that use optical fibers—up to a kilometer long—to produce tunable laser light. At left, the laser's output—which contains multiple Raman-shifted wavelengths—is taken off a beam splitter and dispersed by an external grating to show the broad range of wavelengths that can be tuned.

Bell Labs has developed some of the world's most transparent glass fibers to *carry* light for communications. We've also devised a way to make these highly transparent glass fibers *generate* light. In fact, they are the basis for a new class of tunable light sources called fiber Raman lasers. They're among the latest, and by far the longest, of many lasers invented at Bell Labs, beginning in 1957 with the conception of the laser itself.

Since the new fiber lasers work best at wavelengths at which they are most transparent, we can make them very long. The longest active lasing medium ever built, in fact, was a fiber Raman laser over a kilometer in length. Studying the ways light and glass interact over such distances is part of our research in lightwave communications.

In these new light sources, a glass fiber with high transparency and an extremely thin light-guiding region, or core, is excited by a pump laser. The pump light, interacting with the glass, amplifies light at different wavelengths through a phenomenon known as stimulated Raman scattering. This light is fed back into the fiber by a reflecting mirror. If gain exceeds loss, the repetitively amplified light builds up and "lasing" occurs.

Fiber Raman lasers have conversion efficiencies of about 50%, operate in pulsed and continuous wave modes, and are easily tunable over a broad wavelength range in the visible and near infrared regions of the spectrum.

We've used these lasers to measure the properties of fibers and devices for optical communications; and studies of the lasers themselves have revealed a wealth of information on frequency conversion, optical gain, and other phenomena. Such knowledge could lead to a new class of optoelectronic devices made from fibers, and better fibers for communications.

Looking back

These long lasers come from a long line of Bell Labs firsts:

1957: The basic principles of the laser, conceived by Charles Townes, a Bell Labs consultant, and Bell Labs scientist Arthur Schawlow. (They later received the basic laser patent.)

1960: A laser capable of emitting a continuous beam of coherent light—using helium-neon gas; followed in 1962 by the basic visible light helium-neon laser. (More than 200,000 such lasers are now in use worldwide.) Also, a proposal for a semiconductor laser involving injection across a p-n junction to generate coherent light emitted parallel to the junction.

1961: The continuous wave solid-state laser (neodymium-doped calcium tungstate).

1964: The carbon dioxide laser (highest continuous wave power output system known to date); the neodymium-doped yttrium aluminum garnet laser; the continuously operating argon ion laser; the tunable optical parametric oscillator; and the synchronous mode-locking technique, a basic means for generating short and ultrashort pulses.

1967: The continuous wave helium-cadmium laser (utilizing the Penning ionization effect for high efficiency); such lasers are now used in high-speed graphics, biological and medical applications.

1969: The magnetically tunable spin-flip Raman infrared laser, used in high-resolution spectroscopy, and in pollution detection in both the atmosphere and the stratosphere.

1970: Semiconductor heterostructure lasers capable of continuous operation at room temperature.

1971: The distributed feedback laser, a mirror-free laser structure compatible with integrated optics.

1973: The tunable, continuous wave color-center laser.

1974: Optical pulses less than a trillionth of a second long.

1977: Long-life semiconductor lasers for communications. (Such lasers have performed reliably in the Bell System's lightwave communications installation in Chicago.)

Looking ahead

Today, besides our work with tunable fiber Raman lasers, we're using other lasers to unlock new regions of the spectrum in the near infrared (including tunable light sources for communications), the infrared, and the ultraviolet.

We're also looking to extend the tuning range of the free electron laser into the far infrared region—where no convenient sources of tunable radiation exist.

We're working on integrated optics—combinations of lightwave functions on a single chip.

Lasers are helping us understand ultrafast chemical and biological phenomena, such as the initial events in the process of human vision. By shedding new light on chemical reactions, atmospheric impurities, and microscopic defects in solids, lasers are helping us explore materials and processes useful for tomorrow's communications.

Also under investigation is the use of intense laser irradiation in the fabrication of semiconductor devices. The laser light can be used to heat selective areas of the semiconductor and anneal out defects or produce epitaxial crystalline growth. Laser annealing coupled with ion implantation may provide a unique tool for semiconductor processing.

We've played an important part in the discovery and development of the laser—an invention making dramatic improvements in the way our nation lives, works and communicates.

Bell Laboratories

From Science: Service

and then a new equilibrium will establish itself." Szilard adds in the 1960's: "He had certainly more foresight than the rest of us, though it is not clear whether what we have now is an equilibrium or whether it is something else." This volume is one authentic and indispensable document about our leaders and our times, and likely about our fate still to come.

Lotte Jacobi and her sensitive, intimate and loving eye and camera looked at Einstein the man for many years. If you would see him in shirtsleeves and with trousers rolled up, or joyously windblown in a little sailboat on Long Island Sound (although not where the fateful letter was drafted), you have only to pick photographs out of the set of prints she now offers. They show the man in his home, content or burdened by the mail, smiling or grave with guests, in the fearful time between his 50th and 60th years. They are enduring evocations, perhaps more fitting than a colossal bronze for this quiet, good, beset and penetrating man.

LIGHT WATER: HOW THE NUCLEAR DREAM DISSOLVED, by Irvin C. Bupp and Jean-Claude Derian. Basic Books, Inc., Publishers ($10). The Boeing 707, the first American jet transport and an enduring worldwide commercial success, arose out of a single prototype that became both the 707 and the Air Force tanker KC-135. Tankers accounted for most of the early production and still account for more than half of the total production run. The successful military investment led to a resounding civil achievement.

Sharing this kind of military parentage are the American-designed reactors moderated by light water (as distinct from heavy water). Such reactors dominate the Western world of nuclear power; counting both completed power stations and those under construction in late 1977, 90 percent of all power reactors were of this type. Their military legacies are the great gaseous-diffusion plants that provide their enriched uranium and the original development of the reactor type itself for naval service. But reactors at sea, most of which are submerged, are an order of magnitude lower in power than the usual land-based power reactor, and the economic questions are entirely distinct. "To the scientific community, light-water reactors were the accidental by-product of early military experimentation directed at submarine propulsion." The reactor type that works well in propelling submarines is not the clearly rational choice among alternative designs for civil use. In the real world technical rationality has taken second place to "military, political and commercial constraints."

This candid and well-informed book chronicles the rise of light-water reactors worldwide to dominance of the

SOMETIMES A PRODUCT JUST KEEPS GETTING BETTER AND BETTER.

The Kodak Carousel projector is a good example of such a product. Tried, proved, and improved over time.

It uses gravity to gently drop each slide in place. It's whisper-quiet, too.

It features a curved-field projection lens. For sharp edge-to-edge projection.

It uses the Transvue slide tray that lets you see the number of the slide shown, even in the dark.

It comes in many models. So you can choose just those features you want. Such as remote control. Automatic focus. Even automatic slide changing at timed intervals. And there is a wide variety of accessories made for it. Everything from special-purpose lenses, trays, stack loader to special "presentation aids."

See the Kodak Carousel projectors at your photo dealer's. And enjoy the projector that keeps getting better.

KODAK CAROUSEL PROJECTORS Kodak

market. It assesses the nuclear debate in Western Europe and the U.S. not by the usual survey of apparent issues of technical substance—safety, waste disposal and the proliferation of nuclear weapons—but by examining the string of decisions in their contexts. The Franco-American collaboration of the authors brings together two men who have had a long professional association with their national nuclear power programs, one of them technically trained and the other one closer to political economy. They write, more in sorrow than in anger, about the "abuse of a technology" by its advocates, so that now, although "the world needs nuclear power," the option to exploit it is in acute danger.

The narrative is in two parts: flood and ebb. The tide crested during the early 1960's. When Jersey Central Power & Light bought a light-water turnkey plant from General Electric at the end of 1963 without Federal subsidy, economic analysis promised that it could deliver cheaper power than any other generating system. Westinghouse soon matched the price pretty well, and the "Great Bandwagon Market" for light-water reactors began. Nearly 50 plants were ordered in 1966 and 1967 by U.S. utilities; four manufacturers shared the business. By the early 1970's there was worldwide intoxication with light water. Britain and Canada apart, American designs had swept the market.

The French decisions are outlined in detail. The nine turnkey projects (with firm prices guaranteed by the manufacturers) had the effect of ending the efforts of designers in Europe and Japan. Everyone wanted to be the licensee of an American company. Finally "some 18 different European companies or consortia" were offering light-water reactors for sale. The low cost proved to be an illusion—an illusion of promise rather than performance, of partial achievement rather than of complete and tested systems all the way from mine to waste disposal, and of skimped attention to safety (or at least to instilling a conviction of safety in the public mind).

The Organization of Petroleum Exporting Countries (OPEC) played a strong supporting role. Fuel costs rose precipitately after 1973. Coal and uranium were competitive in 1970; in 1976, with coal at four times its 1970 price, coal-generated power was still competitive with nuclear power from light-water reactors. In constant dollars the nuclear plants that began operating in 1975 were on the average three times as expensive as the turnkey plants had been. Those early plants had in fact been loss leaders. The "learning curve" had not brought the price down; the economies of scale were not easily realized. "By 1975 the steel and concrete needed per megawatt for a 1,200-MW plant approximately equaled the 1960 requirement for a 200- to 300-MW design." Clearer safety meant thicker walls.

Today the issue is deep in symbol, emotion and myth. Not only is there a general shift away from a consensus in favor of economic growth but also the circumstances of light-water commercialization have made the technology peculiarly vulnerable. Scientists, engineers, politicians, bureaucrats and businessmen in a dozen countries were caught up in a self-fulfilling interlock of hopes. Nuclear power had been a "theology," light water its "sanctified technology." The U.S. Atomic Energy Commission and the Congressional Joint Committee on Atomic Energy, institutionally allied with manufacturers and their salesmen, had done much to confuse hope with reality.

Is there a chance for fission power? The authors think a compromise all around might yet work. They are clearly disturbed by a public climate that appears to require "full accommodation of all minority objections to any positive action," and they believe this "reactionary prospect" might be mitigated by concessions, thereby retaining a significant long-term role for fission.

The first issue for compromise might be waste disposal: it is an urgent problem, and yet not so acute as to demand instant action. Perhaps here a consensus might be found that is workable, unhurried and impartially judged to be safe enough. The second issue is the economic one. Most people want cheap power, but no one can yet fairly assess relative costs because there simply is not enough experience. A temporary suspension of judgment that would allow a test of both nuclear and coal-fired power might be workable.

The authors seem to be on firmer ground when they suggest there are real benefits in "letting sleeping breeders lie." Many people who balk at a "vision of hundreds of breeder reactors and a plutonium-based electricity system" might support the renewed growth of light-water capacity. In Europe the issue is even more symbolic, the authors believe. Where are the alternatives to the full-blown scenario of nuclear power as Europe's only choice? They suggest that critics there are concerned not so much with the effects of light-water reactors as with the proper structure of modern society and the place of high technology within it.

It was no 20,000 tons of coal that fell on Hiroshima. In all this strange story there is something deeper still, little faced even by these analysts. The fear of an unknowing acceptance of the impure and the dangerous, and the suppressed anxieties of nuclear war, play some part in this curious blend of overconfidence and suspicion. High technology holds many other dreams. Consider a future with a thousand big rigs afloat in warm tropical seas, turbines working against

cold water from the depths, their power used to make hydrogen, ammonia or even methanol. Such a scheme would share the very same ambience of giant corporate construction, but it lies far from the real and the perceived risks of uranium. Would it come closer to acceptance? This is a useful, serious and critical account, without evident hyperbole or animus. The full tale is yet to be told.

BIOGEOGRAPHY AND ADAPTATION: PATTERNS OF MARINE LIFE, by Geerat J. Vermeij. Harvard University Press ($25). There is a sense of history about any acts, even trivial ones, that extinguish barriers long established. More than one reader will have mixed two samples of seawater, taken perhaps from the Pacific and Atlantic coasts, with a feeling of innovation. One of Dr. Vermeij's experiments, carried out on Guam, consisted in offering to local species of crabs some living Guamanian snails and some related Jamaican snail shells that moved because they were animated by local hermit crabs. (That is a lot easier than fetching the snails alive from the West Indies, and other tests had ensured the validity of the method.)

In every case the West Indian species were vulnerable to crushing by the crabs "at much larger sizes (both absolute and relative) than were their . . . counterparts in Guam." The effect was clear, a factor of two or three in weight for most cases. Indeed, it seems generally true that the armor of the Indo-Pacific shells of many kinds is better than that of their West Indian cousins, and so is the weaponry of the predators. (A sample pun out of an abundant supply may warn the reader: "Tropical predators are more fully endowed with specialized shell-destroying equipment than most of their counterparts and may in general characterize themselves as, 'We shell overcome.'")

An enthusiastic and imaginative University of Maryland zoologist has produced this readable book in the style of Darwin himself. It is full of argument based on simple experiments and expert comparative studies, innocent of fancy apparatus, biochemistry or even mathematics. Photographs of purposeful series of shells and tables of species found at different sites offer most of the technical support the author needs.

The gradients of differences among related forms are perhaps stranger from east to west than they are from north to south. Here sounds one grand theme of the book. North-south differences arise from temperature and from concomitant limits on living function. The little beach clams of southern India burrow into the warm sand within two seconds, whereas their more sluggish cousins in chilly Scotland take three times longer. Yet at both latitudes dislodging waves roll in every few seconds. The British clam therefore lives only below the low-water line, whereas its tropical cognates are found intertidally.

Quite generally the high latitudes channel living forms sharply by simple physical circumstance. Cold-blooded forms cannot be very mobile in the cooler climates; the chemistry of calcification there makes thick and baroque shells more costly; diversity of form increases in the Tropics. With diversity, of course, comes coevolution. The arms race of armor against lance becomes decisive for form. Environmental conditions are the less limiting ones; the more pervasive conditions are the forces of predation, at least among the shelly beasts that are the motif of this work.

Periwinkles, snaily bivalves and the plants and animals (such as barnacles and algae) that live on other organisms each get a detailed chapter. In all these groups there is a general tendency for the development of spines, ridges, thickness of shell and a general architecture resistant to predators to be greater at the lower latitudes. The predators follow suit. A crayfishlike animal from the coast of Thailand has one pair of arms fitted with ends like hammers for pounding the shell of the prey until it breaks, a breathless drama of the arms race. On the other hand, it appears that soft bottom dwellers, which can hide anywhere and live in a low-oxygen environment where motion is costlier, show less concern for armor. The argument is supported by counts of related species, by surveys of shells healed after partial failure and by sequences of species.

One striking graph plots the percent (among tropical intertidal gastropods in five regions) of species that have an outer shell with a toothed lip or that have an inflexible operculum. The trend is clear: the Indo-Pacific species has the most armor, the eastern Atlantic one the least. Why? The final answer to this puzzle is not in hand, but a tentative explanation can be given. Advanced coevolution in the Indo-Pacific has yielded the greatest diversity, the strongest symbioses and the highest level of prey-predator warfare. It is tempting to conclude (one cannot yet be sure) that the ancient Pacific world has been less disturbed since Tertiary times and that a long mutual adaptation of the eater and the eaten has elaborated the game. The Atlantic theater, which was surely disturbed severely in recent glaciated epochs, has not provided enough time for the subtle warfare of the genetic plan.

It is obvious that Dr. Vermeij is well prepared to examine the possible effect of the opening of barriers in the sea by man, at Suez and potentially with a sea-level, saltwater Panama Canal. He foresees a considerable effect at Panama, and he explains why. But why some species, such as the horseshoe crab, perceive their environment as being invariant over hundreds of millions of years, whereas other species respond to change

U.S. NAMES RABBIT DIESEL #1 IN ECONOMY.

We always knew we made terrific economy cars.

Now it's official.

Because according to the 1979 EPA Guide, a Rabbit Diesel gets the highest mileage in America: an incredible 40 estimated MPG, and a whopping 50 estimated highway MPG.

(Use the "estimated MPG" for comparison to other cars. Your mileage may vary with speed, weather, and trip length. And your highway mileage will probably be less.)

We could go on and on about all the innovations we've built into our newest Rabbit.

Like the fact that it goes on and on (amazingly!) without ever needing a conventional tune-up. (There are simply no spark plugs, points, condensers, or carburetors to tune.)

Our wizardry doesn't stop with the engine either.

There's front-wheel drive for better tracking.

There's more room for people than 31 other cars you could buy.

And a Rabbit Diesel goes like a bat out of you-know-where. In fact, it set 31 world's records for diesels on a track in Miramas, France.

Come test it for yourself.

For years we've built cars that use very little gas.

Now we've built one that doesn't use any at all.

VOLKSWAGEN DOES IT AGAIN

at a rate that is orders of magnitude higher, is far from clear. This Darwinian book—there is hardly higher praise—is a step toward understanding, which will probably challenge our technical methods more than the broad insights of that clever naturalist on the *Beagle* did.

THE COSMOGRAPHICAL GLASS: RENAISSANCE DIAGRAMS OF THE UNIVERSE, by S. K. Heninger, Jr. The Huntington Library ($17.50). In 1559 the young William Cuningham of Norwich, Doctor in Physicke, brought out in London his handsome folio under this same main title, "conteinyng the pleasant Principles of Cosmographie, Geographie, Hydrographie, or Navigation." Elizabeth the Queen was the patron for the book. In the modern volume at hand, also impressive although not outsized or lavish, a scholar at the University of British Columbia has reproduced Cuningham's elaborate title page as his own, with its many figures such as Astronomia, Strabo and Ptolomeus and with appropriate new letterpress material inserted into the old woodcut. The volume it introduces is a delight, a "portfolio of diagrams which in one way or another depict the world as someone in the Renaissance viewed it." Every figure is taken from a book printed before 1700, a few from early books of the 15th century. Each figure is analyzed and set in context by the text, which is, however, not a series of captions but rather a connected and readable exposition illustrated in detail; the book is not a history or a thesis but rather a display of the images themselves with explanation.

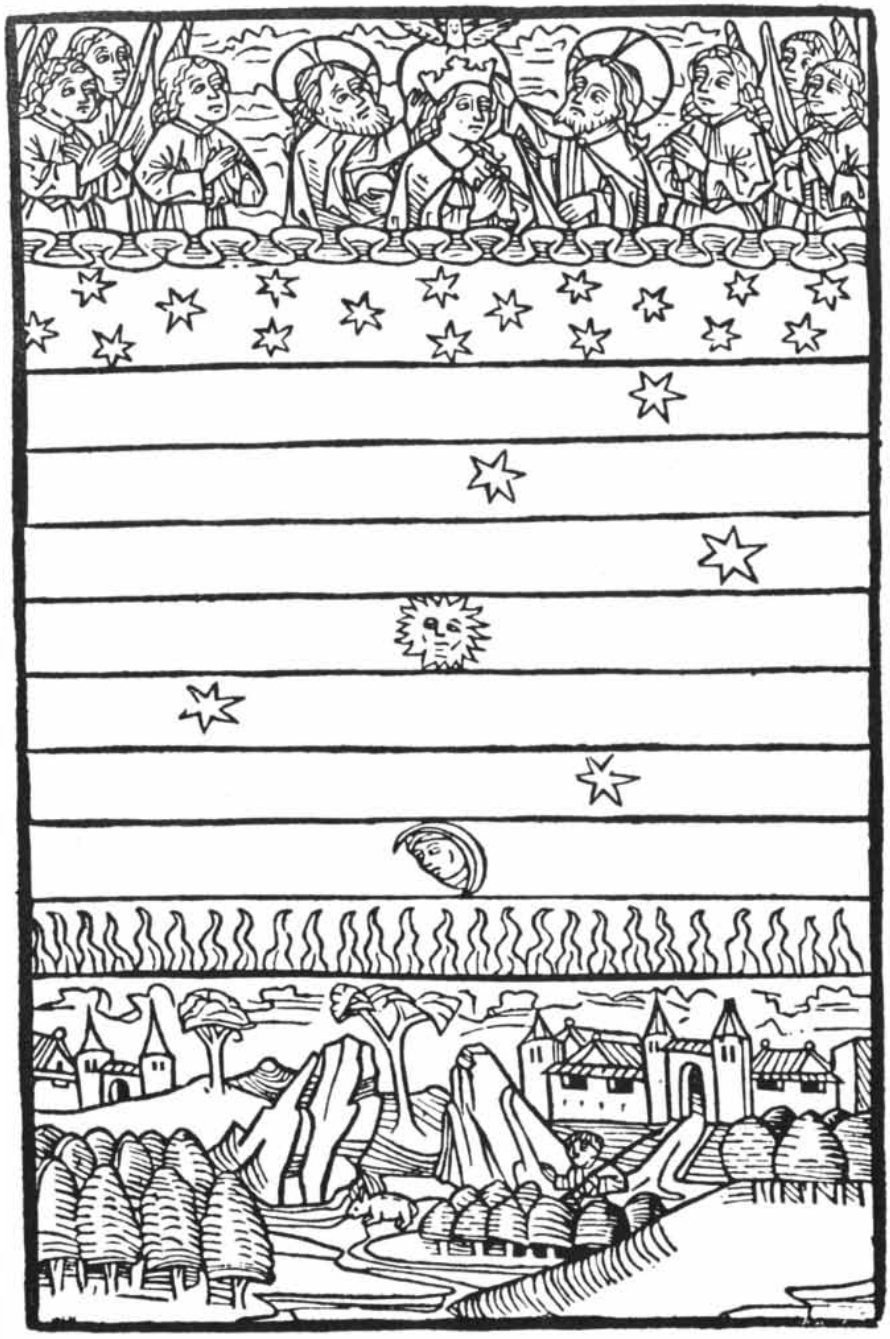
There is a clear synoptic structure: the first three chapters (after the introduction) exhibit the figures of authors whose intent was to describe our universe in a way consonant with what we perceive through the senses; the second half of the book presents those who sought rather to uncover some underlying theoretical articulation of the whole. The first half shows great attention to concentric circles, enlivened by the creatures of the universe, whereas the second half emphasizes less evident symmetries with much play of matrix tables, densely labeled polygons and the human figure mapped into the cosmos.

Professor Heninger and his librarian friends have given us a work delighting the eye, a store of images for our visual time that sends a strong and refreshing wind across many a dustier treatise. There are some 150 illustrations, all well reproduced but without reference to the original scale.

Here are a few of them, described to give a taste of the feast. A woodcut from a 1499 edition of Konrad von Megenberg's *Buch der natur,* a popular encyclopedia first published in 1350, shows the universe in layers with our own place of earth, air and water at the bottom. Above us is the layer of fire, "which be-

ing a pure element is not visible," above that in order are the seven planets and the fixed stars, all apparent to our eye, and past the stars is the Triune God flanked by angels, perceptible only to the mind. In this figure both theory and observation are served.

From Sir Edward Sherburne's critical translation of *The Sphere* of Manilius, published in 1675 ("a landmark in English neoclassicism"), we see, elegantly engraved by the famed Wenceslaus Hollar, six pages, each bearing a beautiful and easily read diagram of a cosmic system, from Ptolemy to Tycho and Copernicus. It would be hard to omit the diagram from a 1512 edition of a treatise by the Catalan mystic Ramón Lull. We see the high staircase rising to a noble building of wisdom, the steps marked as a scale of beings from mere stones to flames, plants, animals, man, angels and God. The famous island observatory of Tycho, all circles and squares, is shown as he represented it in 1598; in its context it is as much a symbol as an ideal research station. It goes without saying that the book is well served by a scholarly apparatus (except for that little matter of scale).



THE UNIVERSE is seen in the medieval encyclopedia *Buch der natur* (1349–50) as a hierarchy of the elements, the planetary spheres and the Christian heaven. The illustration is reproduced in *The Cosmological Glass: Renaissance Diagrams of the Universe,* by S. K. Heninger, Jr.

Established 1845 SCIENTIFIC AMERICAN March 1979 Volume 240 Number 3

The Professionalization of the U.S. Labor Force

The increased number of college graduates in the U.S. has altered the composition of its work force. Professionals and managers now hold one out of four of the nation's jobs

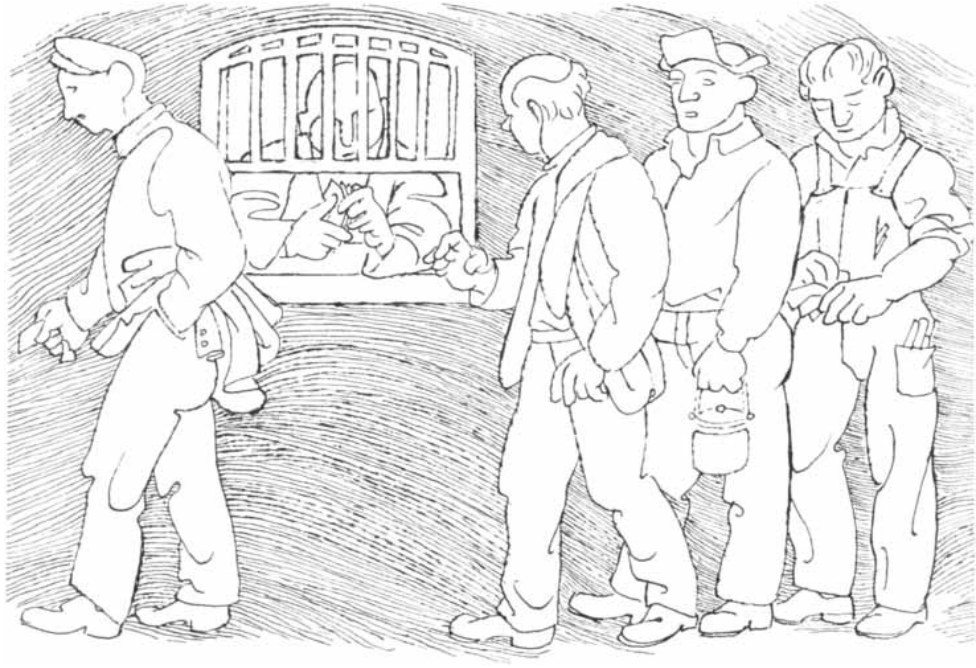
by Eli Ginzberg

Since the end of World War II a steadily increasing percentage of young Americans have been graduating from college. By 1976 one out of four members of the age cohort of 25 to 29 years old had earned the bachelor's degree, compared with one out of 20 before 1940. Now, some 30 years later, may be a good time to assess the cumulative impact of this unprecedented commitment of human and material resources to higher education and career advancement.

The wielders of power in society have often provided routes of upward mobility whereby young people of the requisite talent and ambition could make their way from the lower classes to the top, there to cement or augment the power of their forerunners. Such disparate cultures as the China of Confucius and the Britain of Victoria used education as the ladder. A farmer's son could try the examinations set by Peking; a shopkeeper's son could scale the walls of Oxford. Compared with those precedents the American establishment (a constantly shifting group) made education an open gate.

In the U.S. labor force people listed in two census categories now command one out of four jobs, and these include substantially all the preferred jobs, as jobs are reckoned by income, status and career prospects. The categories "Professional, technical and kindred" and "Managers and administrators" also provide the most comprehensive count of the population of college graduates, granted that some college graduates are also found in lower-ranking jobs among sales, clerical, blue-collar and even service workers. An economy once dominated by entrepreneurs and self-employed professionals ("their own bosses") is now dependent on managers and professionals mainly employed on salary by large organizations, public and private.

A variety of developments reflect the influx of educated and highly educated personnel into the U.S. labor force. Consider the shrinking percentage of blue-collar workers in the labor force, with the near-doubling (inflation discounted) of the gross national product since World War II; the expansion of research and development in governmental and business budgets and the flowering of major new industries built on the technologies of solid-state electronics and information processing; the transformation of family-owned big businesses into multinational conglomerates under multilayered middle-management administration; the substantial growth of the not-for-profit private and governmental sectors of the pluralistic U.S. economy; the proliferation of the control functions of the Federal Gov-



CHARACTER OF U.S. LABOR FORCE in the late 1940's is suggested in this drawing by Bernarda Bryson. Over the past 30 years the percentage of blue-collar workers in the labor

ernment into many aspects of the economy and the environment, and the more active initiative of state and local governments in the provision of services to the public, particularly in health and education.

By no means, of course, is there room at the top, in the councils of the establishment, for all the able and qualified people American society has been bringing up through its educational system. It is probable that most managers and professionals never had the illusion of such advancement or that they put it aside early in their careers. On the other hand, government officials, professors, journalists, health administrators, systems analysts and the like do not see themselves as minions of their employers. With many of them educated and trained for six or seven years beyond high school, so that they have had time to learn to think critically, they do not automatically accept the values, goals and patterns of behavior of the society around them or of the organizations they join. It was the economist Joseph Schumpeter's perceptive insight that the enlistment of such people in the management of the capitalist system might ultimately undermine it. Clearly their judgment and authority narrows the discretion of top management in the public agencies and business organizations in which they are employed. To the political life of the country they have brought new kinds of issues and new styles of promoting and advancing them.

In the 1950's, following the rapid increase in college enrollment induced by the stipends of the "G.I. Bill of Rights" to veterans of World War II, the U.S. higher education system entered on a period of sustained expansion. State legislatures encouraged state universities, on the model of the California system, to proliferate into chains of campuses. A booming economy brought private universities into the market with $100-million fund-raising campaigns. The Federal Government, from hugely expanded outlays bearing the powerful sanctions of "defense" and "health," had increasingly large sums available to support scientific enterprise in the universities; that flow of funds climbed to $1.5 billion by 1967.

This was a far cry from the way of life of university science in the 1920's; in those days, the physicist I. I. Rabi recalls, he could get his work started at Columbia University only because another prospective Nobel prizewinner, Harold C. Urey, generously shared his research funds with him. Throughout the nearly three decades of its expansion, as the late Allan M. Cartter of New York University observed in the mid-1960's, the higher education system was itself creating a market for its own graduates, particularly those of doctoral rank. The teaching staffs of the universities and colleges more than doubled, increasing from 247,000 in 1950 to 562,000 in 1977.

The number of college graduates rose from 286,000 in 1955 to more than 900,000 in 1977, and the vast majority of these went to swell the ranks of managers and professionals throughout the economy. In 1940 college graduates were comparatively rare in the labor force, in the ratio of 1:22. Once one added up those engaged in the classic professions (physicians and other medical practitioners, lawyers and the clergy), teachers and engineers there remained only about 620,000 college-trained professionals to be accounted for. Today, with college graduates making up nearly a fourth of the total labor force, the corresponding number of additional professionals to be accounted for is six million.

The Current Population Survey for 1977 shows the category "Professional, technical and kindred" at 13.7 million. This compares with seven million, shown in a comparable data source, for 1958—a 97 percent increase in a single generation. From the 13.7-million total one can subtract 1.078 million technicians in engineering, science and other fields, who by definition are not professionals. In the adjusted total of 12.6 million the "Learned professions" of law (462,000), medicine (724,000) and the clergy (347,000) add up to 12 percent. Among those closely allied to the learned professions can be counted the 562,000 members of the professoriate and the much larger subcategory of "Writers, artists and entertainers," numbering 1.141 million.

"Teachers, except college and university," account for 3.042 million, to which can be added 208,000 librarians and 175,000 vocational and educational counselors for a combined subtotal of 3.425 million in the teaching professions. What can be called the "helping" professions, more closely allied to medicine, bring together 1.285 million nurses, dietitians and therapists, 462,000 health technologists and 444,000 social and recreational workers in a subtotal of 2.191 million. The subtotals of the clas-

force has shrunk, and the professional and managerial segment has evolved from a small group of entrepreneurs and self-employed professionals to sizable group of salaried workers, mostly employed by large public and private organizations (*see illustration on next page*).

sic and allied professions and the teaching and helping professions yield a grand total of 8.852 million people who can be grouped together as professionals rendering their services to people.

The rest of the 3.730 million professionals can be reckoned, by contrast, as rendering services to organizations—business, not-for-profit and governmental. Among the occupational groupings under this heading the 1.267 million engineers and 868,000 accountants are the largest and most familiar. Here appear also the new professions, outgrowths of postwar developments in technology, the computer specialists numbering 371,000 and the operations and systems analysts numbering 122,000. Among the professionals who serve big organizations, finally, should be counted a large percentage of the lawyers and perhaps as many as 700,000 "Writers, artists and entertainers," as well as specialists in personnel and labor relations, physical, biological and social scientists, architects and a variety of research workers. The adjusted total of 4.884 million in this large-organization support group comprises 40 percent of all professionals.

Between the category of professionals and the other major category of managers and administrators there is a constant flow of people, principally in the managerial direction. Thus a lawyer, an accountant or an engineer is elected a corporate officer; a bench chemist becomes a plant manager or a physician takes on the administration of a medical center. The managerial category, shown at 9.7 million in the Current Population Survey for 1977, has increased about 42 percent above the 1958 figure of 6.8 million, or at about half the growth rate of the professionals. The gross figure conceals, however, something of a social revolution in the ranks of U.S. managers and administrators. In this 20-year period the number of self-employed managers (principally small businessmen) declined by 51 percent to a total of 1.745 million, while the number of salaried managers increased by four million, or more than 110 percent, to 7.548 million. (Farm managers are not included in this breakdown.) The slope of the trend to professional management is plainly steep.

About 1.3 million of the managers and administrators are employed in the public sector and the private not-for-profit sector, carrying responsibilities such as the administration of bureaus of the Federal Government, universities, offices of philanthropic foundations or hospitals. Some two million managers are engaged in goods production (in manufacturing, construction, mining, transportation and public utilities), where the principal value is added to the material product of the economy. Finance, insurance and real estate employ 900,000. Selling at retail and wholesale accounts for four million, in a ratio of 2:5 to the total managerial category. Services (and miscellaneous industry sectors) account for the rest.

Considering the entire professional and managerial categories together, in their combined total of 22.3 million [*see top illustration on page 53*], it can be seen that most professionals and managers are employed by business and governmental organizations. Fewer than two million managers (mostly small businessmen) and probably fewer than one million professionals are self-employed. The substantial role of the governmental and the private not-for-profit sectors in the country's total economic activity is reflected in the fact that two out of five managers and professionals are employed in these sectors. What is more, the real source of the dollars that support the work of the professionals is not always their direct employer. The Federal Government supplied some $10 billion of the total $27.8 billion expended by industry for research and development in 1977 and may therefore be regarded as the ultimate employer of a third of the research scientists and engineers on the industrial payroll.

Given the increase in the number and

PROFESSIONALS AND MANAGERS, or those people classified in the U.S. census categories "Professional, technical and kindred" and "Managers and administrators," hold an important new place in the nation's labor force, as is suggested in this illustration. Numbering about 22.3 million in 1977, professionals and managers have come to command the large majority of what are generally considered to be the most desirable of the nation's jobs: those jobs that offer the greatest rewards in terms of income, status and career advancement.

importance of the jobs to which education provides the major entry, it is appropriate to ask whether women and members of racial minorities are now to be found in greater numbers in the top fourth of the labor force. After all, education is supposed to be the great equalizer of opportunity, and education is largely furnished at public expense.

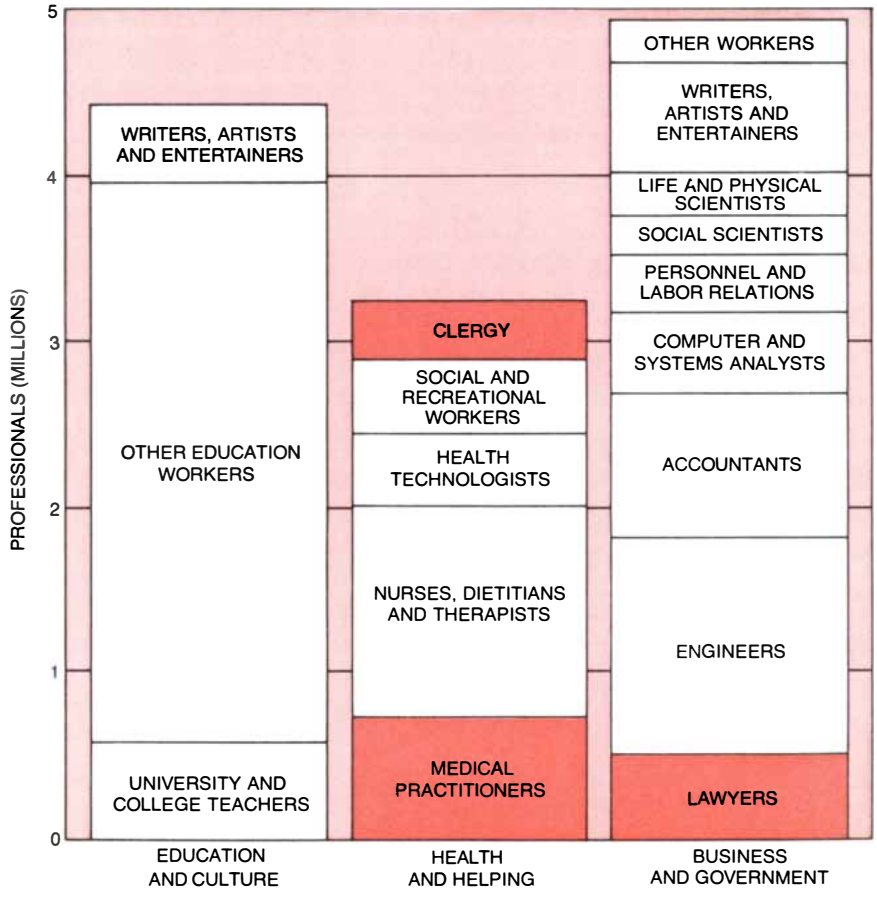
For women it can be said that there has been some improvement, with notable percentage gains in the most prestigious occupations where they had earlier been grossly underrepresented. To begin with, women are overrepresented in the professions compared with their numbers in the labor force as a whole. The reason is that certain professions, such as schoolteaching and librarianship, nursing and social work, have been reserved for them as "women's work." Hence as long ago as 1950, when women made up 29 percent of the labor force, they held 40 percent of the 4.5 million professional jobs. In 1977, in spite of the huge expansion of the professional labor force, the female component increased to 43 percent, and the number of women professionals had grown from 1.8 million to 5.8 million. This large increase has until recently made scarcely a ripple in the learned professions. The ratio of men to women in engineering still exceeds 50 to one, in the law 20 to one, in medicine 10 to one and in the sciences six to one.

That some considerable improvement may be realized in the future is suggested by the fact that enrollment in colleges is now for the first time 50 percent female. Young women are majoring in the sciences and engineering and going on to graduate study in those fields in much larger numbers than ever before. In medical and law schools the percentage of women enrolled is now running well ahead of their current representation in those professions.

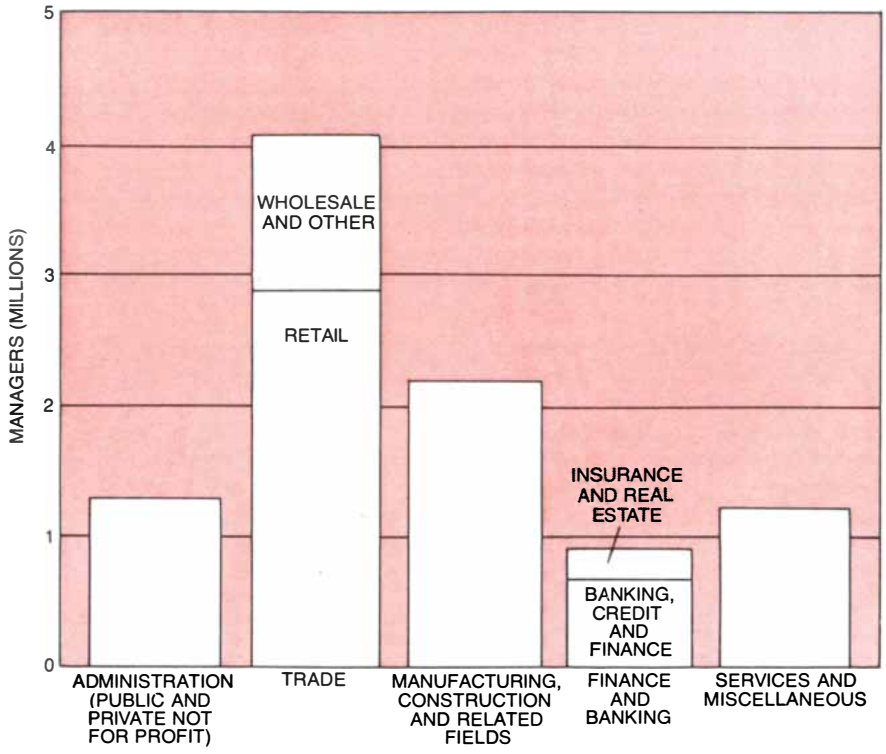
Progress for the U.S. black minority is reflected in the increase in the percentage of employed blacks who are in the professions. That indicator moved from 4.8 to 11.7 percent between 1960 and 1976, compared with the increase for whites from 12.1 to 15.7 percent. Again, as in the case of women, blacks are grossly underrepresented in the learned professions. In the universe of engineers, scientists and allied personnel numbering approximately two million there are only 32,000 blacks, and they are outnumbered by 36,000 Asians. Among the fewer than 250,000 scientists and engineers of doctoral rank only 1,700 are black, and they are outnumbered by 10,000 Asians. Against the cumulative disadvantages of poor preparation in elementary and secondary schools, attendance at colleges with weak science departments and a historic lack of job and career opportunities, the members of the black minority will find their way into the scientific and engi-

OCCUPATION GROUP	PERCENT OF LABOR FORCE	
	1958	1977
PROFESSIONAL, TECHNICAL AND KINDRED	11.0	15.1
MANAGERS AND ADMINISTRATORS	10.8	10.7
CLERICAL	14 5	17.8
SALES	6.3	6.3
SERVICE	11.9	13.7
BLUE-COLLAR	37.0	33.4
FARM	8.5	3.0
TOTAL	100.0	100.0

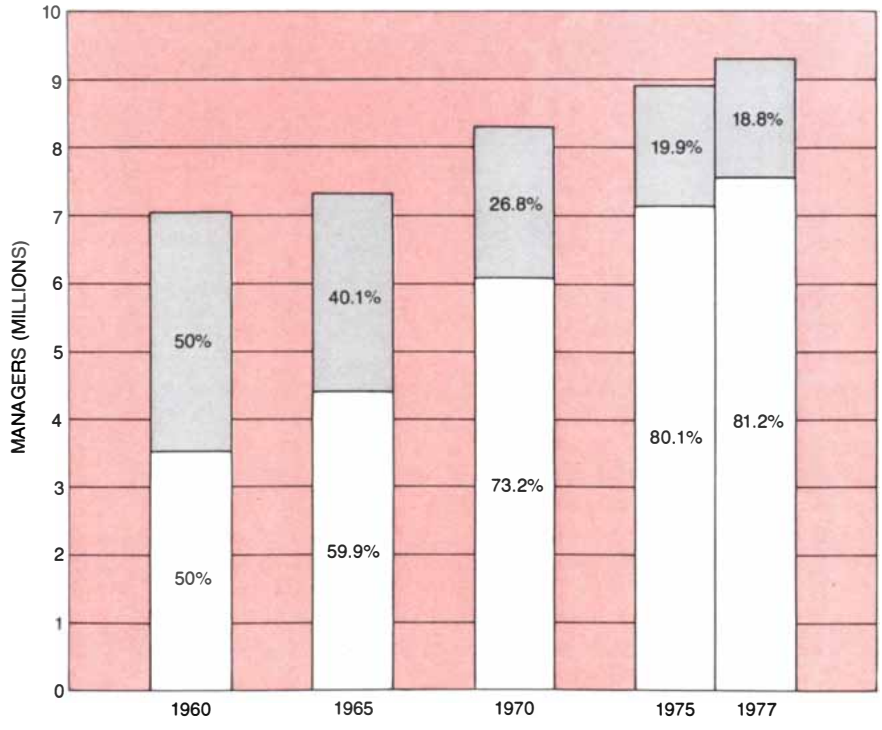
CHANGING COMPOSITION OF THE LABOR FORCE is due largely to increased opportunities for higher education provided since the end of World War II. From 1958 through 1977 the total number of people in the employed segment of the labor force grew by 44 percent, from 63 million to 90.5 million. By 1977 there were 12.6 million professionals (and 1.078 million technicians) and 9.7 million managers and administrators, increases of respectively 97 and 42 percent over the figures for 1958. In this 20-year period proportion of professionals in the labor force increased by about 4 percent and proportion of managers actually dropped slightly, while proportions of blue-collar workers and farm workers each dropped by more than 3 percent.



ONE OUT OF EIGHT of the 12.6 million people classified as professionals in 1977 were employed in the classic professions of law, medicine and the clergy (*dark colored blocks*). There are some areas of overlap in the broad categories into which the various professional jobs have been organized in this chart. For example, medical practitioners (including dentists and veterinarians) are generally self-employed professionals, who could also be classified in the business and government category. The 1.441 million professionals who are writers, artists and entertainers have been divided into two groups in order to differentiate between independent workers and workers employed by organizations. In addition a number of the social, biological and physical scientists have been included in the classification of university and college teachers.



MANAGERS AND ADMINISTRATORS, who numbered 9.7 million in 1977, were distributed among five major areas of the U.S. economy as is shown in this chart. Some 1.3 million, or 13 percent, were employed as administrators and officials in the public and private not-for-profit sectors, working in fields such as school and health administration, union organization, postal services and so on. Most of the remaining 8.3 million managers and administrators work in profit-making enterprises, and about two million of them are self-employed. Farmers, many of whom are classified as self-employed managers, have not been included in this breakdown.



NUMBER OF SALARIED MANAGERS in the U.S. labor force (*white*) grew by four million, or 110 percent, from 1960 through 1977, while the number of self-employed managers (*gray*) dropped by about 1.8 million, or 51 percent. The increase in the percentage of salaried workers in the managerial segment of the labor force over this period reflects the continual movement of professionals into the management of large public and private organizations; the decrease in the percentage of self-employed managers reflects the thinning of the ranks of small businessmen. Managers and administrators of farms have not been included in this breakdown.

neering professions in increasing numbers only with sustained private and public encouragement.

In the managerial occupations the position of women is weaker and reflects gains less sizable than those in the professions. In 1950 only 15 percent of the then 6.4 million managers and administrators were women, about half their percentage in the labor force. Most of them were self-employed: the owners or co-owners of small shops. By 1977 women managers had doubled in number and had increased their representation in the managerial category from 15 to 22 percent. The shift from self-employment to salaried employment proceeded at a higher rate, although it lagged far behind the rate for men.

A sensitive indicator of the change in the status of both women and blacks is provided by figures that show their representation among the professionals and managers employed by enterprises with more than 100 employees. These larger enterprises offer better-paying and securer jobs with, by definition, better prospects for promotion. A study conducted by the Conference Board shows striking percentage gains for both women and blacks in these jobs between 1966 and 1975 (better than twice the rate for white men), but again in modest absolute numbers that reflect the past history of discrimination against them [*see bottom illustration on opposite page*].

The ultimate summary indicator of relative status in the U.S. economy is of course earnings. In 1976, compared with overall median male earnings of $13,500, male professionals earned just under $17,000, self-employed professionals about $26,000 and physicians on the average about $50,000. Corresponding figures for male managers and administrators were an average of $16,700 for the entire category and $20,500 for salaried managers in manufacturing. According to a National Science Foundation survey for 1974, scientists and engineers earned $19,300, more than 40 percent above the median for all male workers, with $22,600 going to those who held managerial positions. The Federal Government paid scientists and engineers between $2,000 and $3,000 more per year than industry or universities. In all these positions women earned on the average $3,000 less than men.

There are signs now that the growth of these occupations, having arrived at one out of four jobs in the labor force, may be slowing down. It is significant that the ranks of the professoriate appear to have closed, that newly minted Ph.D.'s, even in the sciences, have great difficulty finding good academic appointments. The Federal Government has throttled down for more than a decade the increase in its outlays for university science, with a consequent reduction of enrollment in graduate schools of the

arts and the sciences. The effort to put brakes on the rising cost of health services as well indicates that the steep rise in the number and proportion of professionals and salaried managers will moderate and may even decline.

The full significance of the professionalization of the top fourth, the commanding ranks, of the country's labor force remains to be studied and comprehended. It is plain that the character and quality of the management of business has changed. Consider the difference between the management structure of the Standard Oil Company and that of its successor, the Exxon Corporation of today. Nelson Rockefeller once remarked that both of his grandfathers, John D. Rockefeller and Winthrop Aldrich, were high school dropouts. The ownership and directorate of Exxon depends on a bureaucracy of highly educated managers and professionals who hold 28 percent of the jobs on that company's payroll. A similar transformation in the character of personnel is to be seen in the swelling of the payroll of government—Federal, state and local—during the same period. Nothing can happen in organizations thus staffed, whether they are organized for profit or not, without the participation or at the very least the acquiescence of the middle management and its staff of professionals. As John F. Kennedy once said, in response to a proposal he found attractive: "I like it! But I don't know whether the Government of the United States will agree."

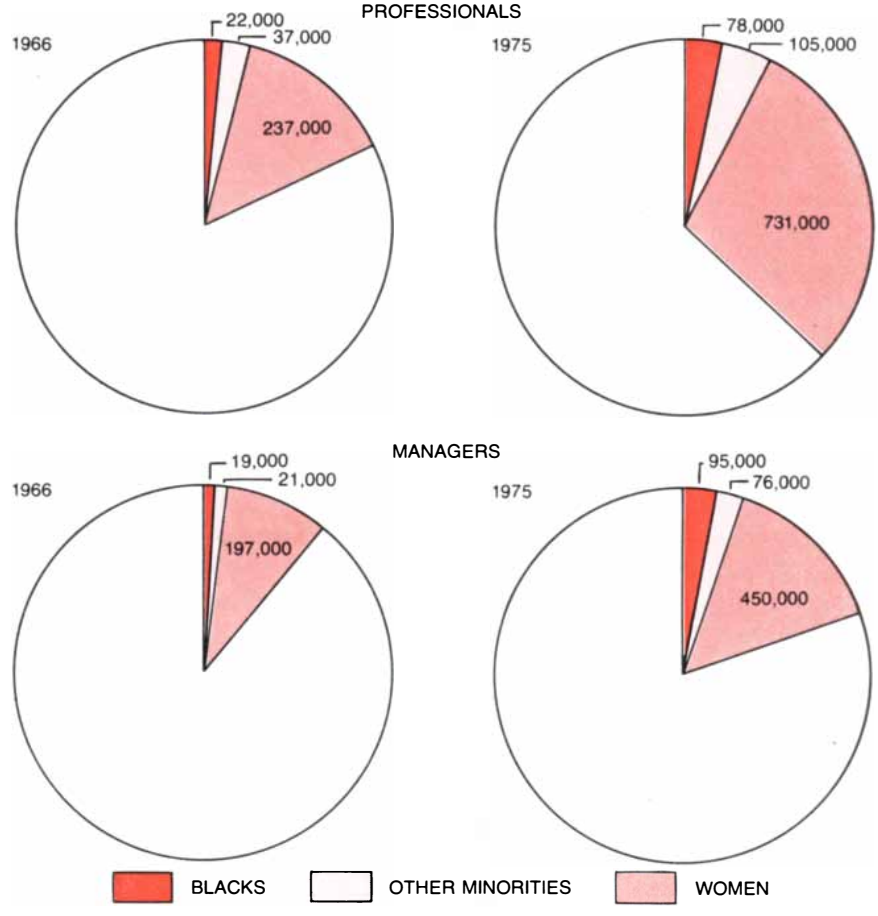
In the political arena the same kind of educated expertise has given a new quality to public discourse and a new dimension to political action. Ralph Nader has successfully fought the General Motors Corporation and has tilted with substantially the entire business establishment. He and his colleagues have shown consumers how to use law and politics to compel corporate enterprises to respond to consumer dissatisfaction. Rachel Carson and the environmentalists who came after her have enlisted legislatures and the courts to halt new capital construction, particularly by the public-utilities industry, wherever a scenic view might be spoiled, an endangered species threatened or a beach polluted. The continued vulnerability of the U.S. energy economy is a testimony to their effectiveness.

At the same time the U.S. establishment finds itself beleaguered on its other flank by neoconservative activists such as Milton Friedman, Irving Kristol and Alfred Laffer. Their advocacy of broad-scale deregulation of business, "Proposition 13" cuts in tax levies and reduction of income transfers to the poor could, if they were enacted, exacerbate existing tensions and conflicts in American society and reduce the space for compromise and accommodation. Either way the new movers and shakers in American society have demonstrated an impressive capacity for exciting and molding public opinion.

No establishment can ensure its survival without the recruitment of talent from outside its ranks. If it does not succeed in the full co-optation of the newcomers, however, it may leave itself vulnerable to them as they proceed to advance their own interests and aims and succeed in usurping decision-making power. It remains to be seen whether the demands of the new American mandarins can be met without subversion of the risk-taking, profit-seeking, efficiency criteria on which the country's business system has long rested.

	PROFIT		NOT FOR PROFIT	
	MANAGERS	PROFESSIONALS	MANAGERS	PROFESSIONALS
EDUCATION AND CULTURE	*	.441	.391	3.987
HEALTH AND HELPING	*	.700	.175	2.562
BUSINESS AND GOVERNMENT	8.320	4.161	.767	.781
TOTAL (MILLIONS OF PEOPLE)	8.320	5.302	1.333	7.330

BUSINESS AND GOVERNMENT ORGANIZATIONS employed 63 percent of the 22.3 million people classified as managers and professionals in 1977. The breakdown of the three major areas of industry into profit and not-for-profit sectors reveals the substantial role that government and private not-for-profit organizations play in the total economy of the U.S.: in 1977 they employed nearly 40 percent of the managers and professionals. Here 700,000 of the professionals involved in health care have been placed in profit sector under assumption that they are self-employed; categories with a small number of people are marked with asterisk.



NUMBERS OF BLACKS, OTHER MINORITIES AND WOMEN working as professionals and managers in firms with at least 100 employees are given for the years 1966 and 1975. Over this 10-year period the total number of professionals employed in such firms increased from 1.692 million to 2.440 million, while the total number of managers increased from 2.083 million to 3.180 million. The rates of increase in the numbers of women, blacks and other minorities employed as professionals and managers during this period are several times the rate recorded for white males. Absolute numbers for 1966 are so low, however (reflecting past history of discrimination against these groups), that gains seem more substantial than they really are.

Apollo Objects

They are asteroidlike bodies whose orbits cross the orbit of the earth. Some 30 of them are now known. They may be nuclei of comets that have lost their volatile components

by George W. Wetherill

In 1937 a body about a kilometer in diameter, later named Hermes, passed within 800,000 kilometers of the earth, no more than twice the distance of the moon. It has not been seen again. About once in every century a similar object can be expected to travel past the earth at less than the lunar distance. And once in every 250,000 years, on the average, the earth and such a body will collide. The impact of the collision will release energy equivalent to 10,000 10-megaton hydrogen bombs and will make a crater some 20 kilometers in diameter. Fortunately such catastrophes are so infrequent that none has been recorded within human history.

Hermes, the asteroid discovered crossing the earth's orbit in 1937, was only the third object of its class to be identified. The first such object had been discovered five years earlier, in 1932, by Karl Reinmuth of the University of Heidelberg in the course of a photographic search for ordinary asteroids. Similar objects had undoubtedly been photographed still earlier, but they had never been properly identified. The rapid motion of an earth-approaching body makes a trail on a photographic plate that is not readily distinguishable from the trail of a meteor or an artifact of photography. Reinmuth was the first to show that the object he had photographed was an asteroid whose elliptical path crossed the earth's orbit before reaching its perihelion (closest approach to the sun). The object was given the provisional designation 1932 HA and was subsequently named Apollo for the Greek solar deity. The term Apollo objects was later given to the entire class of bodies with earth-crossing orbits.

In the same year, by coincidence, another small asteroid, Amor, was discovered with an orbit similar to Apollo's except that its perihelion was just outside the earth's orbit at a distance of 1.08 astronomical units from the sun. (One astronomical unit, abbreviated A.U., is defined as the mean distance between the sun and the earth.) It was subsequently recognized that the orbits of objects with perihelions so close to 1 A.U. will often evolve into earth-crossing orbits. Astronomers now designate the class of bodies with perihelions smaller than the somewhat arbitrary value of 1.3 A.U. as Apollo-Amor objects.

Since 1932 a total of about 28 earth-crossing Apollo objects and a slightly smaller number of Amor objects have been discovered. The qualification "about" is necessary because in a few cases the orbits were not determined well enough for it to be certain that the objects belonged in the Apollo-Amor class. Even when an orbit seemed well defined, the usual fate of the first Apollo-Amor objects to be discovered was to be promptly lost. Apollo itself was not relocated until 1973, and Adonis, the second such body to be identified (in 1936), was not sighted again for 40 years, during which period it had crossed the earth's orbit more than 30 times.

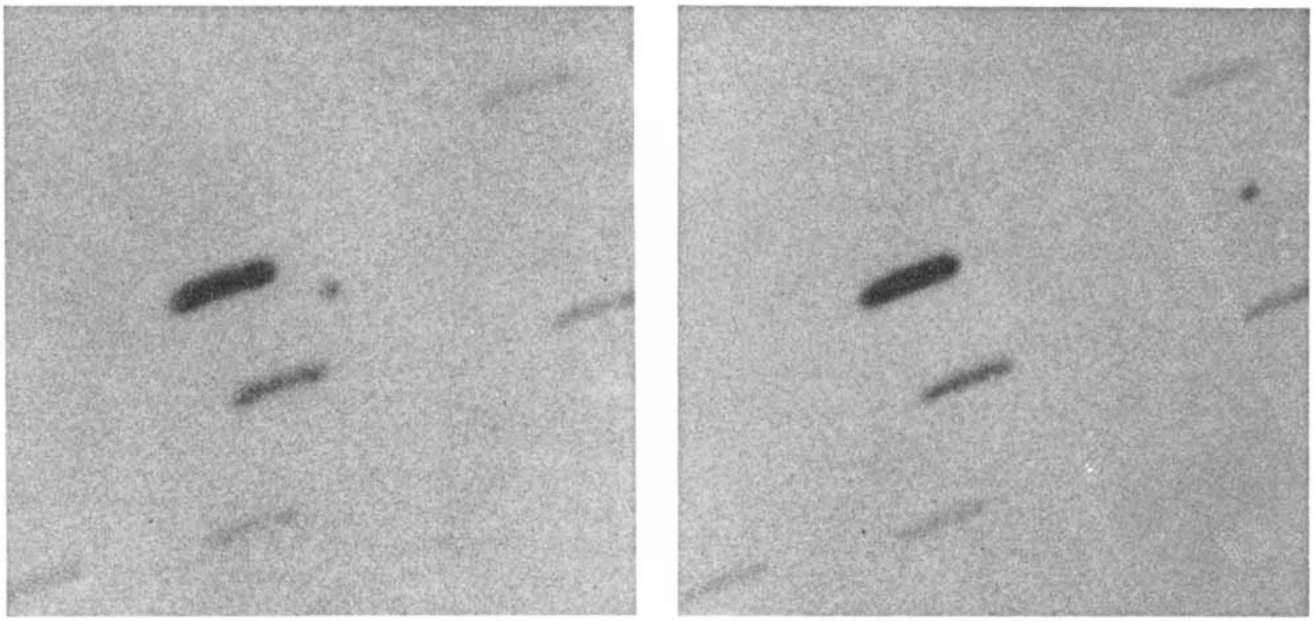
Until recently scant attention was paid to Apollo-Amor objects even by the minority of astronomers specializing in planetary studies. It was only gradually recognized that these small asteroids have an importance to earth and planetary science all out of proportion to their size and number. Bodies of the Apollo-Amor type have been the principal producers of craters larger than five kilometers in diameter on the earth, the moon, Mercury, Venus and Mars (with the possible exception of Mars). The surface density of such craters in turn is the principal means available for determining the relative age of different regions on these neighbors in space. Crater counts also make it possible to correlate the geology of one planet with that of another.

A large proportion, probably most, of the meteorites falling on the earth are fragments of Apollo objects produced in collisions as the eccentric orbits of the objects carry them through the asteroid belt, which lies between the orbits of Mars and Jupiter. Meteorites are the most important source of detailed data on the early history of the solar system. The interpretation of these data calls for information about the sources of the meteorites.

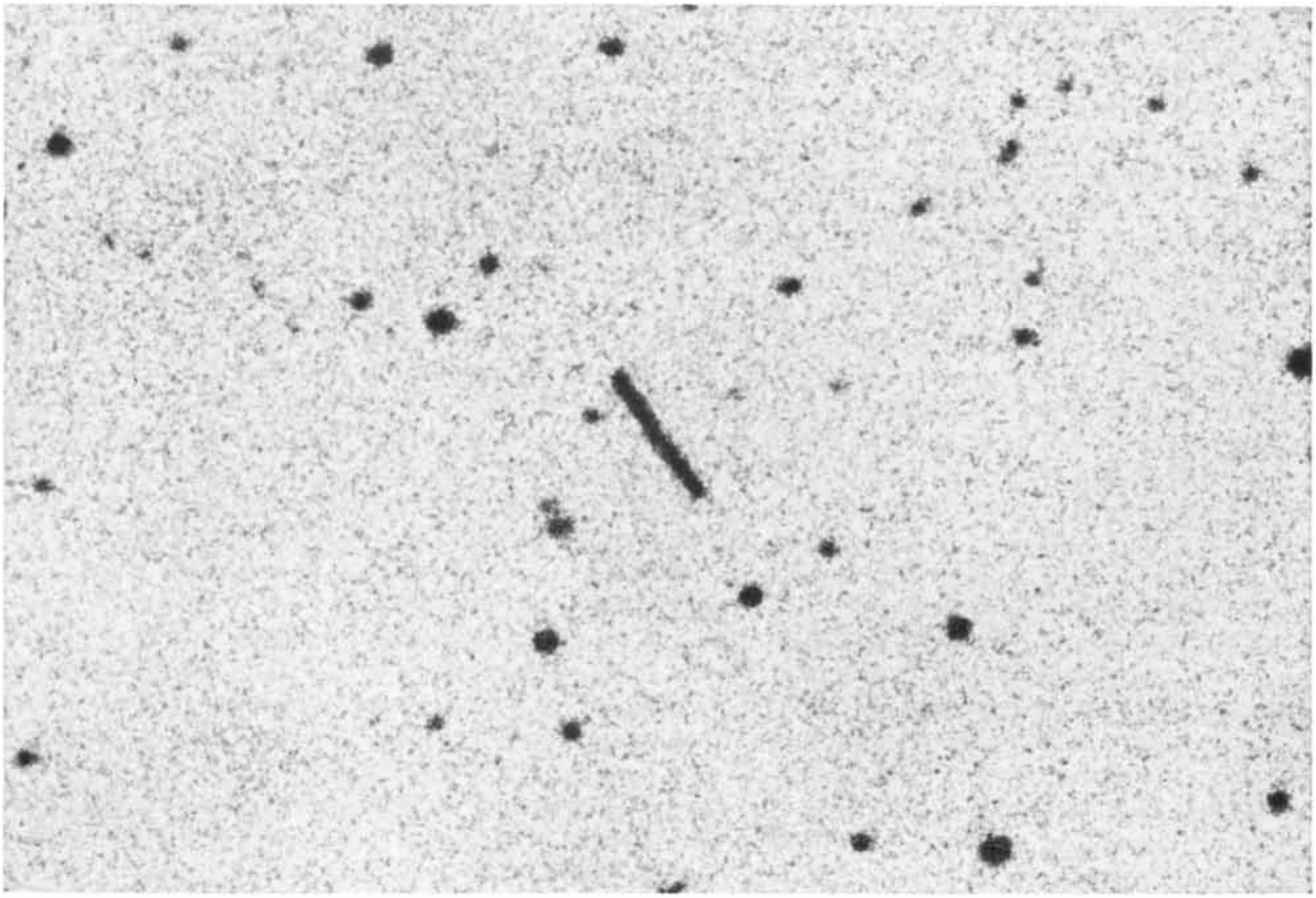
One can say with some confidence that the Apollo-Amor objects are not random samples of bodies somehow ejected from the broad asteroid belt, which occupies a region roughly between 2.1 and 3.5 A.U. from the sun. The first asteroid, Ceres, was discovered in 1801 by Giuseppe Piazzi, an Italian monk. The largest asteroid known, Ceres is about 1,000 kilometers in diameter and travels in an orbit about 2.8 A.U. from the sun. More than 2,000 asteroids have now been catalogued, the great majority of which lie beyond the orbit of Mars (1.5 A.U.). The asteroid belt is estimated to include a total of some 400,000 objects more than a kilometer in diameter. As we shall see, some of these bodies may become Apollos. On the other hand, according to a recent and perhaps surprising hypothesis, many Apollo-Amor bodies are not actually asteroids at all but are the outgassed remnants of comets, the most primitive objects in the solar system.

Within the past decade interest in Apollo-Amor objects has greatly increased. When astronomers observe a fast-moving object on their photographic plates, they often quickly report their discovery to the International Astronomical Union Central Bureau for Astronomical Telegrams and to other astronomers. Ephemerides indicating location, velocity and direction of travel are calculated and distributed so that follow-up observations can be made. In some cases the newly discovered object can be recognized on old plates. Once an object has been identified on several plates its orbit can usually be calculated with sufficient accuracy to ensure its being found on future close approaches.

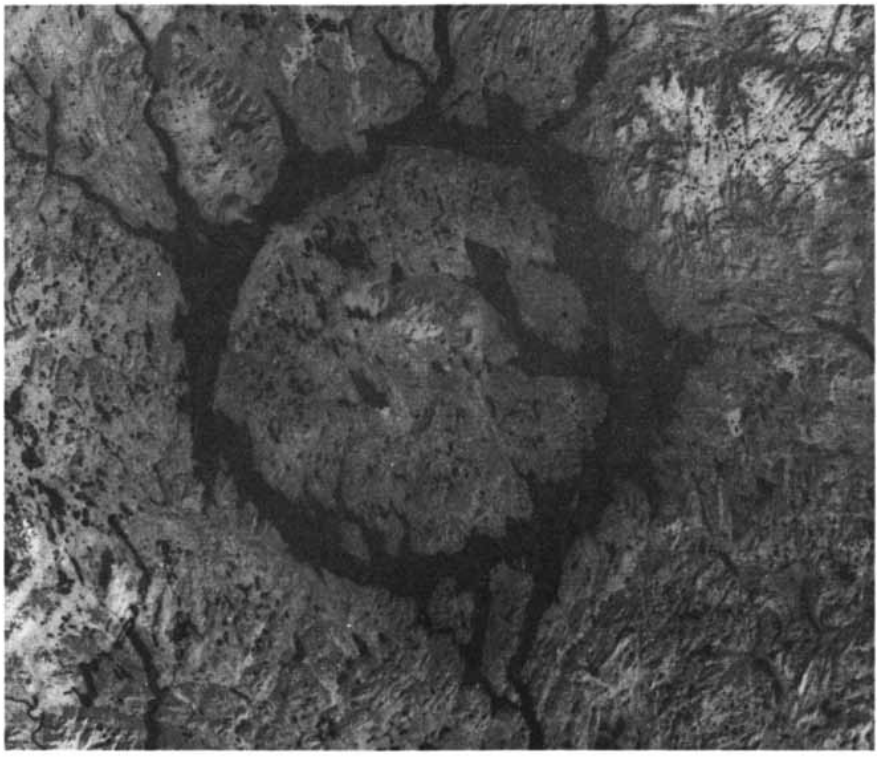
In 1973 the first observational program dedicated to a systematic search for Apollo-Amor objects was begun at the Palomar Observatory by Eugene M. Shoemaker and Eleanor F. Helin of the California Institute of Technology. Working with a wide-field Schmidt telescope of medium size about four nights a month, they have discovered five



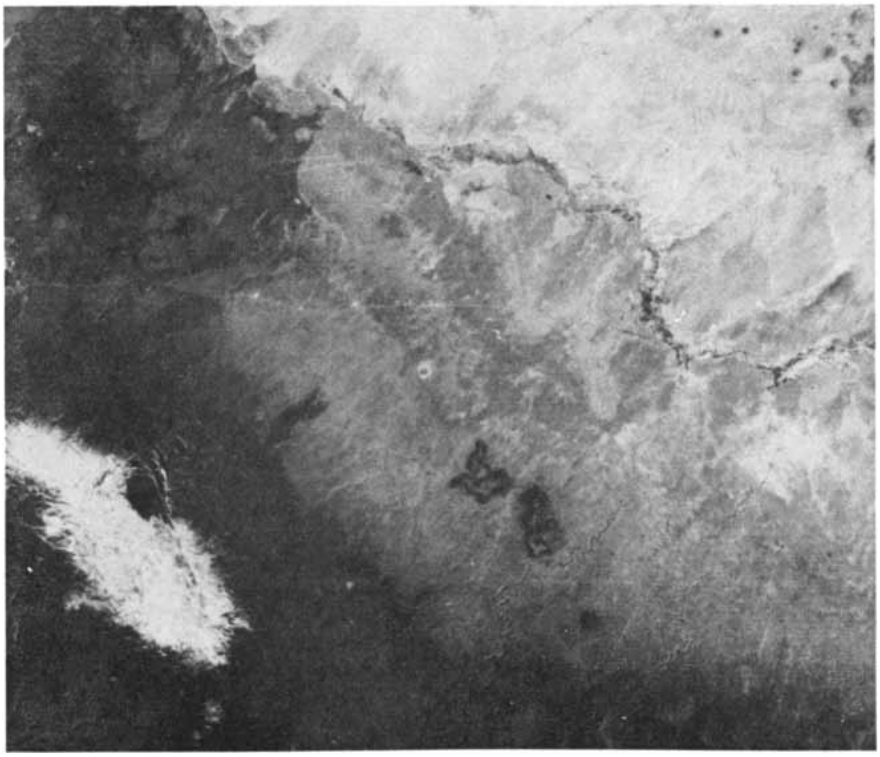
APOLLO, the first Apollo object known, was discovered in 1932. It was then lost. These negative prints record its rediscovery 41 years later, on March 28, 1973. Apollo is the faint round image in the center of the picture at the left. In the picture at the right, which was made an hour later, it has moved to the northwest. The bright surrounding stars in the constellation Virgo are streaks because the telescope was guided at the Apollo object's predicted speed during the two 25-minute exposures. If the star images had been held fixed, the track of the faint Apollo object, which was then 100 million kilometers from the earth, would have left no visible record. The recovery photographs were made with the Harvard College Observatory's 155-centimeter reflector by Richard F. McCrosky and Cheng-Yuan Shao, who were prepared to conduct a search extending over several months. By remarkable coincidence Apollo was found in the very first photograph.



2100 RA-SHALOM, which made the short streak in this negative print, was the third of four Apollo objects discovered last year. It was found by Eleanor F. Helin of the California Institute of Technology on a plate made on September 10 with a 46-centimeter Schmidt telescope on Palomar Mountain. The finding was actually a rediscovery of an object photographed several years earlier and then lost. The number 2100 signifies that the object is the 2,100th to be listed in the official catalogue of asteroids of all kinds. "Ra," the name of the ancient Egyptian sun god, and "Shalom," the Hebrew word for peace, were bestowed on the object at the time of the Camp David meeting.



GIANT ERODED CRATER in northeastern Quebec was formed about 210 million years ago by the impact of a large body similar to the Apollos. To excavate such a crater, which is about 70 kilometers across, the asteroid must have had a diameter of some three kilometers. The photograph was made by the Earth Resources Technology Satellite (ERTS). The visibility of the crater is enhanced by the dark water of Lake Manicouagan, which now fills the crater's circular depression. In the center of the formation is a peak of shocked rock. Many craters in Canada were preserved under layers of sediment that were later scraped off by glacial action.



METEOR CRATER, near Flagstaff, Ariz., is shown at the same scale as the crater at the top of the page. Barely visible in the center of the picture, Meteor Crater is some 1,300 meters across. It appears to have been formed about 25,000 years ago by an iron meteorite less than 100 meters in diameter. The impact released energy equivalent to an explosion of 10 megatons.

Apollo objects and one Amor object, together with a number of other interesting bodies, including a new comet.

Several other observing programs employing larger Schmidt telescopes, although not expressly directed toward Apollo-Amor searches, have also contributed significantly. Working with the 1.2-meter Schmidt telescope on Palomar Mountain, Charles Kowal has discovered five Apollo-Amor objects, and a group conducting an all-sky survey at the European Southern Observatory in Chile has discovered three.

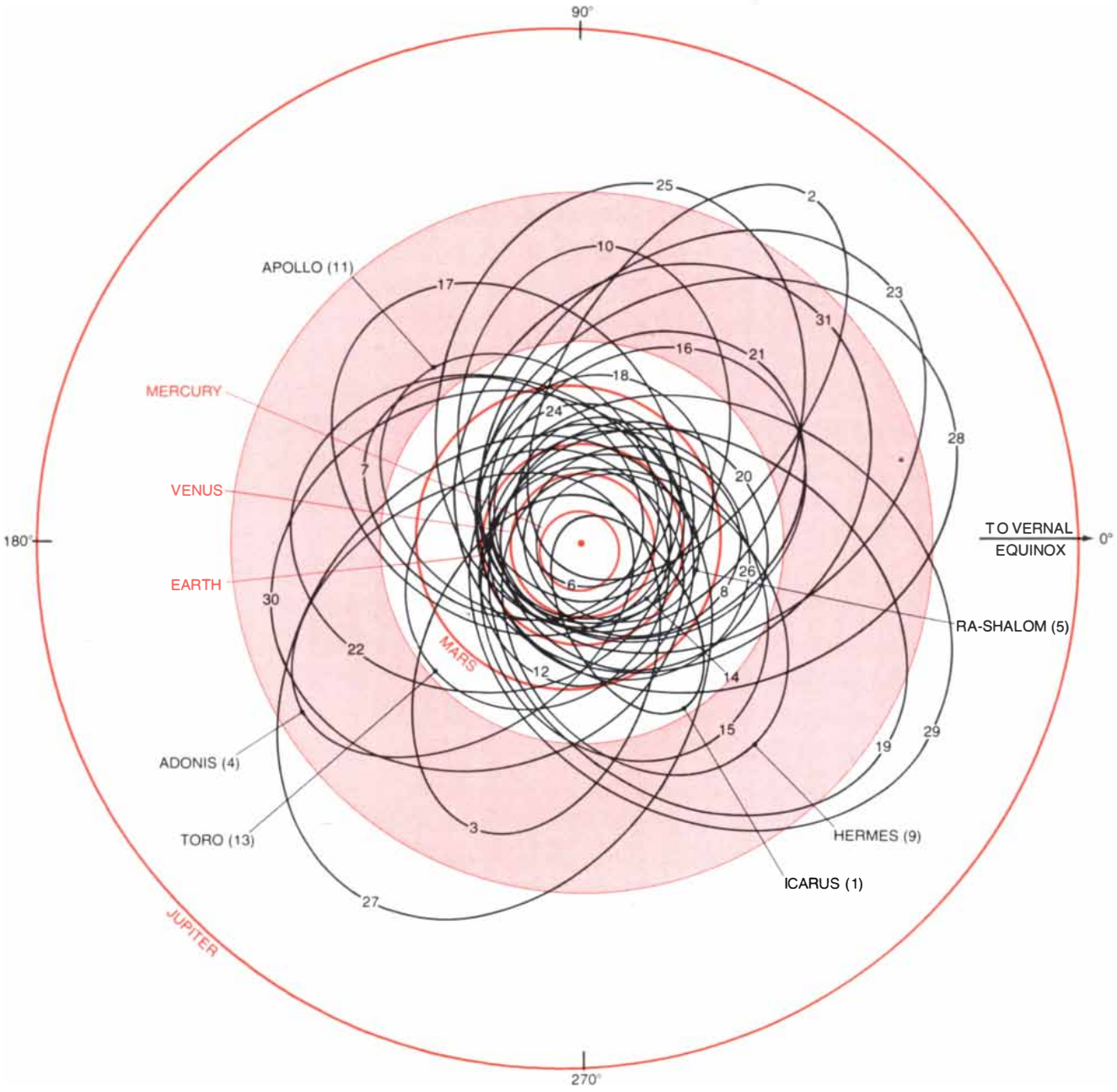
In contrast to the situation of a decade ago almost half of the Apollo-Amor objects are fully certified asteroids, designated by number. These days few are hopelessly lost. Most of the 40-odd objects have highly eccentric orbits that extend out into the main asteroid belt, and in half a dozen cases even beyond it. At least three of the objects, however, have an orbit of only moderate eccentricity, with an aphelion (greatest distance from the sun) inside the orbit of Mars.

Any quantitative discussion of the role Apollo-Amor bodies play in planetary cratering or as a source of meteorites calls for a knowledge of their size and abundance. Their angular diameter is well below the limits of telescope resolution, so that their dimensions must be determined indirectly. That can be done by measuring their geometric albedo: the fraction they reflect of the solar radiation that falls on them. By combining an object's albedo and its measured brightness with its known distance from the earth and the sun one can compute its cross-sectional area and therefore, assuming it to be roughly spherical, its approximate size.

The albedo can be determined in several ways. In a steady state the incident energy that is not reflected from an object must be absorbed and then reradiated, principally at infrared wavelengths. Hence by measuring the brightness of an object at infrared wavelengths and comparing it with the brightness at visible wavelengths one can determine the reflectivity of the object. Laboratory measurements have also demonstrated that the degree of polarization of light reflected from a surface depends jointly on the angle of reflection and the albedo. In addition the "color" of an asteroid can be determined by measuring its relative brightness at different wavelengths in the ultraviolet, visible and near-infrared regions of the spectrum. Such measurements often allow at least a tentative identification of asteroids with plausible assemblages of minerals found in meteorites and in terrestrial and lunar rocks. Laboratory measurements of the albedo of similar mineral assemblages can provide another check on an asteroid's albedo.

It is easy to find fault with any of these methods for inferring the albedo of asteroids, but the remarkable fact is that many of the albedos have been determined by more than one of the methods and almost all the results are in good agreement. Most of the measurements have been made on typical bodies in the asteroid belt. Only 11 Apollo and Amor objects have so far been subjected to albedo measurements. For this small sample, however, the geometric albedos are similar to those of ordinary asteroids, ranging from about .05 to .2. (For comparison the geometric albedo of the moon is .12.) If an albedo of .16 is taken as being typical of Apollo-Amor objects, the majority of those discovered turn out to have a diameter of between one kilometer and two kilometers. With an albedo of .16 an object about a kilometer in diameter will have an astronomical magnitude of +18 when it is located at a distance of 1 A.U. from both the earth and the sun. The magnitude of an asteroid adjusted to this standard condition is termed its absolute magnitude.

The total number of Apollos and Amors with an absolute magnitude brighter than +18 (that is, with a diameter of more than a kilometer) can be estimated by comparing the number actually discovered with an estimate of how complete the telescopic search has been. The first such attempt was made by Ernst J. Öpik of the Armagh Observatory in Ireland in 1963, when 10 Apollos were known. Öpik concluded that there must



ORBITS OF 31 APOLLO OBJECTS are shown in relation to the orbits of Mercury, Venus, the earth, Mars and Jupiter. They are numbered in the order of their increasing distance from the sun at perihelion, or closest approach to the sun, which is also the order in which they are listed in the table on page 64. The asteroid that comes closest to the sun is Icarus. Only three of the Apollo objects (Ra-Shalom, No. 6 and No. 14) occupy orbits that lie entirely within the orbit of Mars. The majority pass through the main asteroid belt (*broad band in color*), which is estimated to include nearly half a million objects upward of one kilometer in diameter. Fragments of Apollo objects that have collided with objects in the asteroid belt may supply a large proportion of the meteorites that fall on the surface of the earth.

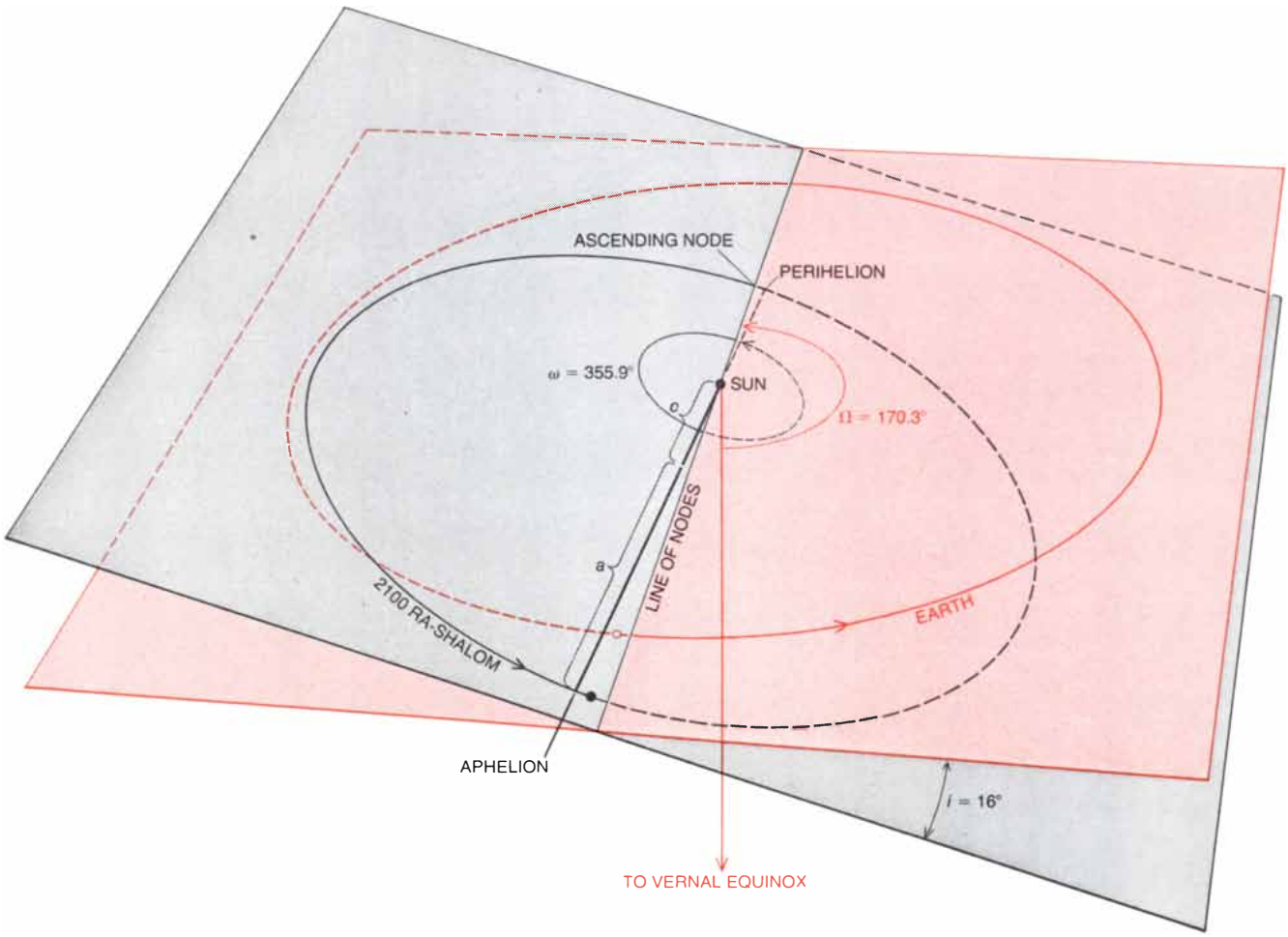
be at least 43 Apollos and that "there may be many more." Since about 28 Apollos have been discovered, and the discovery rate, now about four per year, is increasing, there must indeed be many more than 43.

A more recent estimate on the basis of the completeness-of-search method has been made by Shoemaker, who has concluded that the total number of Apollo objects brighter than magnitude +18 is 750 ± 300. In 1967 a rather different estimation approach was taken by Fred L. Whipple of Harvard University, who reasoned that if the total number of Apollos was small, it would be highly probable for some of them to be rediscovered by chance. As of 1967 no chance rediscoveries had yet been made. The algebra of combinations and permutations can be used to set a probable lower limit on the total number of Apollo objects. Whipple first estimated there were more than 50, and a repetition of this calculation in 1973 gave a value of 100. On the basis of the most recent data, including the chance rediscovery in 1978 of one Apollo (Ra-Shalom), the method argues strongly against there being as few as 200 and leads to an estimate in agreement with Shoemaker's.

It is a matter of chance whether or not an Apollo object will strike the earth. At the present time none of the known Apollo objects is on a collision course with the earth. All the Apollo-Amors, however, are under the continuous gravitational influence of the nearby planets, particularly Jupiter, which causes the asteroidal orbits to precess. As a result of precession the major axis of an elliptical orbit is gradually rotated through 360 degrees in space. This means that all objects with a perihelion inside the earth's orbit and an aphelion beyond the earth's orbit will sooner or later be in an orbit that intersects the orbit of the earth. Therefore a typical Apollo object will find itself in an orbit that intersects the earth's orbit about once in every 5,000 years.

In most cases the earth will be at some other point in its orbit when the Apollo object is at the point of intersection, and so the two bodies will not collide. There is, however, a calculable probability that both the earth and the Apollo object will be at the point of intersection simultaneously. For any given Apollo object the collision probability is only about 5×10^{-9} per year, or once in 200 million years. If we assume that there are between 750 and 1,000 Apollo objects larger than a kilometer in diameter, it works out that roughly four should strike the earth in every million years. Although such an impact rate is low on the 5,000-year scale of recorded human history, it is high on the scale of billions of years of geologic time.

No catastrophe resulting from an Apollo impact has happened within recorded human history. The most recent known encounter with an asteroidal object of significant size took place between 25,000 and 50,000 years ago, when a crater slightly more than a kilometer in diameter and nearly 200 meters deep, Meteor Crater in Arizona, was excavated by an iron asteroidal



ORBITS OF THE EARTH AND APOLLO OBJECTS often pass through each other like two links in a chain. Since the orbits of the Apollo objects precess fairly rapidly, however, their orientation with respect to the earth's orbit can change significantly in a few hundred years. The orbital plane of 2100 Ra-Shalom is inclined by 16 degrees (the value *i*) to the orbital plane of the earth. The relative positions of the two bodies are shown for the night of Ra-Shalom's discovery, September 10. The size and shape of an asteroid's elliptical orbit are given by the length of the semimajor axis, *a*, and the eccentricity of the ellipse, *e*, defined as *c*/*a*, where *c* is the distance from the sun to the center of the ellipse. The orientation of the asteroid's orbital plane with respect to the earth's is given by Ω, the longitude of the ascending node, defined as the angular distance measured eastward in the plane of the earth's orbit, or ecliptic, from the vernal equinox to the

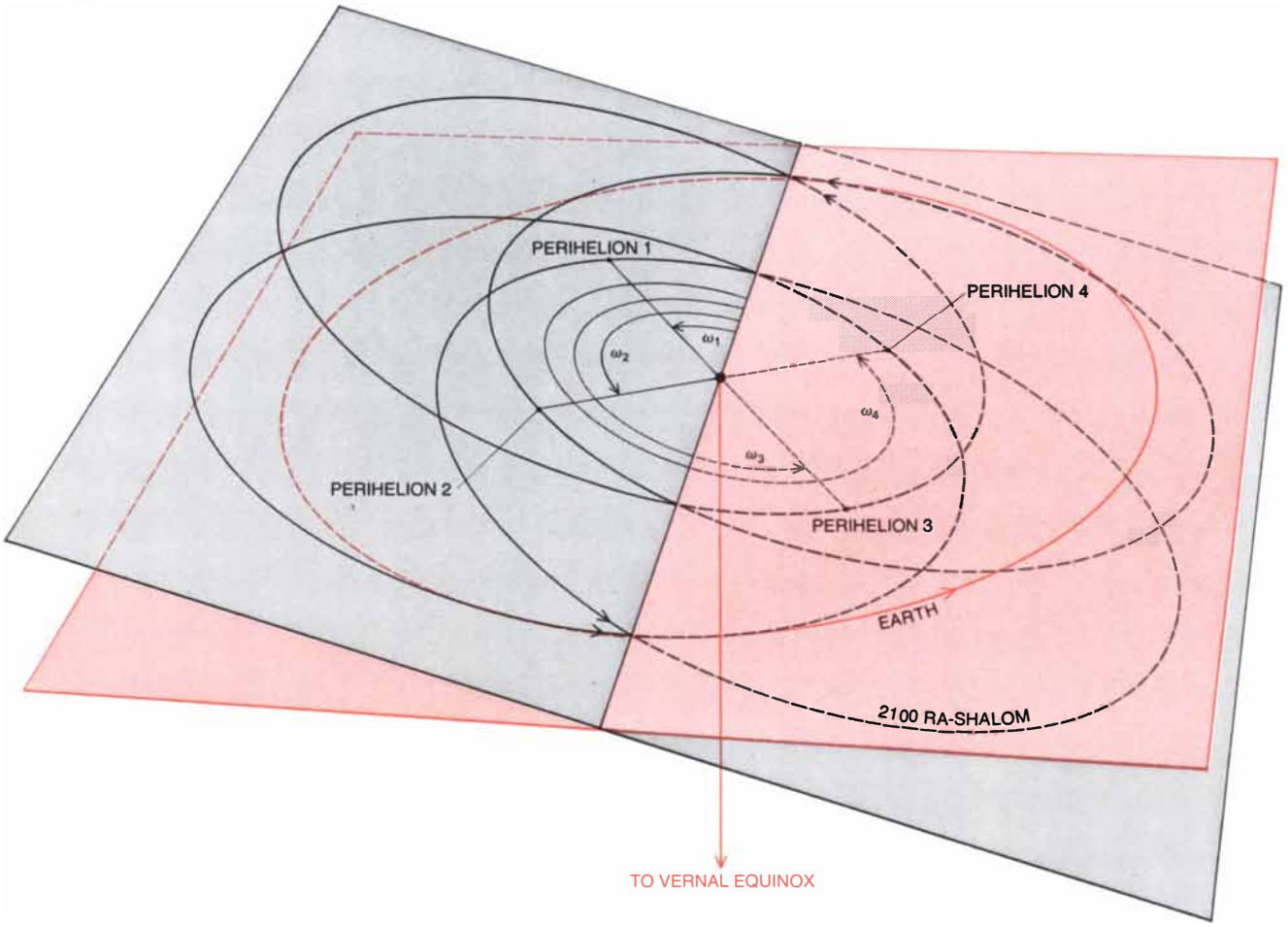
body that was probably not more than 100 meters across. The youngest-known crater larger than 10 kilometers in diameter provides the basin for Lake Bosumtwi in Ghana. The event that created it occurred about 1.3 million years ago. Craters as small as Meteor Crater are rapidly eroded. Only the largest craters leave any trace after 500,000 years. If it were not for the erosional effects of weathering, geological processes and living organisms, the earth would look as heavily cratered as the moon.

When these rare but statistically inevitable collisions take place, the consequences are dramatic. In the reference frame of the earth the kinetic energy of a one-kilometer Apollo object of typical density (3.5 grams per cubic centimeter) and moving in a typical orbit is about 4×10^{27} ergs. That is 100,000 times the energy released by the detonation of a one-megaton nuclear warhead. All this energy must be dissipated in one way or another on impact with the earth. Only a small fraction of the total energy is required to vaporize the impacting body, so that most of the energy will be expended in pulverizing the material of the earth's surface and ejecting it outward at high velocity. Hence the impact will excavate a crater similar to the craters that dominate the landscapes of the moon, Mars and Mercury.

There are no direct data on the dimensions of craters produced by explosions of 10^5 megatons. One can, however, extrapolate empirical scaling laws based on much smaller explosions and carry out computer simulations of explosions of asteroidal magnitude. Such calculations indicate that an impact releasing 4×10^{27} ergs of energy will make a crater about 22 kilometers in diameter.

If one accepts the estimate that the earth is struck roughly four times in a million years by objects capable of producing craters 20 kilometers in diameter, where, one may ask, is the evidence? One might expect the earth to be liberally peppered with craters. It turns out that the evidence is there, but until rather recently it escaped attention.

Close inspection of aerial photographs discloses many circular features, often quite faint, on the landscape all over the world. Detailed geological study has shown that a number of them are remnants of impact craters originally larger than 20 kilometers in diameter. Their origin is demonstrated by the discovery of geological structures one might expect to find in deeply eroded explosion craters. There is also clear evidence that rocks in the region have been subjected to extreme shock. The evidence includes the presence of igneous rocks that have recrystallized after having been melted by sudden impact and the presence of forms of quartz, such as coesite and stishovite, that can be created only by a combination of high temperature and high pressure. Macroscopic evidence includes "shatter cones," structures of quartzite that flare outward and downward, away from the direction of impact. In some cases the chemical "signature" of a nonterrestrial impacting body can be identified in the crater ejecta. Only rarely, however, does



point where the asteroid crosses the ecliptic from south to north. The orientation of the asteroid's ellipse in its orbital plane is given by ω, which is the angle between the ascending node and the perihelion point, measured in the asteroid's direction of motion. For Ra-Shalom a is .83 astronomical unit (or .83 times the mean distance from the earth to the sun), e is .436, Ω is 170.3 degrees and ω is 355.9 degrees. Perturbations caused primarily by the gravitational attraction of the planet Jupiter will induce the angle ω to rotate through 360 degrees in about 10,000 years. The cloverleaf pattern in the illustration at the right shows that at four values of ω, separated by a few thousand years, the orbit of the asteroid will intersect the orbit of the earth. Simultaneously, and at about the same rate, the line of nodes will also precess through 360 degrees (not shown). Approximately once every 250 million years a typical Apollo object and the earth will collide.

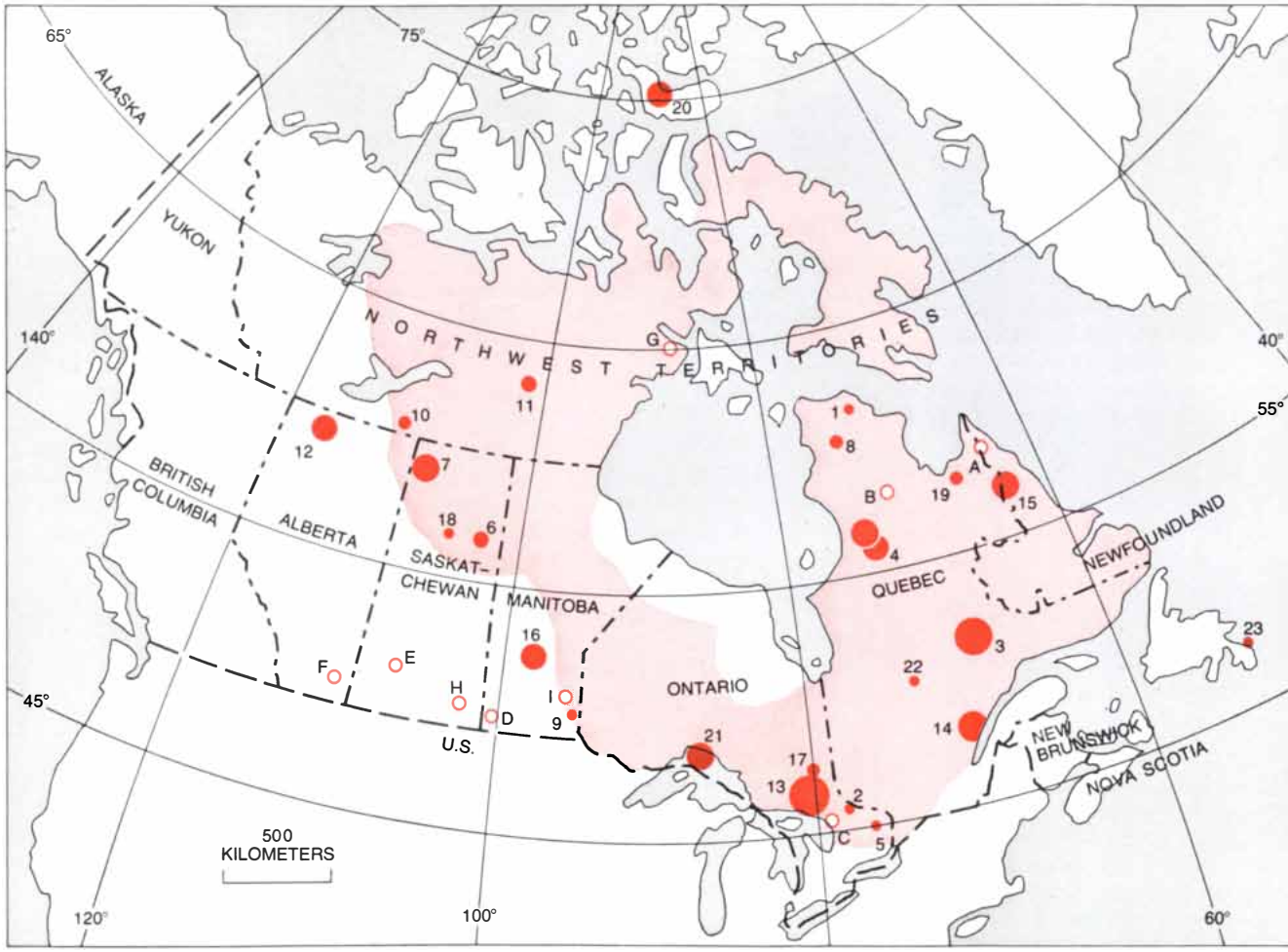
one find actual fragments of the impacting object; it will have been vaporized by the temperatures reached in the explosion.

Such telltale evidence of impact origin has now led to the identification of more than 50 impact structures, most of them, not surprisingly, in regions of the earth inhabited by geologists with an interest in such structures. About half of the well-established craters have been found in central and southeastern Canada, which constitutes only about 1 percent of the earth's surface. Since there is no natural way in which extraterrestrial bodies can be targeted on a small region of the earth, the abundance of Canadian craters can only be attributed to the diligence of the search program mounted by the Canadian Department of Energy, Mines and Resources in Ottawa.

Most of the well-studied craters are geologically young, being no more than a few hundred million years old. Only two Precambrian craters are known (one near Sudbury in Ontario and the other near Vredefort in South Africa), even though Precambrian time encompasses more than 80 percent of the earth's history. Unless a crater is unusually large all traces of its presence will be eroded away in less than 600 million years. Moreover, craters made by asteroidal bodies that fall in the oceans, which cover more than 70 percent of the earth's surface, are understandably difficult to identify. All things considered, it is now recognized that the rate of crater formation per unit area on the earth is at least as high as it is on the moon.

Many of the identified terrestrial craters are no more than 100 meters in diameter and so overlap the size range of craters that can be excavated by the largest meteorites known. They can properly be called meteorite craters. About half of the known terrestrial craters, however, are more than 10 kilometers in diameter and therefore must have been made by bodies much larger than any meteorites yet found. In other words, they must have been created by the impact of bodies comparable in size to the telescopically visible Apollo objects. A plausible estimate, based on the distribution of known craters, is that about 1,500 Apollo-like objects have struck the earth since the end of the Precambrian era 600 million years ago. Considering the present uncertainty in the total number, size and velocity of Apollo objects, this estimate is in reasonable agreement with the estimate given above of four collisions per million years.

How can one be sure it is correct to identify the large craters on the earth and the moon with the only recently recognized Apollo objects? The answer is that objects that can make large craters must be large enough to be seen with telescopes. There are only two kinds of astronomically observable objects that cross the orbit of the earth:



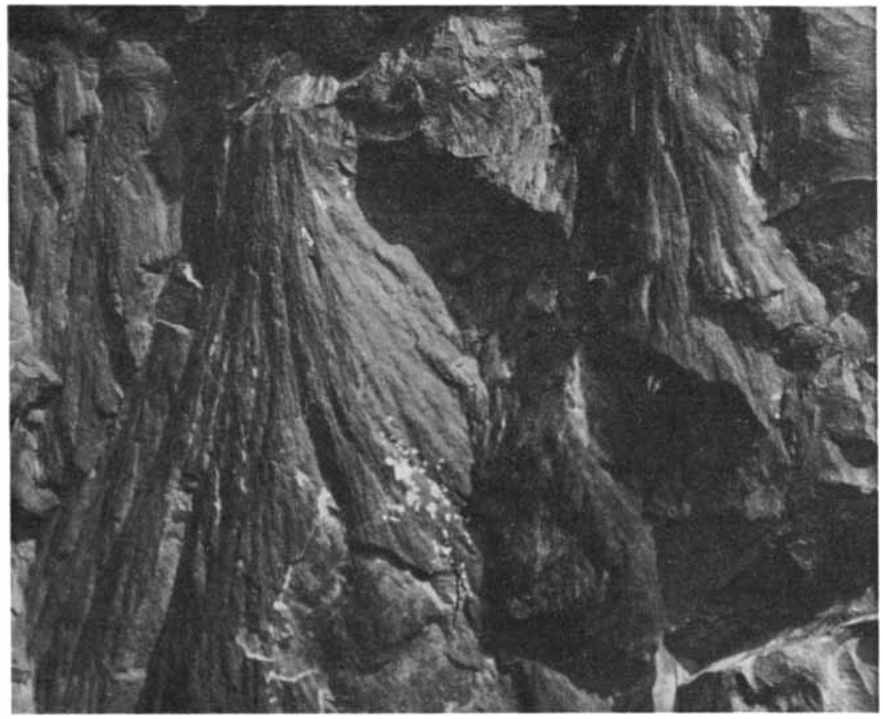
NEARLY HALF OF THE ANCIENT CRATERS that are well established have been found in central and eastern Canada, a region that represents only about 1 percent of the earth's land area. The filled circles show the location and relative size of 23 well-established impact craters, numbered in the order of their recognition; the open circles identify possible impact structures. The youngest-known crater (No. 1) is less than five million years old; the oldest-known (No. 13, near Sudbury, Ontario) is about 1.8 billion years old. The latter crater is one of only two craters that can be traced to the Precambrian era, which ended 600 million years ago. Crater No. 3 is the Manicouagan formation in the top photograph on page 56. Richard Grieve of the Canadian Department of Energy, Mines and Resources has made a count of the well-established ancient craters in North America and Europe; on the basis of his count one can estimate that in the past 600 million years 1,500 Apollo objects one kilometer in diameter or larger have struck the earth, about 70 percent of which presumably landed in the oceans. The map is based on a compilation made by Earth Physics Branch of the Department of Energy, Mines and Resources.

Apollo objects and comets. Unlike an asteroid, the solid nucleus of a comet contains ices or other substances that as the comet swings close to the sun volatilize and produce a nebulous coma, or head, and often a tail as well. Comas often expand to a diameter of thousands of kilometers. The diameter of the nucleus within the coma can be estimated by comparing the observed and the calculated values of the quantity of gas expected to be emitted as a function of the comet's distance from the sun. A more direct clue to size is the brightness of a comet nucleus when it is so far from the sun that the obscuring coma is absent. Such estimates indicate that comet nuclei are in the same size range as the one inferred for Apollo objects: from one kilometer to 10 kilometers.

Since both Apollo objects and comets are capable of striking the earth and the moon and leaving craters, how can one estimate the relative bombardment rates for the two types of objects? If one considers only estimates of the size and number of Apollo objects, it appears that they alone are plentiful enough to account for the number of large craters produced on the moon since the end of the flooding of the mare ("sea") regions by basaltic magmas some three billion years ago. And as we have seen, the estimated Apollo abundance also agrees with the estimated number of large craters produced on the earth since Precambrian times.

It can also be shown that if the nuclei of comets are no larger than Apollo objects, comets can have produced only a few percent of the larger craters observed on the moon. The fraction could be much bigger, however, if observed comets are actually larger than the prevailing estimates of their size and if as a result there are many unobserved comets more than a kilometer in diameter. In that case comet impacts could account for up to 35 percent of the larger lunar (and terrestrial) craters. Recently A. H. Delsemme of the University of Toledo and L'ubor Kresák of the Astronomical Institute at Bratislava in Czechoslovakia have independently presented evidence that the number of faint comets is actually lower than might be expected, which argues that the cometary contribution to cratering is indeed small.

Having drawn the distinction between Apollo objects and comets, I shall now blur it by mentioning a possibility to which I shall return in more detail. As was originally suggested by Öpik, many Apollo objects may be "extinct" comets, that is, the nonvolatile residues of formerly active comet nuclei that have been thoroughly outgassed by repeated approaches to the sun. If this is the case, many lunar and terrestrial craters are cometary in the sense that the comet responsible for them had earlier metamorphosed into an Apollo object and therefore is already included in the calculations of Apollo cratering rates.



SHATTER CONES, rocks that have been fractured in a conical pattern by impact, have been found at most well-established crater sites. These cones were found in an outcrop on the south shore of Keeley Lake near the Sudbury basin, the remnant of a crater whose original diameter is estimated to have been 140 kilometers. The structure is No. 13 on the map on opposite page.

In what way, if any, are the large crater-forming Apollo objects related to the myriad smaller bodies, ranging from large meteors down to micron-size particles, that continuously bombard the earth's atmosphere? It appears likely that these smaller bodies are derived from sources of several kinds. One can say quickly, however, that a large fraction of the observed meteors (as distinct from meteorites) are not related to objects of asteroidal appearance but are derived from observable comets. The cometary meteors are those seen in the periodic meteor showers, such as the Perseids, which follow the orbit of Comet Swift-Tuttle, and the Draconids, which are associated with the short-period Comet Giacobini-Zinner. Meteors are also associated with active comets that have been observed to disintegrate and disappear, for example Biela's comet, which gave rise to the Andromedids. It is therefore plausible that at least some of the meteor showers seemingly unrelated to known comets actually represent the debris of comets that disintegrated within the past few thousand years, while they were still active.

On the other hand, a large fraction of the observed meteors and fireballs bear no obvious relation to meteor showers. These include the meteorites that survive passage through the earth's atmosphere. It is very unlikely that any of the meteorites now in museum collections arrived as part of a shower, even though some could be remnants of dispersed cometary streams. One can determine how long meteoritic bodies have been in orbit by measuring the buildup in them of the products of nuclear reactions triggered by galactic cosmic rays. It turns out that almost all meteorites have been close to their present size of less than a meter for millions of years. Even if the meteorites originated in streams as comet remnants, they would have been dispersed into random orbits by gravitational interaction with planets over a few thousand years.

It is certain, however, that not all meteors are the remnants of comets. For example, one can show that a small but significant fraction of meteors, presumably consisting of a material strong enough to survive passage through the atmosphere, must represent fragments produced by collisions among ordinary asteroids in the main asteroid belt between Mars and Jupiter. As a result of gravitational perturbations induced primarily by Jupiter, some of the fragments will be "pumped up" into highly eccentric orbits that cross the orbit of the earth. Any fragments that happen to enter the earth's atmosphere will incandesce and be seen as meteors; those that survive and reach the ground will be meteorites. Obviously there is no possibility of directly identifying such objects with their origin in the asteroid belt by means of their orbits.

A good qualitative case can be made

for believing that some meteors and meteorites are direct fragments of Apollo objects. At least 19 of the Apollo objects have orbits that either enter the main asteroid belt or extend beyond it. Hence they will inevitably collide with the large volume of debris concentrated in that part of the solar system. In addition to several hundred large bodies with diameters of between 25 and 1,000 kilometers the region is estimated to contain as many as 10^{10} objects ranging in size from a kilometer down to a few meters. Collisions between these numerous small asteroids and Apollo objects must be frequent.

In a typical encounter the smaller object will make a crater on the larger Apollo. The collision will impart a velocity of about 50 meters per second to the ejected material, which will fly off into space because the escape velocity from a one-kilometer body is quite low (less than a meter per second). The additional velocity of 50 meters per second imparted to the debris will not, however, be sufficient to change its orbit appreciably from that of the Apollo object itself, which moves around the sun with a velocity of about 20 kilometers per second. Hence the collision fragments, given a few million years, will have a good chance of entering the earth's atmosphere.

If an Apollo object collides with an asteroid as large as 100 meters in diameter, the probability is high that the Apollo itself will be shattered. Then the swarm of remnants will be as likely as the original Apollo to hit the earth. The probable time scale for impact on the earth, however, will be much longer than the time needed to disperse the stream of fragments into independent and essentially random orbits of their own. As a result meteors and meteorites produced by either the cratering or the shattering of an Apollo object are much more likely to arrive individually and sporadically than to arrive as members of a coherent stream. This, of course, agrees with the observations.

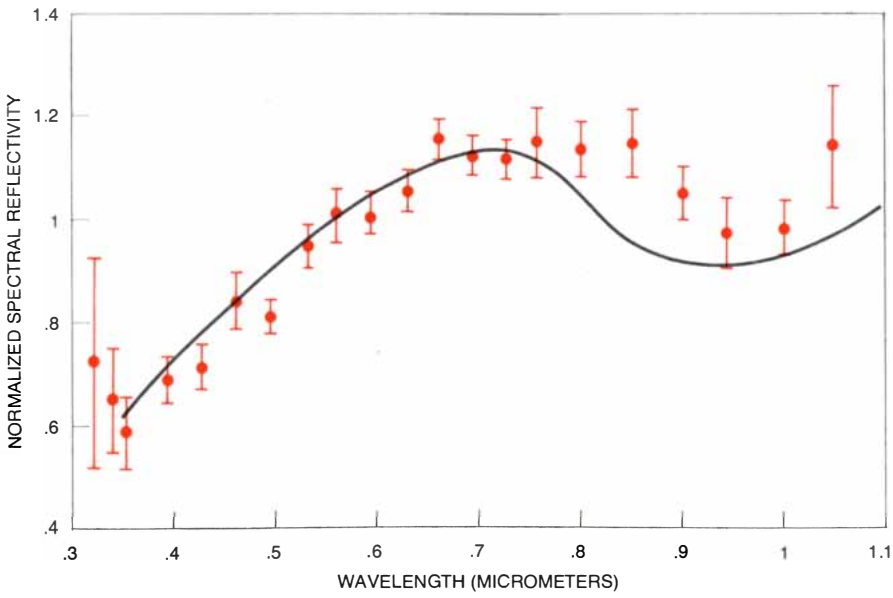
I have sought to quantify the qualitative argument I have just presented in order to see what fraction of the meteoric and meteoritic material estimated to bombard the earth annually might be accounted for by Apollo collisional events in the main asteroid belt. My estimate is that in the size range between 100 grams and 1,000 kilograms (one metric ton) some 400 metric tons of Apollo fragments should enter the earth's atmosphere every year. If the fragments have the mechanical properties of typical stony meteorites, about 100 of the 400 tons should survive their passage through the atmosphere and be recoverable as meteorites. Although this calculation is uncertain by a factor of at least 10, the 100-ton value is quite close to current estimates of the rate of impact of meteorites as determined by the actual recovery of meteorites, by bright-meteor camera networks and by the impact rate recorded by seismometers placed on the moon.

Unless Apollo material is peculiarly fragile and susceptible to destruction in the earth's atmosphere it seems virtually certain that a large fraction of the 2,000 meteorites in various collections are Apollo fragments, and it is possible that nearly all of them are. The competition from other possible sources is not strong. I have already argued that the recovered meteorites do not belong to streams associated with active comets. The only alternative sources remaining involve mechanisms capable of perturbing collisional fragments from bodies in the main asteroid belt into earth-crossing orbits by resonant gravitational interactions with neighboring planets, mainly Jupiter. The yield of meteorites from such mechanisms is difficult to calculate, but the most reasonable estimates suggest they would supply only about a tenth as much earth-bombarding material as the yield from Apollo fragmentation.

If we accept the fact that a great many, and perhaps most, meteorites have arisen from Apollo objects, what can the meteorites tell us about the nature of the Apollo objects? About 85 percent of all meteorites found belong to the class of stony objects known as chondrites. Chondrites are sometimes called undifferentiated meteorites because their content of the less volatile elements closely matches the abundance ratios of elements in the sun and in the solar system as a whole. In fact, the study of chondrites has provided most of the data on the general abundance of the elements and their isotopes, the age of the solar system and conditions in the early solar system.

One piece of supporting evidence relating Apollo objects to chondrites was provided by Clark R. Chapman of the Planetary Science Institute in Tucson, Thomas B. McCord and Michael J. Gaffey of the University of Hawaii and Carle Pieters of the Johnson Space Center in Houston, who measured the spectral reflectance of 1685 Toro, the first earth-crossing body to be so examined. The reflectivity of the Apollo object across the entire spectral region from a wavelength of .3 micrometer in the ultraviolet to one of 1.1 micrometers in the infrared closely follows the reflectivity of typical chondrites. After rising to a peak at about .7 micrometer both spectra show a dip in the vicinity of .95 micrometer before rising again. In the case of the chondrite the dip corresponds to a prominent absorption band attributable to pyroxene, a mineral consisting chiefly of silicates of magnesium and iron. One can infer that pyroxene is also plentiful in 1685 Toro.

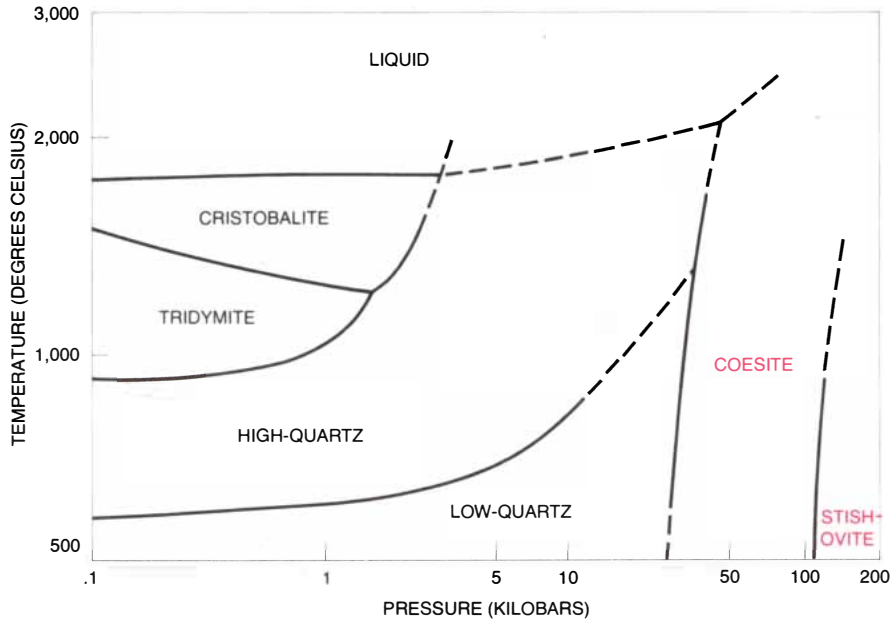
Albedo measurements of other Apollos support the belief that they are chondritic. Three Apollos of very low albedo are of particular interest, including the last two to be discovered: Ra-Shalom (whose name expresses a hope for universal peace) and 1978 SB, just discovered at the Crimean Astrophysical Observatory. The albedos of these two objects seem to be low enough to resemble



SPECTRAL REFLECTIVITY OF APOLLO OBJECT designated 1685 Toro (*vertical bars*) closely matches the reflectivity of a typical chondrite, the commonest type of stony meteorite (*solid curve*). The broad dip in the spectral curve of the meteorite in the region of .95 micrometer corresponds to a strong absorption band of the mineral pyroxene. Presumably pyroxene is a major constituent of 1685 Toro, the only Apollo studied in spectral detail. Measurements were made by Clark R. Chapman of the Planetary Science Institute, Thomas B. McCord and Michael J. Gaffey of the University of Hawaii and Carle Pieters of the Johnson Space Center.

the albedos of the carbonaceous chondrites, which are chemically the most undifferentiated type of meteorite.

Now we are ready to ask: Where do the Apollo objects come from? The first thought, that they may have existed ever since the solar system was formed 4.5 billion years ago, can easily be shown to be wrong. I have already remarked that the probability of an Apollo object's striking the earth is roughly 5×10^{-9} per year, and nearly the same probability applies to its striking Venus. Therefore after several hundred million years there would be few survivors of a population of earth-crossing objects. Nevertheless, could the present estimated family of 750 Apollos represent the remnant of a population that was initially huge? The answer is no. If a huge population had existed, the record of craters on the moon would reveal it. The flux of cratering projectiles has been nearly constant for the past three billion years. If anything, there is some evidence that the rate of cratering has actually increased slightly during the past 600 million years.



PRESENCE OF HIGH-DENSITY FORMS OF QUARTZ (coesite and stishovite) is a clear indication that terrestrial rocks have been subjected to high pressures by the shock of a high-velocity impact. Silica has a number of crystalline forms that are stable in different ranges of temperature and pressure; the two forms of quartz are the commonest. The high-temperature forms tridymite and cristobalite are metastably preserved by rapid cooling in some volcanic rocks. Coesite and stishovite, which can be created only at pressures above 30 kilobars (30,000 times atmospheric pressure), have been found in vicinity of many terrestrial impact craters.

It is therefore clear that Apollos are being injected into the inner solar system from some source at a fairly constant rate. In round numbers about 15 new Apollos larger than a kilometer must be supplied every million years to balance the estimated removal rate implied by four earth collisions and three Venus collisions per million years. In the same period possibly one Apollo will be removed by collision with some other body (Mercury, the moon or Mars), and the other Apollos will be perturbed into hyperbolic orbits and thereby ejected completely from the solar system. There would seem to be only two sources for the required 15 Apollos. Either they are small asteroids from the main asteroid belt that somehow have strayed into earth-crossing orbits or they are the outgassed residues of formerly active comets. It is now understood, at least qualitatively, how both of these sources might yield a supply of Apollo objects.

Small asteroids similar in size to the Apollo objects are continually produced in the main asteroid belt as the collision fragments of larger bodies. The small asteroids will themselves be destroyed by collisions, resulting in a steady-state number. The steady-state population of main-belt asteroids larger than a kilometer is about 400,000. It is estimated that they are being created and destroyed by collisions at the rate of about 400 per million years. If there were a way to place about 4 percent of the newly created bodies in earth-crossing orbits, it would suffice to maintain the Apollo population.

It is quite unlikely, however, that even this small fraction of objects can be ejected from the main asteroid belt and sent directly into earth-crossing orbits by the collisions that created them. The change in velocity needed to transfer material from a typical asteroid orbit into an earth-crossing orbit is about six kilometers per second. The associated kinetic energy is more than 10^{11} ergs per gram. The energy required to fragment a body of Apollo size is no more than about 10^{8} ergs per gram. Therefore any collision capable of supplying the kinetic energy needed for sending an Apollo-size object into an earth-crossing orbit would most likely shatter the collision fragment into small pieces. Furthermore, experience with terrestrial explosion craters shows that only a minuscule proportion of the ejected material reaches a velocity of six kilometers per second; most of the material does not reach a velocity of 200 meters per second. As a result even if an occasional large fragment could survive ejection at six kilometers per second, the number of such fragments would be very small.

It is therefore necessary to seek gentler mechanisms that accomplish the transfer into an earth-crossing orbit. Several are known, and all have the property that gravitational perturbations accomplish almost all the required velocity change. Of these mechanisms the most important is likely to involve various kinds of resonant interactions that are capable of increasing the orbital eccentricity of the small body so that it comes under the gravitational influence of Mars.

In this connection particular importance may be attached to the large group of asteroids near the inner edge of the asteroid belt, between 2.1 and 2.6 A.U. from the sun, that occupy orbits intersecting the plane of Jupiter's orbit at a low angle and are completely stable on the time scale of the age of the solar system. James G. Williams of the Jet Propulsion Laboratory of the California Institute of Technology has shown that the orbits of these asteroids have semimajor axes whose values are such that their precession rate is nearly in resonance with the precession rate of the coupled system of the larger planets, dominated by Jupiter and Saturn. Under the influence of their giant neighbors the asteroids in this particular group have been driven into orbits that, although they are stable, oscillate considerably in eccentricity.

The same stability would not be guaranteed, however, to Apollo-size fragments ejected from such asteroids by an accidental collision. If the fragment were increased in velocity by as little as 200 meters per second and were placed in an orbit whose semimajor axis was only slightly smaller than that of the parent body, it would enter the critical region where resonances with the coupled Jupiter-Saturn system would begin to amplify the fragment's orbital eccentricity, driving it eventually into occasional close approaches to Mars when that planet was near its aphelion.

As a result of successive close approaches gravitational perturbations induced by Mars will "random walk" the semimajor axis of the potential Apollo

object into a position where the object's orbital eccentricity will be amplified by further resonant interactions until it finally crosses the earth's orbit and becomes a true Apollo. The entire process should take less than 100 million years, and doubtless some Apollos are created in this way. The best estimate, however, is that such a process might account for only about one Apollo object per million years. The probabilities must be stretched quite a bit to account for 15 Apollos.

The alternative hypothesis, and the one I favor on the basis of all the present evidence, is that the majority of Apollo objects are former comets. It is known that comets emit large volumes of volatile material, principally water vapor and carbon dioxide. Presumably these substances are originally present in the comet nucleus either in the form of simple ices or bound in more complex compounds. Such volatiles, however, are not the sole constituents of the nucleus. When the comet approaches the sun, much fine-grained dust can be detected flowing out along with the gases. Typically as much dust is released as gas. When the larger fragments of nonvolatile dust are captured by the earth, they produce the cometary meteor showers.

Some meteors that are almost certainly of cometary origin weigh as much as several tons. Such massive bodies can hardly be swept along with the usual stream of volatilized gas, and so they must arise from more catastrophic events, perhaps associated with a split-

	OBJECT	PERIHELION DISTANCE (ASTRONOMICAL UNITS)	APHELION DISTANCE (ASTRONOMICAL UNITS)	ORBITAL INCLINATION (DEGREES)	ESTIMATED DIAMETER (KILOMETERS)	YEAR OF DISCOVERY
1	1566 ICARUS	0.19	1.97	23	1.0	1949
2	1978 SB	0.35	4.11	12	8.0	1978
3	1974 MA	0.42	3.09	38	4.0	1974
4	2101 ADONIS	0.44	3.30	1	0.8	1936
5	2100 RA-SHALOM	0.47	1.20	16	3.0	1978
6	1976 UA	0.47	1.22	6	0.2	1976
7	1864 DAEDALUS	0.56	2.36	22	2.0	1971
8	1865 CERBERUS	0.58	1.58	16	1.0	1971
9	HERMES	0.62	2.66	6	1.0	1937
10	1981 MIDAS	0.62	2.93	40	1.0	1973
11	1862 APOLLO	0.65	2.29	6	2.0	1932
12	2063 BACCHUS	0.70	1.45	9	1.0	1977
13	1685 TORO	0.77	1.96	9	4.0	1948
14	2062 ATEN	0.79	1.14	19	1.0	1976
15	1977 HA	0.79	2.40	23	0.6	1977
16	PL-6743	0.82	2.42	7	0.4	1960
17	1976 WA	0.82	3.03	23	2.0	1976
18	1620 GEOGRAPHOS	0.83	1.66	13	2.0	1951
19	1947 XC	0.83	3.67	1	2.0(?)	1947
20	1959 LM	0.83	1.84	3	6.0(?)	1959
21	1950 DA	0.84	2.53	12	2.0	1950
22	1866 SISYPHUS	0.87	2.92	41	5.0	1972
23	1973 NA	0.88	3.98	68	3.0	1973
24	1978 CA	0.88	1.37	26	2.0	1978
25	1863 ANTINOUS	0.89	3.63	18	2.0	1948
26	1975 YA	0.91	1.69	61	1.0	1975
27	PL-6344	0.94	4.21	5	0.4	1960
28	1978 DA	1.02	3.92	16	2.0	1978
29	1915 QUETZALCOATL	1.05	3.99	21	1.0	1953
30	1917 CUYO	1.06	3.23	24	2.0	1966
31	1580 BETULIA	1.12	3.27	52	6.0	1950

TWENTY-EIGHT BODIES are now recognized as Apollos. An additional three objects, the last three in the table, are not strictly earth crossing at present but will pass within the earth's orbit as a result of perturbations caused by Jupiter. Sixteen of the 31 are designated by number, indicating that their orbital elements are known precisely enough to qualify them for listing in the catalogue of asteroids. Most of the rest are identified by the year of their discovery followed by two letters; the first letter indicates the half month of the year (beginning with *A* for January 1–15) in which the discovery was made and the second the order of discovery in that period. Two objects are designated PL, which stands for Palomar-Leiden Survey. Ra-Shalom was originally designated 1978 RA, signifying that it was the first object (A) to be discovered in the first half of September (R), 1978. The discoverers of an asteroid have the privilege of giving it a name.

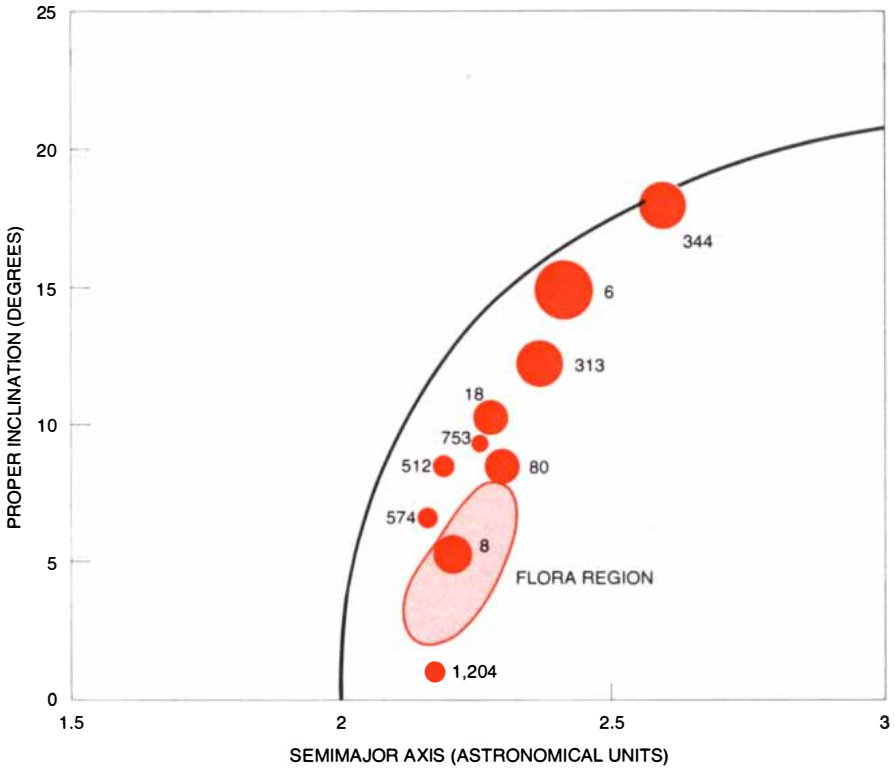
ting of the cometary nucleus, an event that is occasionally observed. In the absence of such a catastrophe the larger objects would simply remain with the comet, making the nucleus richer in nonvolatile material.

Comets that appear for the first time have spent much of their life traveling far beyond the most distant planet. They become luminous for the first time when their orbit, perturbed by a neighboring star, has been deflected to a perihelion close to the sun. In most cases subsequent perturbations induced by the planets will cause the comets to escape from the solar system before their volatile ices are much depleted. A small fraction of such long-period comets are deflected, however, into smaller stable orbits that typically carry them near the sun every seven years or so.

In the course of hundreds or thousands of perihelion passages the comet nucleus is steadily depleted in volatile substances. It has been observed that such short-period comets are markedly less active than comets in large and highly eccentric orbits. In a few cases the short-period comets have appeared to be entirely "asteroidal" in that during some perihelion passages they show no coma or tail. It is therefore no great extrapolation to imagine that a comet can lose all its free volatiles and present itself simply as an asteroid.

The mere depletion of volatiles, however, would not convert a comet into a typical Apollo. Almost all short-period comets have aphelions beyond the orbit of Jupiter, and all known Apollos have aphelions within that orbit. There are nonetheless exceptions. Several short-period comets have aphelions slightly inside Jupiter's orbit. One of them, Comet Encke, has an aphelion far inside, at 4.1 A.U. from the sun.

The activity of Comet Encke has declined noticeably during historical times, and the comet seems well on its way to becoming inactive. If it does become totally inactive, perhaps during the next few centuries, it will be indistinguishable from a low-albedo Apollo object. The most recently discovered Apollo, 1978 SB, is such an object, and it happens that its orbit is quite similar to Comet Encke's. Quite possibly Comet Encke and 1978 SB are genetically related. The existence of meteor streams in similar short-period orbits that give rise to meteorites with physical properties essentially identical with meteorites thought to be derived from Comet Encke argues that the comet's orbital evolution is not a unique event. Physical models capable of explaining such orbital evolution have recently been presented by Whipple and his colleague Zdenek Sekanina. They show that in a comet that is rotating, orbital changes can arise from the dynamical reaction to escaping gas and dust.



POTENTIAL SOURCE OF APOLLO OBJECTS is a group of bodies that are clustered in low-inclination orbits near the innermost part of the main asteroid belt between Mars and Jupiter. Such asteroids are designated here by their catalogue numbers; the adjacent filled circles indicate their relative size. It can be estimated that the group includes some 5,000 bodies (*shaded area*) belonging to the Flora region (named for the largest body in the region, asteroid No. 8). Although all the objects in the group occupy stable orbits, they lie close to the instability "surface" indicated by the curved black line. In 1969 James G. Williams, then a student at the University of California at Los Angeles, discovered that bodies whose inclination and semimajor axis placed them on this surface would be perturbed by Jupiter into highly eccentric earth-crossing orbits. When asteroids in the group shown here collide, as they must from time to time, Apollo-size fragments could be ejected into orbits lying on or close to instability surface.

The evidence provided by one comet—Comet Encke—and several cometary meteor streams may seem to offer only a tenuous basis for the belief that Apollo objects may be of cometary origin. Actually even if all Apollo objects are extinct comets, the current existence of one active comet such as Comet Encke may be more than one has a right to expect. It is perhaps a happy accident that Comet Encke did not fade away a few centuries ago, in which case there would be no compelling observational basis for thinking that comets can turn into Apollos, even if they do. A single Encke-like comet every 65,000 years would supply Apollos at the rate of 15 per million years, the rate needed to maintain the current population in a steady state. Since the active lifetime of a short-period comet is only a few thousand years, most of the time there should be no active comets in Encke-like orbits.

Looking into the future, it is possible that Apollo objects could be turned to practical use. Some of the objects are in orbits so similar to the earth's that they are as accessible as the moon. Moreover, they are small enough so that one could consider the possibility of using them as ready-made and maneuverable space stations. In a scheme originally proposed by Arthur C. Clarke and examined in some detail by Henry H. Kolm of the Massachusetts Institute of Technology the asteroids would be equipped with "mass drivers," engines that would consume the substance of the asteroid itself to propel it into new orbits. Others have suggested that the Apollos could be mined to supply raw materials for the construction of space vehicles.

Before such exotic projects were undertaken it might be well to reflect that the total amount of material in all the Apollos is only about 10^{19} grams, a mass equivalent to no more than the top two or three centimeters of the earth's continental surface. It is likely that the Apollos, together with a few comets, represent the only accessible source of nonterrestrial carbon compounds and water in nearby space, a source that requires about 25 million years to be replenished. Perhaps it would be better if the region of space occupied by the Apollo objects and their close relatives the Amor objects were declared an inviolable "wilderness area."

Cancer Metastasis

Metastasis, and not the primary tumor, is usually the threat in cancer. Apparently only certain very malignant cells possess characteristics that enable them to travel through the body and establish new tumors

by Garth L. Nicolson

Most cancer patients are not killed by their primary tumor. They succumb instead to metastases: multiple, widespread tumor colonies established by malignant cells that detach themselves from the original tumor and travel through the body, often to distant sites. If a primary tumor is detected early enough, it can usually be eliminated by surgery, radiation or chemotherapy or some combination of those treatments. Unfortunately the metastatic colonies are harder to detect and eliminate, and it is often impossible to manage all of them successfully. From a clinical point of view, then, metastasis can well be considered the conclusive event in the natural history of cancer. Moreover, the ability to metastasize is the property that uniquely characterizes a malignant tumor.

A tumor is initiated by the transformation of a normal cell into one that escapes the host's usual controls on growth and differentiation. The transformed cell proliferates to form a tumor, which may be benign or malignant. A tumor is generally considered benign if it remains similar in structure to the tissue from which it is derived, if it grows slowly by simple expansion and remains encapsulated by a layer of connective tissue, and if its cell nuclei divide almost normally, with few abnormal chromosomes. A malignant tumor, on the other hand, is usually atypical in tissue structure, grows rapidly and does not remain encapsulated, displays many abnormal nuclear divisions and chromosomes—and invades the surrounding normal tissue, shedding cells that have the ability to colonize new sites. Whereas distinctions between the benign and the malignant state based on rate of growth and degree of "abnormality" are not absolute, the ability to invade surrounding tissue and to colonize distant sites absolutely defines a malignant tumor, or cancer. Understanding the biology of cancer metastasis is therefore of primary importance both for understanding the nature of cancer and for learning how to prolong or save the lives of cancer patients by destroying the most life-threatening metastatic tumor cells.

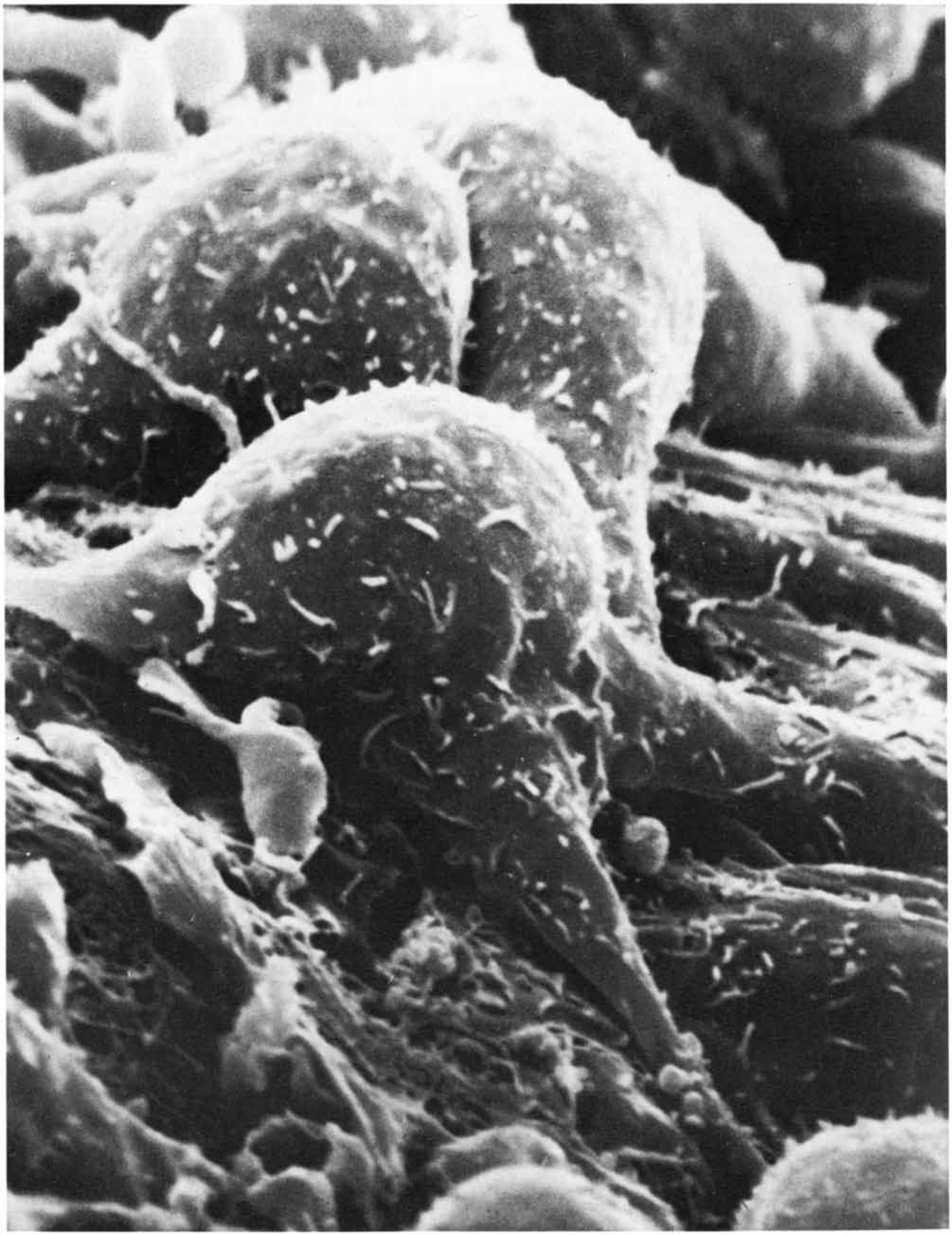
Cancer metastasis comprises a complex series of events involving interactions among malignant cells from the primary tumor and between those cells and the body's normal cells. Irving Zeidman of the University of Pennsylvania School of Medicine has categorized the sequential steps that lead to the colonization of a distant metastatic tumor: (1) extension into surrounding tissues, (2) penetration of body cavities and vessels, (3) release of tumor cells for transport to other sites, (4) reinvasion of tissue at the site of arrest and (5) manipulation of the new environment to promote tumor-cell survival, vascularization and tumor growth.

First the malignant cells extend from the original site, invading the neighboring normal tissues. Some invading cells penetrate into a body cavity (such as the abdominal or chest cavity) or into the blood, lymph or spinal fluid and are released. In the blood or lymphatic vessels single cells or groups of cells (called tumor emboli) can be transported to sites far away from the primary tumor. Cells that survive transport in the lymphatic system are usually arrested in regional lymph nodes; those that survive in the bloodstream are arrested in venules or capillaries. Once arrested, the malignant cells can escape the hostile environment of the lymph or blood by invading adjacent tissue. Finally the new micrometastases become vascularized, and their new network of blood vessels supplies the nutrients that enable the cells to proliferate and form a larger tumor mass. Throughout the metastatic process the malignant cells must resist destruction by the host's various antitumor defense mechanisms.

Several theories have been advanced to account for the first step in metastasis, the invasion of adjacent tissues. George Eaves of the University of Leeds School of Medicine has proposed that invasive behavior can be accounted for largely by mechanical considerations. When he injected a quick-setting dental plastic into tissue samples, the plastic infiltrated into zones of least resistance, a behavior resembling that of many malignant tumors. Yet simple mechanical models do not seem to account for the extremely rapid invasive behavior of some tumors, and experiments with tissue cultures make it clear that highly invasive malignant cells will penetrate a sheet of normal cells even in the absence of any mechanical pressure. Dale R. Coman of the University of Pennsylvania School of Medicine has suggested that it is the motility of malignant cells that enables them to infiltrate surrounding tissue, but there are various normal cells that are highly motile and not invasive.

The presence in tumor tissue of high concentrations of certain degradative enzymes has suggested that the invasive behavior of tumor cells may be facilitated by their enzymatic destruction of the normal connective matrix between cells. A. R. Poole and his collaborators at the Strangeways Research Laboratory compared the protein-digesting enzymes secreted by cultured cells from benign human breast tumors, from malignant ones and from normal breast tissues; Bengt Sylvén has performed similar experiments at the Karolinska Institute in Sweden. Poole found that one proteinase, cathepsin B, accumulated at a much higher rate in the malignant-cell cultures. Other studies, however, have demonstrated little difference in the release of such enzymes from tumor tissues and from normal tissues, so that invasion cannot be explained by enzymatic action alone. It probably depends on various combinations of mechanical, cell-motility and enzymatic mechanisms.

Having extended from the original tumor, malignant cells must become detached from it before they can metastasize. Leonard Weiss and George H. Poste of the Roswell Park Memorial Institute believe some tumor cells or normal cells may lyse, or burst, increasing the extracellular levels of degradative



METASTATIC MELANOMA CELLS are seen invading the wall of a blood vessel in this scanning electron micrograph made by the author. Mouse melanoma cells that typically metastasize by way of the blood were incubated with a piece of vein tissue and then maintained on it in a laboratory culture. The melanoma cells have attached themselves to endothelial cells lining the inner surface of the vein and have inserted their processes into fissures that open up as endothelial cells under the tumor cells retract. Endothelial attachment and invasion are two stages in the process whereby a blood-borne tumor cell establishes a new tumor colony. Magnification is about 7,000 diameters.

enzymes that release clumps of malignant cells. There is also some evidence that the normal adhesive forces tending to bind cells together are reduced in the case of malignant cells. Coman found, for example, that shaking clumps of malignant breast-cancer cells separated more cells from the clump than similar agitation of clumps of normal breast-tissue cells.

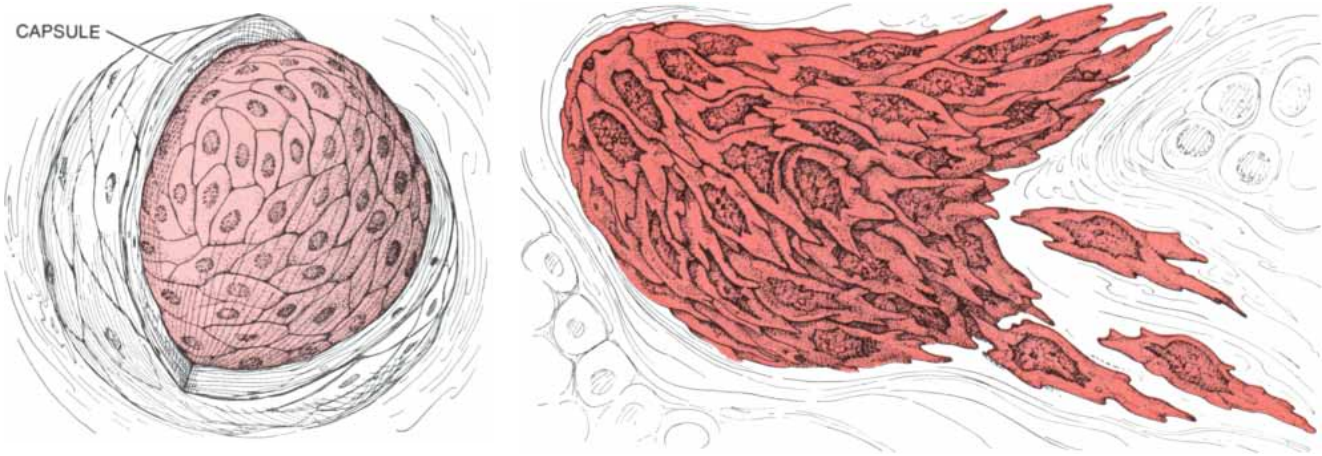
The release of malignant cells into the blood or lymph can lead to widespread tumor metastases. Zeidman and Bernard Fisher and Edwin R. Fisher of the University of Pittsburgh School of Medicine have shown that since the lymphatic and blood systems are connected, tumor cells can pass from the lymphatic vessels to the blood vessels and vice versa. The Fishers labeled carcinoma (epithelial tumor) cells with radioactive chromium and injected them into the peripheral veins of rats. Within an hour after injection live tumor cells could be recovered from the lymph, and injection of the lymph into other animals produced carcinomas. These results showed that malignant cells in the blood are not always trapped in the first capillary bed they encounter but are able to escape from those capillaries and pass from the blood into the lymphatic system, presumably by way of the lymph nodes.

A tumor probably releases malignant cells into the lymph or blood continuously as it grows. Pietro M. Gullino of the National Cancer Institute has monitored the blood leaving transplanted mammary tumors in experimental animals and found that after a tumor has grown to a few grams in weight it releases several million malignant cells a day into the circulation. The presence of malignant cells in the blood (or lymph) does not always lead to the formation of distant metastases, however. A. J. Salsbury of Brompton Hospital in London was one of the first to emphasize that the presence in the blood of tumor cells released from a primary solid tumor does not necessarily mean that a patient will subsequently succumb to multiple metastases. The fact is that the overwhelming majority of the malignant cells released into the blood die very quickly. Only a few seem to be able to survive in the turbulent and otherwise hostile environment of the bloodstream.

Isaiah J. Fidler and Zeidman investigated the fate of blood-borne melanoma cells in mice. (Melanoma is a malignant tumor that develops in the skin of animals and man; it is often highly pigmented by the black substance melanin.) Fidler and Zeidman injected into a peripheral vein mouse melanoma cells whose DNA had been labeled with radioactive iodine. They found that within a few minutes most of the injected cells were arrested in the lung, the site of the first capillary system the cells encountered. Most of the tumor cells died there; a few live cells continued to circulate. Within a day, however, only 1 percent of the injected cells were still alive. After two weeks, when black melanoma metastases could actually be seen in the lungs, only a tenth of 1 percent of the radioactive cells originally injected were still alive. The results indicated that blood-borne metastasis must be a highly selective process and that only a very small percentage of the malignant cells that are able to enter the bloodstream and circulate in it survive to form distant metastases.

While the malignant cells circulate in the blood they undergo a variety of important cellular interactions, as has been demonstrated in my laboratory at the University of California at Irvine and in the laboratories of Fidler (who is now at the National Cancer Institute's Frederick Cancer Research Center), Gabriel J. Gasic and Tatiana Gasic of the University of Pennsylvania School of Medicine and Lance A. Liotta of the National Cancer Institute. These interactions are likely to affect the survival of the cells and their arrest in the capillaries. The circulating malignant cells can, for example, adhere to other tumor cells to form clumps. They can adhere to such normal blood cells as lymphocytes and platelets. And when they enter a capillary bed, they can adhere to the endothelial cells that line the vessels. The larger the clump is, the more likely it is to become arrested in the capillaries and ultimately form a new tumor colony. Having been arrested in the capillaries, some blood-borne tumor cells stimulate the production of fibrin, the fibrous protein that is ordinarily involved in the formation of blood clots. B. A. Warren of the University of Western Ontario has found that the resulting mesh of fibrin tends to protect the malignant cells and enable them to proliferate.

Once the malignant cells are arrested in small venules or capillaries, they must get through the blood-vessel wall and into the tissue outside it in order to survive. The escape of tumor cells from the circulation has been studied by Randall Kramer in my laboratory. It appears that malignant cells induce a dramatic change in the adhesion of endothelial cells (which line the inner surface of blood vessels) to their underlying extracellular matrix: the endothelial cells retract, leaving cell-free spaces through which the tumor cells can escape from the vascular system. Once the tumor cells have invaded the extravascular tissue a new microenvironment must be established. The growth of the new tumor colony is limited until a new blood supply becomes available to deliver nutrients and carry away cellular waste products. The vascularization of the new tumor is stimulated by "tumor angiogenesis factors" that are released by the malignant cells [see "The Vascularization of Tumors," by Judah



TUMOR is a mass of tissue arising from the proliferation of a cell that has been transformed so that it escapes from normal controls on growth and differentiation. In general a benign tumor (*left*) appears almost normal in tissue structure, grows slowly and remains encapsulated by connective tissue; it does not invade surrounding tissue and its cells do not detach and metastasize. A malignant tumor (*right*) is atypical in structure, usually grows faster and is not encapsulated. It invades the surrounding tissue; some cells detach and metastasize.

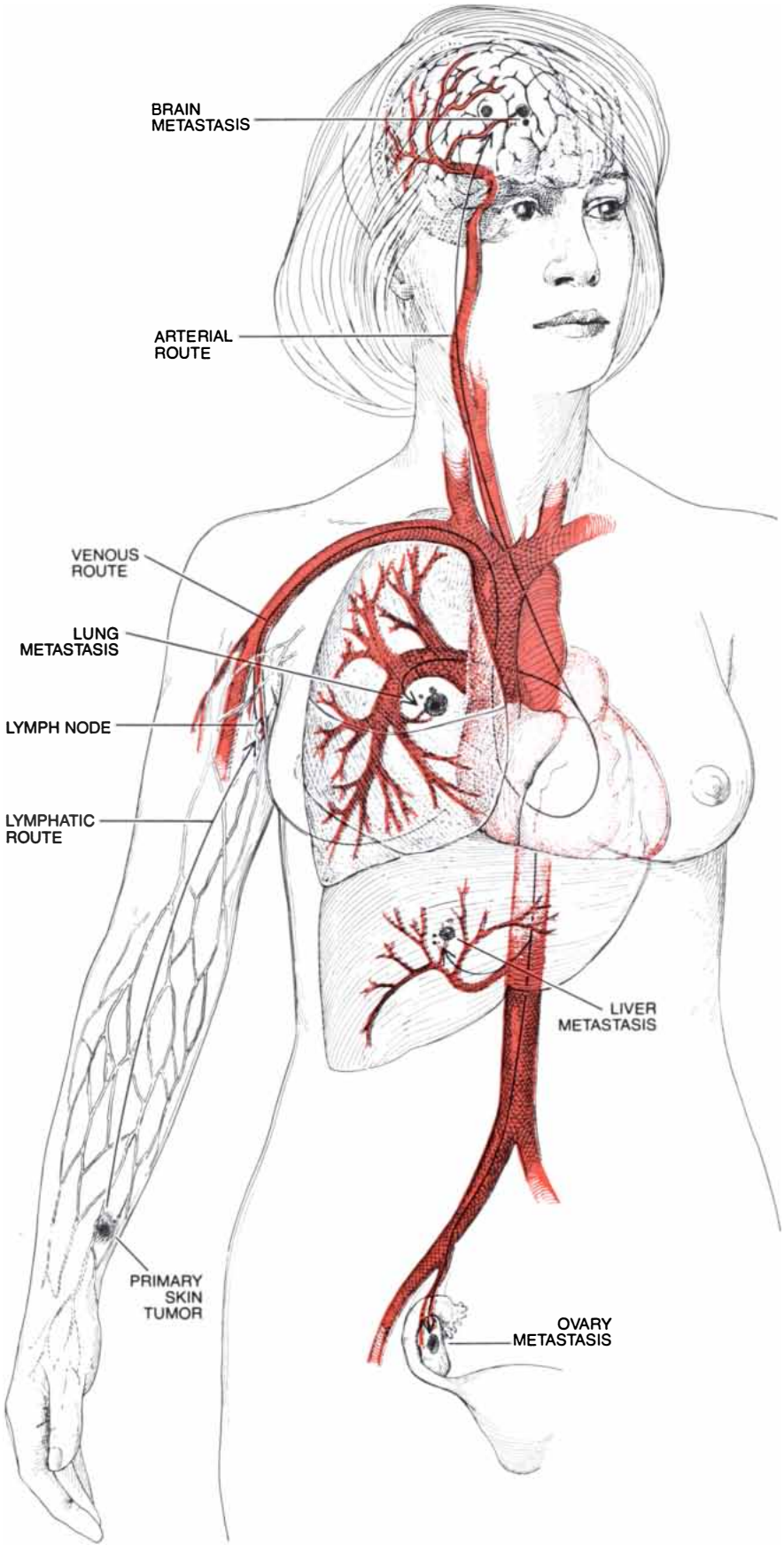
Folkman; SCIENTIFIC AMERICAN, May, 1976]. These factors seem to function by specifically stimulating endothelial-cell proliferation through cell division and the movement of the endothelial cells to form new capillaries.

The process of metastasis, then, is dependent on a number of successful interactions between malignant cells and the normal cells of the host. Probably few of the malignant cells in a primary tumor are capable of accomplishing these interactions and thus completing the entire sequence of steps in metastasis. Only a small fraction of the cells that extend from a primary tumor invade a body cavity or enter the lymphatic or blood system and become detached; of those cells only a few survive transport; fewer still survive to complete the metastatic process, at every step of which they are exposed to a number of the body's defense mechanisms.

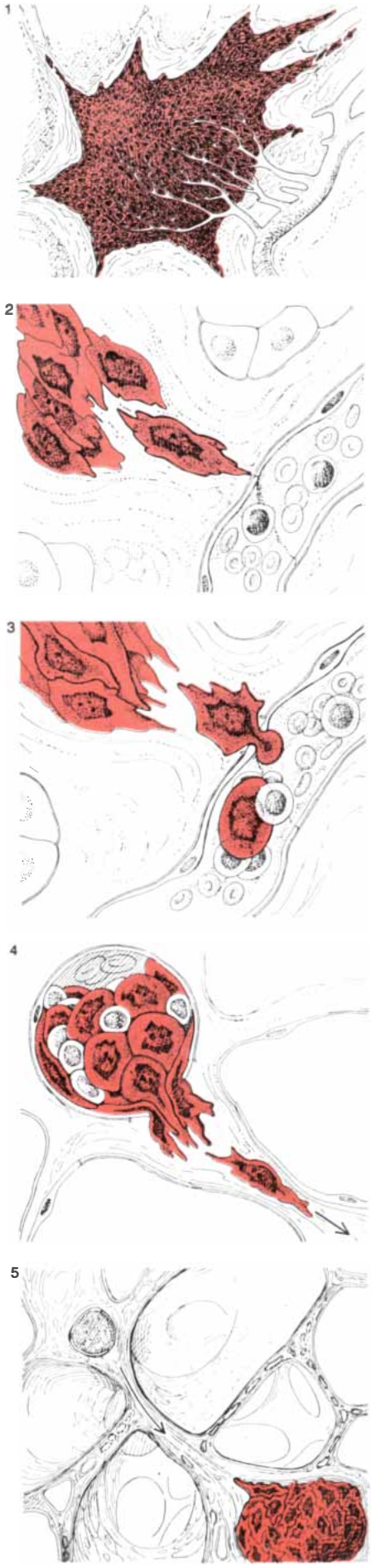
The fact that only a small fraction of a tumor's cells survive to establish new metastatic colonies suggests that the ultimately successful cells may be very special ones. Could it be that they represent a subpopulation of tumor cells endowed with the particular characteristics making for successful metastasis? Might the steady pressure of the host's defenses be responsible for the natural selection of such a subpopulation of particularly malignant variant cells with highly metastatic properties?

This hypothesis can be tested in animal models of metastatic cancer, one of the most useful of which is mouse melanoma. Metastatic melanoma is one of the cancers least susceptible to treatment in man, largely because of the frequency of lymphatic and blood-borne metastasis, and nonhuman melanomas have the same characteristic. Murine *B*16 melanoma, which arose in a strain of black mice maintained at the Jackson Laboratory, can be transplanted from one mouse to another and can also be grown in tissue culture. When *B*16 melanoma cells are implanted in the skin of mice, the melanoma ordinarily metastasizes at a low to moderate rate. In order to achieve a more precise assay of metastatic potential, known quantities of tumor cells are usually injected directly into the circulatory system by way of the tail vein. The cells circulate, arrest, invade, survive and multiply to form a few pigmented tumor colonies in the lungs and fewer still in the ovaries, body cavities, liver and brain.

Fidler, while at the University of Pennsylvania School of Dental Medicine, performed an experiment in which he manipulated *B*16 melanoma cells to select subpopulations with enhanced ability to colonize the lungs. He began by surgically removing the melanoma colonies that formed in the lungs of mice into which *B*16 cells had been injected intravenously. He adapted the tu-



METASTATIC SPREAD of a malignant melanoma, a pigmented skin cancer, can begin when malignant cells from a tumor on, say, the forearm invade a lymphatic vessel and are carried in the lymph to a lymph node. Momentarily arrested there, the cells can enter a blood vessel and then can be carried in the blood through the heart to other parts of the body. Malignant melanoma cells are most likely to metastasize to the brain, the lungs, the liver or the ovaries.



STAGES of blood-borne metastasis are diagrammed. The primary tumor enlarges as the malignant cells proliferate, but it grows rapidly only after vascularization provides an adequate blood supply (*1*). The tumor invades surrounding tissue (*2*). A few cells, singly or in groups, penetrate the basement membrane of a capillary, squeeze between the endothelial cells and enter the bloodstream, where they may interact with blood cells such as lymphocytes (*3*). The tumor cells are carried through the heart to the lung, where they are arrested in a capillary; tumor cells penetrate the capillary wall (*4*), enter the alveolar tissue and proliferate to establish a new tumor colony (*5*).

mor cells to growth in a tissue-culture environment and then harvested the cultured cells, designated as variant melanoma cell line *B*16-*F*1, from tissue culture and injected them intravenously into another group of mice. The new experimental lung-tumor colonies that resulted from this second selection were cultured in turn to derive a new line, *B*16-*F*2. The process of injection, lung colonization, tissue culture and harvesting for reinjection into mice was continued, and with each succeeding cycle the efficiency with which the variant lines were arrested in the lungs and then survived, invaded and grew into lung-tumor colonies increased.

After 10 cycles Fidler established a variant melanoma line, *B*16-*F*10, that formed significantly more lung-tumor colonies than either the parental *B*16 line or the once selected *B*16-*F*1 line when a given number of malignant cells were injected into mice. This family of variant tumor-cell lines that have differing metastatic potentials has been an effective tool in a number of biological and biochemical studies of the sequence of tumor metastasis. Its success as a model for metastasis is probably due to the fact that the original *B*16 tumor line did not consist of a uniform cell population; it was probably heterogeneous with respect to the particular tumor-cell characteristics necessary for metastasis.

Clinical observations indicate that there is a tendency for particular types of primary tumors to metastasize preferentially to particular organs. For example, breast carcinomas tend to spread to the brain or the lung; lung tumors often metastasize to the brain and adrenal glands, prostate carcinomas to bone. Selective patterns of metastasis have also been noted for nonhuman tumors. If the pattern of metastatic distribution for a given cell type is not random, what is the reason?

A hypothesis to explain nonrandom metastasis was put forward early in the century by the cancer investigators James Ewing and Stephen Paget. It held that metastatic distribution was determined by anatomical and vascular conditions and the microenvironment for tumor survival and growth at the site of the metastasis. The implication was that tumor cells thrive in a particular "soil" if they are able to adapt to it, and so the hypothesis has been called the soil hypothesis. More recently it has become clear that the properties of the malignant cells themselves largely influence their patterns of metastasis. As Everett D. Sugarbaker and Alfred S. Ketchem of the University of Miami School of Medicine have stressed, the distribution and survival of individual malignant cells probably depend on a number of factors, including the unique properties of the cells themselves, the environment for survival and growth in specific organs and the antitumor immune status of the host.

In collaboration with Fidler we have studied the specificity with which subpopulations of *B*16 mouse melanoma cells home to the lung and colonize it. Are the tumor cells simply trapped in the lung because it is the first capillary bed they encounter? Or do they actually seek lung tissue? We injected the relatively nonspecific *B*16-*F*1 cells and the highly specific *B*16-*F*10 cells into both the tail vein (from where they would first encounter the lung capillaries) and the left ventricle of the heart, which is downstream from the lungs in the circulatory system, so that the cells would encounter other capillary beds before reaching the lungs. The initial distributions of the radioactively labeled tumor cells were dramatically different, suggesting that they had indeed been initially arrested in capillaries downstream from their different sites of injection. Yet within one day after injection the distributions were similar, and the number of cells of each line that survived was the same, regardless of the point of entry into the circulatory system. By the end of two weeks almost identical numbers of lung tumors were found for both routes, with the *F*10 cells, as expected, forming significantly more tumors than the *F*1 cells. The results indicated that tumor cells destined to form metastases in a particular organ can even detach from an initial arresting point and recirculate until they find the proper environment for arrest, survival and growth.

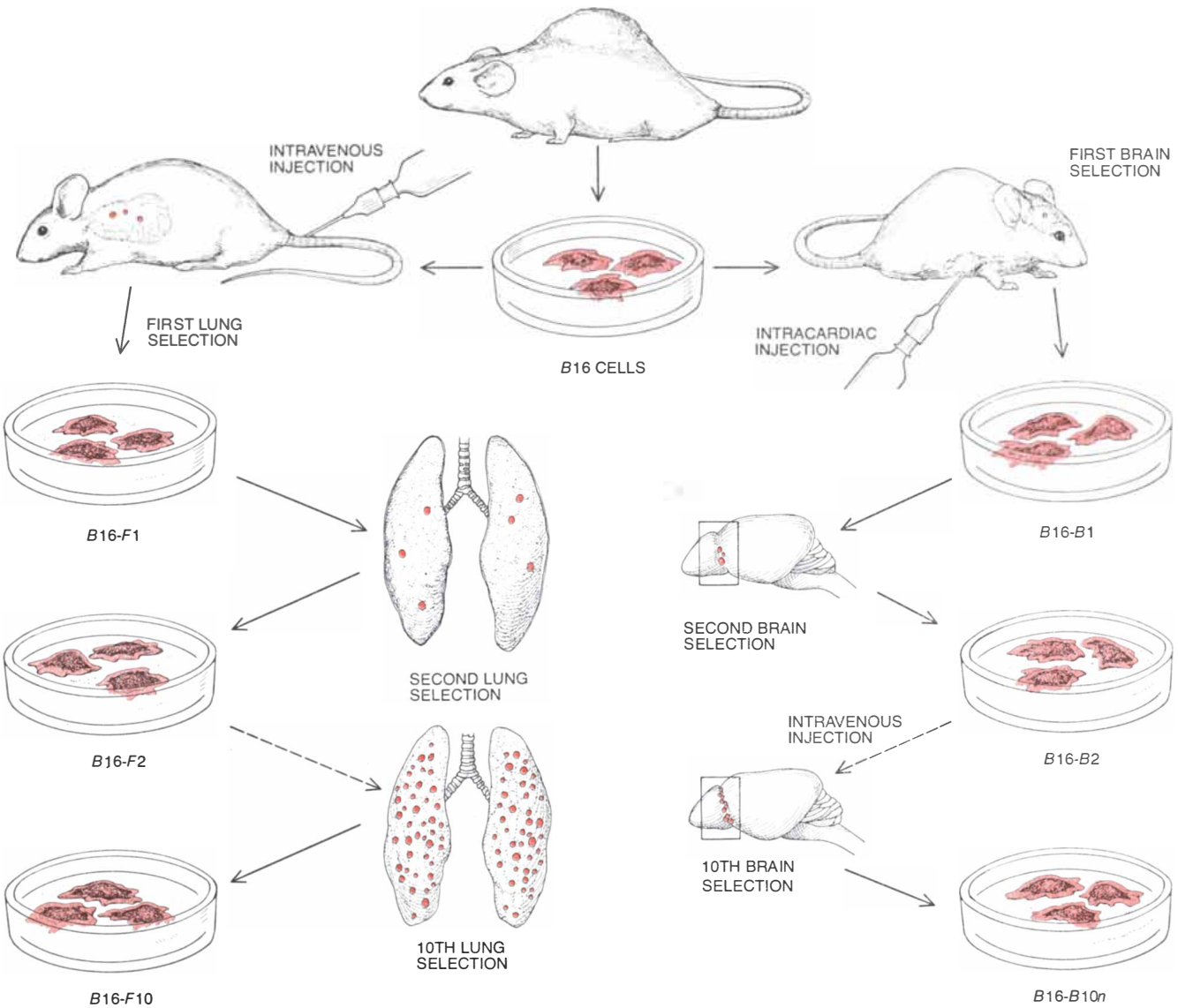
Selection for organ specificity of colonization apparently means not only that cells become more effective colonizers of a particular organ but also that they become less likely to colonize other organs. In this experiment and many others we found that whereas *B*16-*F*1 cells formed some tumors outside the mouse lung (mainly in the lymph nodes, the ovaries, the chest cavity, the liver, the intestines, the brain and the adrenal glands), the highly selected *F*10 cells did not. They colonized only the lung.

In my laboratory Kenneth W. Brunson took the selection process a step further: he began with the *B*16-*F*1 line, which colonizes the lung predominantly

but not exclusively, and from it selected subpopulations that preferentially colonize quite different sites. In one experiment he excised brain tumors, which are rarely formed by the *B*16-*F*1 cells, adapted the cells to tissue culture and obtained a line he called *B*16-*B*1. Then he went through the procedure of injection, tumor excision and culturing that Fidler had developed to obtain the lung-specific *F*10 line, but at each step he excised brain tumors. After 10 selections Brunson had developed a line, *B*16-*B*10*n*, that metastasizes to a particular area of the mouse brain, the rhinal fissure, between the cerebral cortex and the olfactory bulb. By a similar process another tumor-cell line (*B*16-*O*10) was selected that preferentially colonizes the ovaries, a frequent site of human melanoma metastasis.

Is the preferential colonization by *B*-16 lines of such organs as the brain a result of the random adaptation of melanoma cells to the particular environment (as the soil hypothesis would have it) or of unique properties originally possessed by certain tumor cells? Brunson did parallel experiments in which *B*16 cells and successively selected tumor-cell lines were implanted directly in the brain rather than being injected into the bloodstream, thereby eliminating the phases of blood-borne circulation and arrest. After 10 selections for adaptation and growth in the brain, the *B*16 melanoma cells retrieved from brain tumors were no more effective in metastasis to the brain (when they were injected into the skin or the blood) than the initial *F*1 line. In other words, it is not mere adaptability to growth within the brain that is selected for in the *B*1-to-*B*10*n* sequence; it is the ability to home to the brain, to become arrested there and to colonize that site.

Success in identifying the specific properties of tumor-cell lines that enable them to metastasize to particular sites would be a large first step toward developing a way to block metastasis in cancer patients. Although we are far from knowing precisely what cellular properties contribute to successful me-

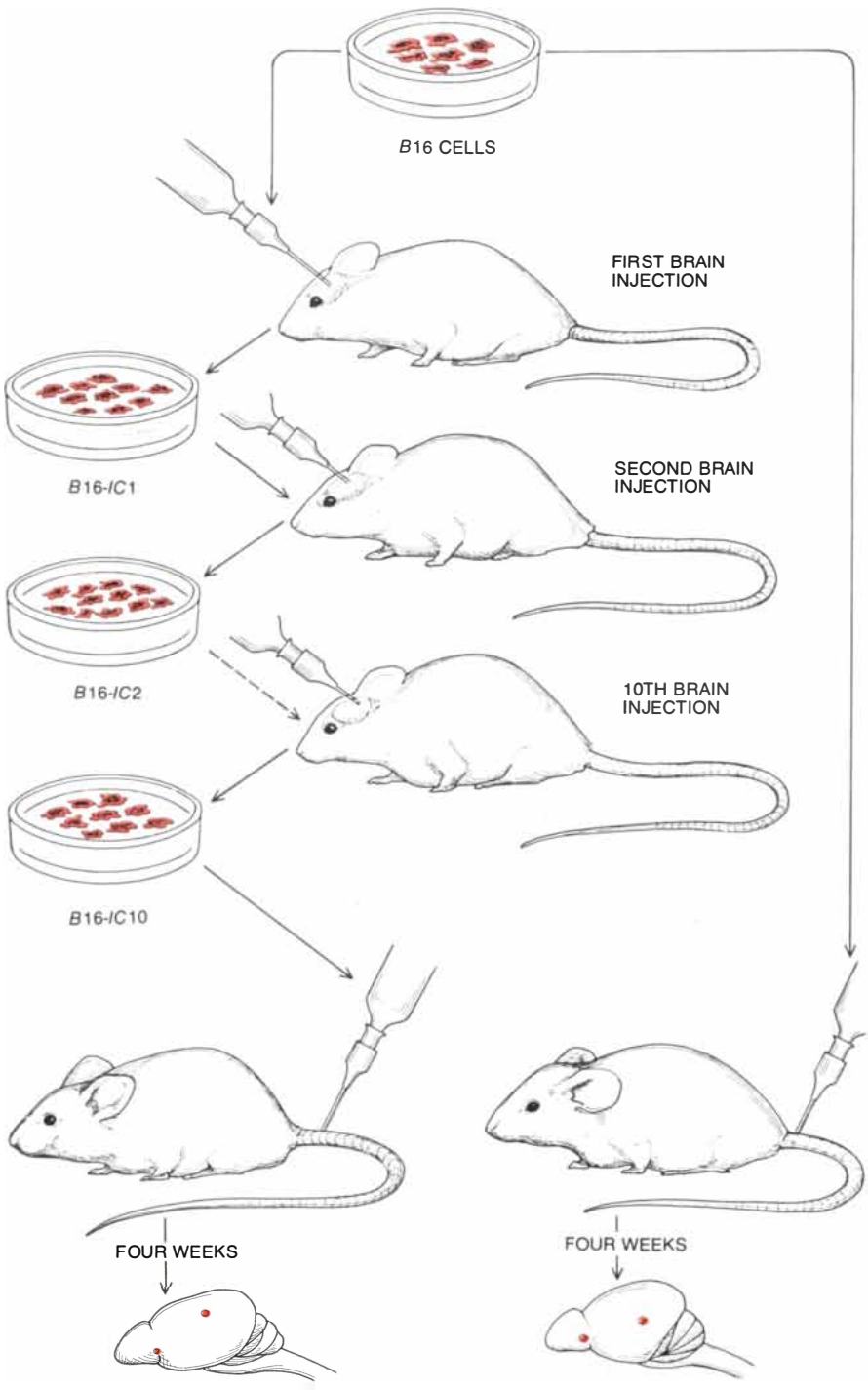


SELECTION OF MELANOMA-CELL LINES with enhanced ability to metastasize to the lungs (*left*) was accomplished by Isaiah J. Fidler, then at the University of Pennsylvania School of Dental Medicine. Kenneth W. Brunson of the author's laboratory selected lines that metastasize to a particular part of the brain (*right*). For the lung selection Fidler took cultured *B*16 melanoma cells from a mouse tumor, injected them into the tail vein of mice and harvested cells from the resulting lung-tumor colonies. These *B*16-*F*1 cells were cultured in turn and again injected intravenously. The highly selected *B*16-*F*10 cells produced by 10 such cycles form significantly more experimental lung metastases than the unselected *B*16 line, and they colonize only the lungs. Brunson injected *B*16 cells into the left ventricle of the heart and harvested the rare brain tumors that were formed. The tumor cells, designated *B*16-*B*1, were cultured and again injected into the heart; this time tumors were collected only from the rhinal fissure near the front of the brain (*rectangle*). The process was repeated with the cells injected intravenously. After 10 selections Brunson obtained the *B*16-*B*10*n* line, highly metastatic only to brain's rhinal fissure.

tastasis, it does appear that the surface of the tumor cell is directly involved in such steps as the arrest of the cell at a particular site.

George Poste and I performed an experiment to demonstrate the cell-surface effect after we found that *B*16-*F*10 cells in tissue culture shed bits of cell membrane in the form of small closed vesicles. We harvested the vesicles and then purified them by centrifugation in tubes containing a gradient of a dense sugar solution. The purified vesicles were added to a culture of *B*16-*F*1 cells along with the polymer polyethylene glycol. In the presence of polyethylene glycol the *F*10 vesicles fused with and became part of the *F*1 cell membrane. When the *F*1 cells enriched with *F*10 membrane vesicles were injected into mice, they metastasized more like the highly metastatic *B*16-*F*10 line than like ordinary *B*16-*F*1 cells, indicating that the cell membrane is indeed involved in metastasis.

In what way, then, do the cell membranes differ in *B*16 lines that colonize different sites? We compared the protein components on the surfaces of various cell lines by incorporating radioactive molecules into the proteins, dissolving the cell membranes in detergent solutions and subjecting the soluble membrane proteins to electrophoresis, thus separating them according to their size and electric charge. Each of the three melanoma lines—those that colonize the lung, the brain and the ovaries—shows a different pattern of specific surface proteins, and in general the amount (or the accessibility) of those specific surface proteins increases with the cell's metastatic potential.

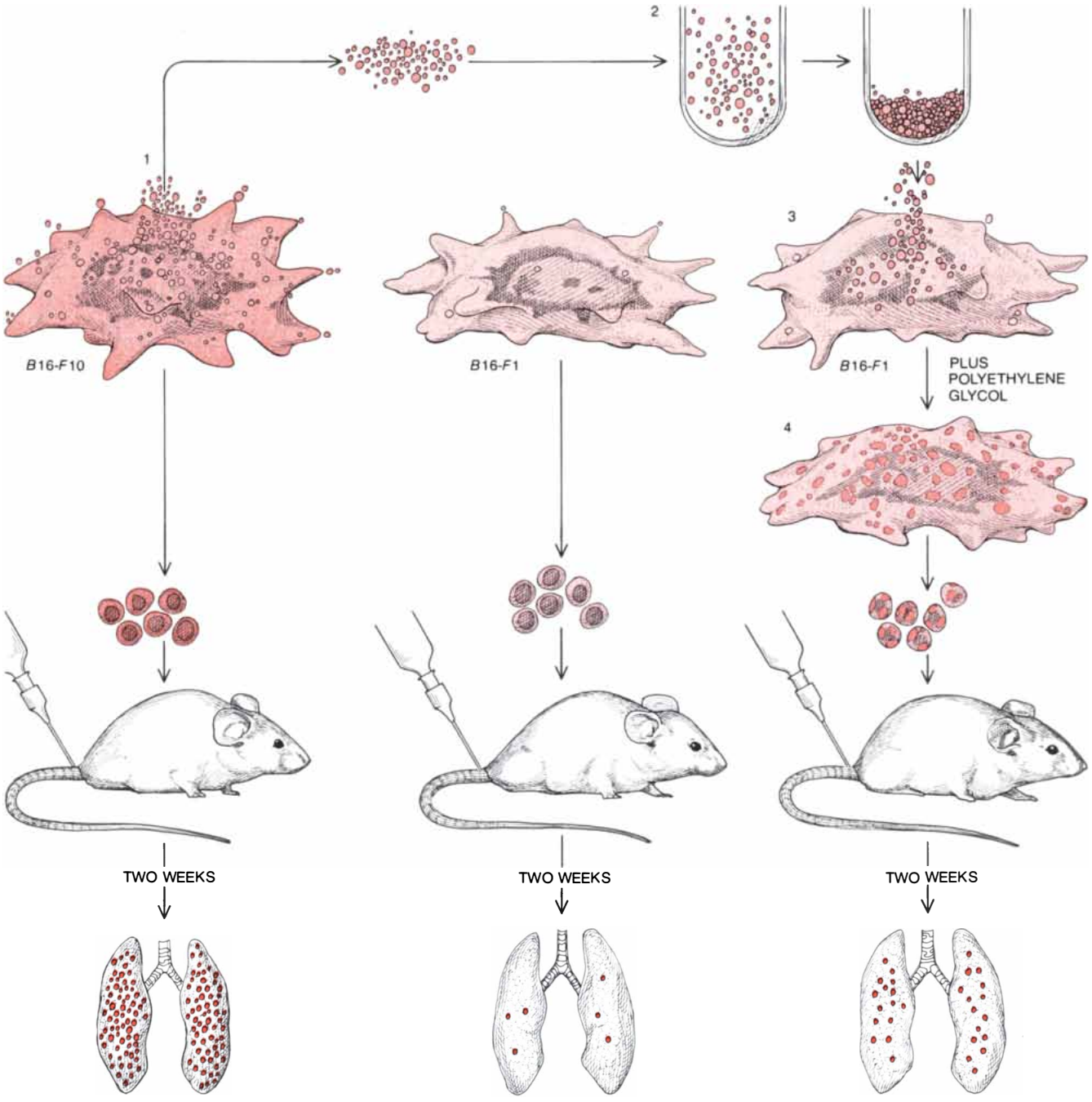


EXPERIMENTAL BRAIN METASTASES

PREFERENTIAL COLONIZATION OF THE BRAIN is not the result of adaptation to the brain environment. This was indicated by an experiment in which *B*16 cells were injected not into the bloodstream but directly into the brain. The melanomas that formed in the brain were excised and their cells were cultured and again injected into the brain. After 10 such cycles the harvested cells, injected into mice, formed no more brain metastases than the original *B*16 cells did when they were injected intravenously. Apparently cells that are highly metastatic to the brain are a subpopulation of the *B*16 line with properties that allow them to arrest in the brain.

Most investigators believe tumors develop from a single transformed cell. Subpopulations possessing different properties can therefore arise only if various cells of a tumor evolve independently even though they are in the same environment. A long period of tumor growth in human patients (or the repeated propagation of tumors by serial transplantation in experimental animals) often leads to progressive changes in the tumor's structure, antigenic properties, metabolic characteristics, growth properties and metastatic potential. Several investigators have called attention to this phenomenon of tumor progression, including Peter C. Nowell of the University of Pennsylvania School of Medicine, Leslie Foulds of the Chester Beatty Research Institute in London, George and Eva Klein of the Karolinska Institute, Emmanuel Farber of the University of Toronto and H. J.-P. Ryser of the Boston University School of Medicine. As a population of tumor cells slowly evolves, either the tumor as a whole changes or subpopulations having divergent properties apparently arise and their frequency in the population increases under selective pressure from the host's defenses, ultimately leading to the emergence of cells with enhanced malignancy. Do subpopulations having different metastatic potentials exist as such in the primary tumor, or does the entire population simply adapt to a given environment?

One way to distinguish between these two possibilities follows the classic experiment of 1943 whereby Salvador Luria and Max Delbrück were first able to distinguish between selection and adaptation as the origin of mutant lines of



CELL MEMBRANE was shown to play a role in determining the metastatic ability of *B*16 cells. Bits of membrane that had been spontaneously released as vesicles (*1*) from the outer membrane of highly metastatic *B*16-*F*10 cells (*dark color*) were harvested from the culture medium, concentrated (*2*) and added to a culture of much less metastatic *B*16-*F*1 cells (*3*). In the presence of polyethylene glycol the *F*10 vesicles fused into the *F*1 cell membranes (*4*). The three cell lines were injected into mice. The *F*1 cells whose cell membrane contained bits of *F*10 cell membrane produced significantly more lung tumors than untreated *F*1 cells, although not as many as the *F*10 cells themselves.

bacteria. Fidler and Margaret L. Kripke repeated that experiment with the unselected *B*16 melanoma line and with a fibrosarcoma (connective-tissue tumor) line from a tumor induced in a mouse by ultraviolet irradiation; we performed the experiment at Irvine with a mouse-sarcoma line. In each case unselected tumor-cell suspensions were divided into two parts. One part was injected intravenously into a large group of mice. The other part was divided into separate cultures from each of which a single cell was isolated and grown to produce a large number of different clones, or subpopulations of cells with an identical genetic constitution. These clones were also injected intravenously into groups of mice.

If the number of tumor colonies in the lungs of mice receiving the various cloned subpopulations turned out to be similar to the number of colonies in mice receiving the uncloned line, it would indicate that the original tumor line was homogeneous; if different cloned subpopulations gave rise to widely different numbers of colonies compared to the uncloned line, it would suggest that the original unselected line was heterogeneous, with cells of low and of high metastatic potential preexisting in it. The latter result was obtained in all three experiments. The implication is that tumor-cell populations are highly heterogeneous with regard to properties governing malignancy, and that a few highly metastatic subpopulations exist in a tumor before the process

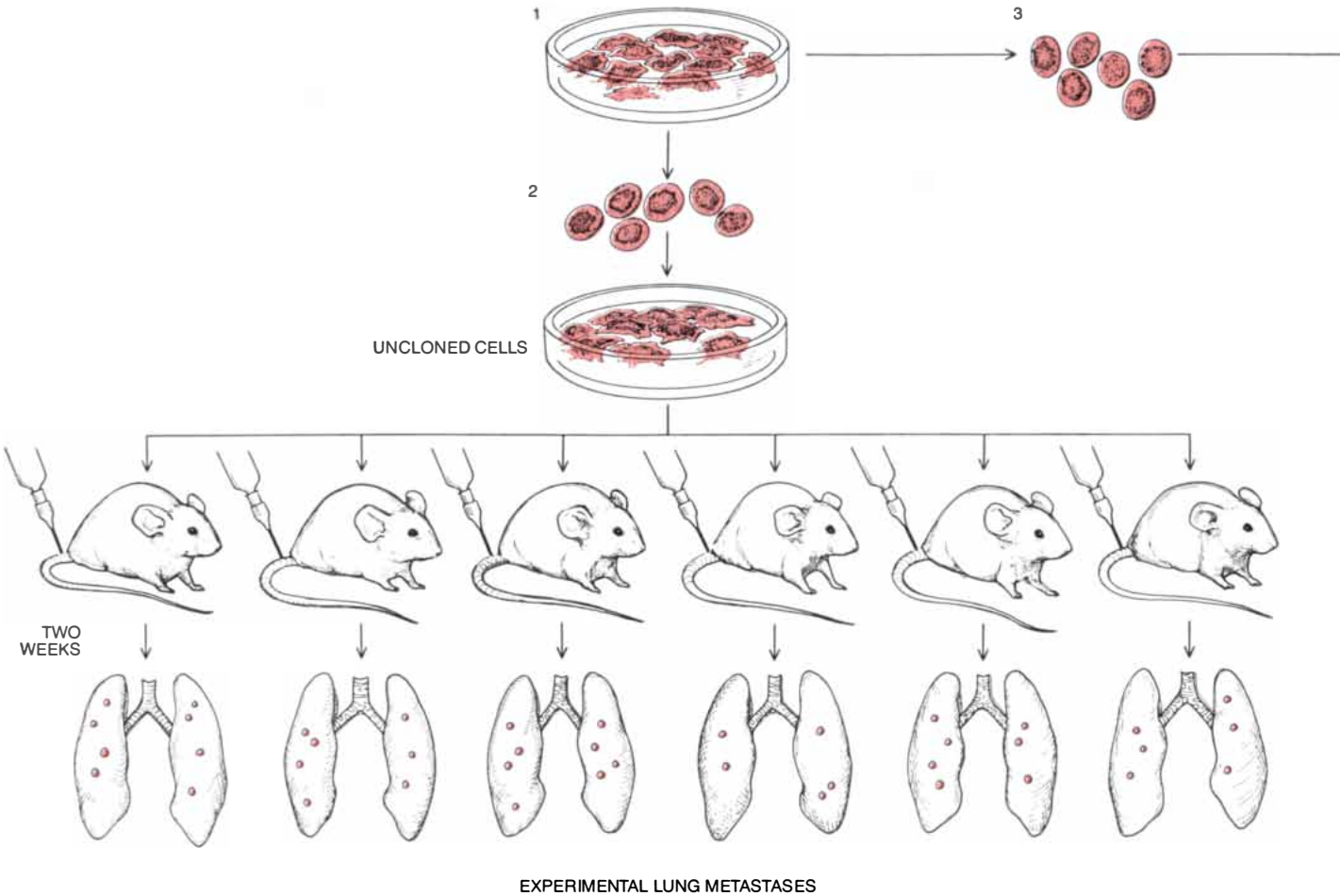
of tumor progression and selection of cell populations with enhanced metastatic properties takes place.

When a normal cell is transformed into a tumor cell, the cell surface displays new (or at least modified) antigens, molecular structures the immune system recognizes as being foreign [see "Cancer Immunology," by Lloyd J. Old; SCIENTIFIC AMERICAN, May, 1977]. The appearance of these tumor-associated antigens can stimulate in the host an immune response that leads to the interception and destruction of the tumor cells. F. M. Burnet of the Walter and Eliza Hall Institute of Medical Research in Melbourne has proposed that cells are transformed periodically but are prevented from forming tumors by a generalized system of "immune surveillance," and that the few tumors that do appear originate from the occasional transformed cells that manage to evade the surveillance system. The efficacy of the immune system in protecting against the appearance and progression of tumors has been questioned by other workers. Richmond T. Prehn of the Jackson Laboratory and Osias Stutman of the Memorial Sloan-Kettering Cancer Center are among those who doubt the generality of immune surveillance as an effective antitumor defense.

Certainly immune responses do affect the progression and metastasis of some human cancers and animal tumors. Several investigators, including Robert W. Baldwin of the University of Nottingham, Luka Milas of the Rudjer Bošković Institute in Yugoslavia, Vilas V. Likhite of the Harvard Medical School and John E. Castro of the Royal Postgraduate Medical School in London, have shown that metastasis of certain tumors can be inhibited when the immune system is stimulated by the introduction of such microorganisms as bacillus Calmette-Guérin (BCG) and *Corynebacterium parvum*. The mechanism of this stimulation is not well understood; it may be effected by a nonspecific incitement of the immune system as well as by a specific response to antigens on the microorganism's surface that are similar to antigens on the tumor cells.

The control of metastasis is not, however, always correlated with the strength of the immune system. Experiments have often failed to show that bloodborne metastasis is any more prevalent in animals whose immune system has been experimentally depressed than it is in animals with an intact immune system. Indeed, the augmentation of immune defenses in some patients and in some experimental animals results in the enhancement of metastasis. When the Fishers depressed the antitumor immunity of mice by injecting antibodies that destroyed the response of lymphocytes (which are an important cellular component of the immune system) to tumor cells, they found that fewer—rather than more—metastases formed. On the other hand, Michael Feldman of the Weizmann Institute of Science in Israel noted that intravenous injection of large numbers of lymphocytes that had been activated to kill carcinoma cells could inhibit the metastasis of that carcinoma in experimental animals.

Such seemingly contradictory results prompted Prehn to suggest that two competing immune systems might be involved, one of them responsible for killing tumors and the other one responsible for stimulating tumors. Prehn found that introducing a small number of antitumor lymphocytes into tumor-bearing mice stimulated tumor growth, whereas the injection of a large number of lymphocytes inhibited tumor growth. Simi-



METASTATIC HETEROGENEITY of a tumor-cell population was demonstrated by experiments conducted in Fidler's and the author's laboratories. In each case an unselected tumor-cell population (*1*) was divided into two parts. One part (*2*) was injected into many groups of mice intravenously. The other part (*3*) was used to produce a large number of clones, each one descended from a single cell; each

lar experiments have been done with the *B*16 melanoma system. The injection into mice of a small number of lymphocytes mixed with *B*16 cells brought about a significant increase in lung metastases, but when large numbers of lymphocytes were mixed with the tumor cells, there were fewer metastases.

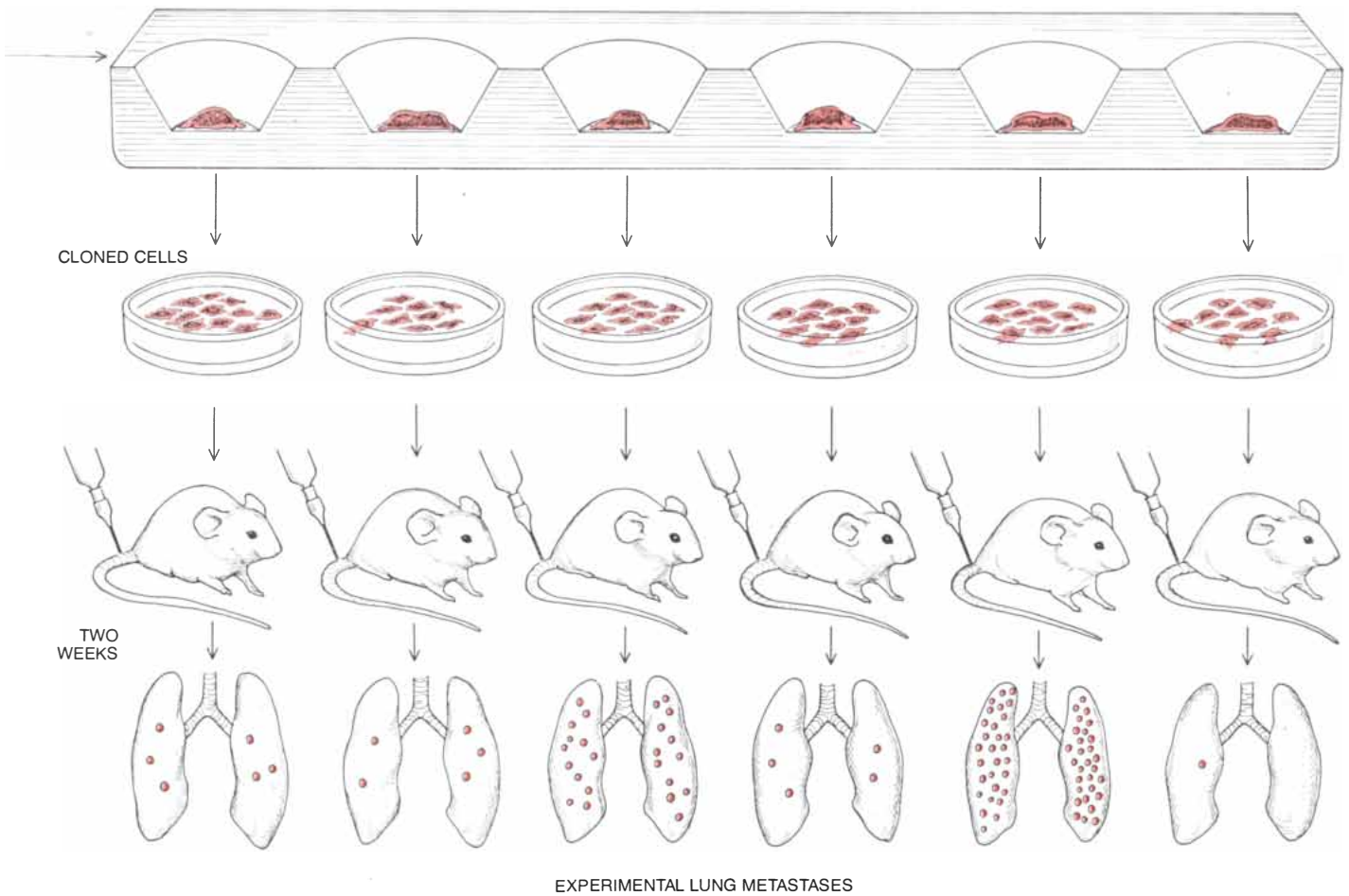
How might lymphocytes in small amounts stimulate the arrest and survival of melanoma cells in the bloodstream? The answer is far from clear, but one pertinent observation is that when lymphocytes and *B*16 melanoma cells are mixed in a suspension, they aggregate into small clumps. As I have mentioned, the formation of clumps enhances the arrest and survival of tumor cells. This could be simply because the clumps become lodged in the lung capillaries more readily or because a group of substances called lymphokines, which can be released by lymphocytes after they interact with tumor cells, have a favorable effect of some kind on the survival of malignant cells.

Why, then, should lymphocytes in large amounts inhibit metastasis? A large amount of lymphocytes is more effective than a small amount in aggregating *B*16 cells, and that would presumably facilitate the arrest and implantation of the tumor cells. On the other hand, large numbers of lymphocytes inhibit metastasis by killing malignant cells. It could be that the antimetastatic effect of large numbers of lymphocytes is actually due to the presence among them of a small subpopulation of "killer" lymphocytes that have become activated specifically to attack the tumor cells. Or perhaps the lymphocytes in various experiments were somehow contaminated with a few functionally different immune cells that were responsible for killing the tumor cells.

In an effort to learn whether or not lymphocytes were directly responsible for the inhibition of metastasis Fidler cultured *B*16-*F*10 melanoma cells with activated lymphocytes. Most of the tumor cells were killed, but a few survived; these were grown and resubjected to the activated lymphocytes, and again the survivors were isolated, grown and plated with activated lymphocytes. After six such selections for resistance to lymphocytes the surviving *B*16-*F*10 cells were completely resistant to killing by lymphocytes. These resistant cells were injected intravenously into mice. They formed fewer metastases than the untreated *F*10 cells—no more, in fact, than the low-metastasis *B*16-*F*1 cells. Moreover, they failed to clump when they were mixed in suspension with lymphocytes. Fidler concluded that the usual effect of lymphocytes on *B*16 cells is to clump them in the circulation and thus facilitate metastasis, and that the inhibition of metastasis observed at high lymphocyte-to-tumor-cell ratios is probably caused by some other type of immune cell that is present with the lymphocytes.

This conclusion is supported by experiments with "nude" mice, a line of hairless mice that lack a mature thymus and therefore have impaired immunity; for example, they accept skin grafts from other species. When we injected labeled *B*16 cells into nude mice, the cells arrested and survived at a lower rate than they did in mice with normal immunity. An injection of lymphocytes from normal mice into the nude mice a day before the tumor-cell injection partially restored the normal pattern of metastasis. Again the indication is that at least in the *B*16 system lymphocytes stimulate experimental metastases.

It is true that in other experiments, in-

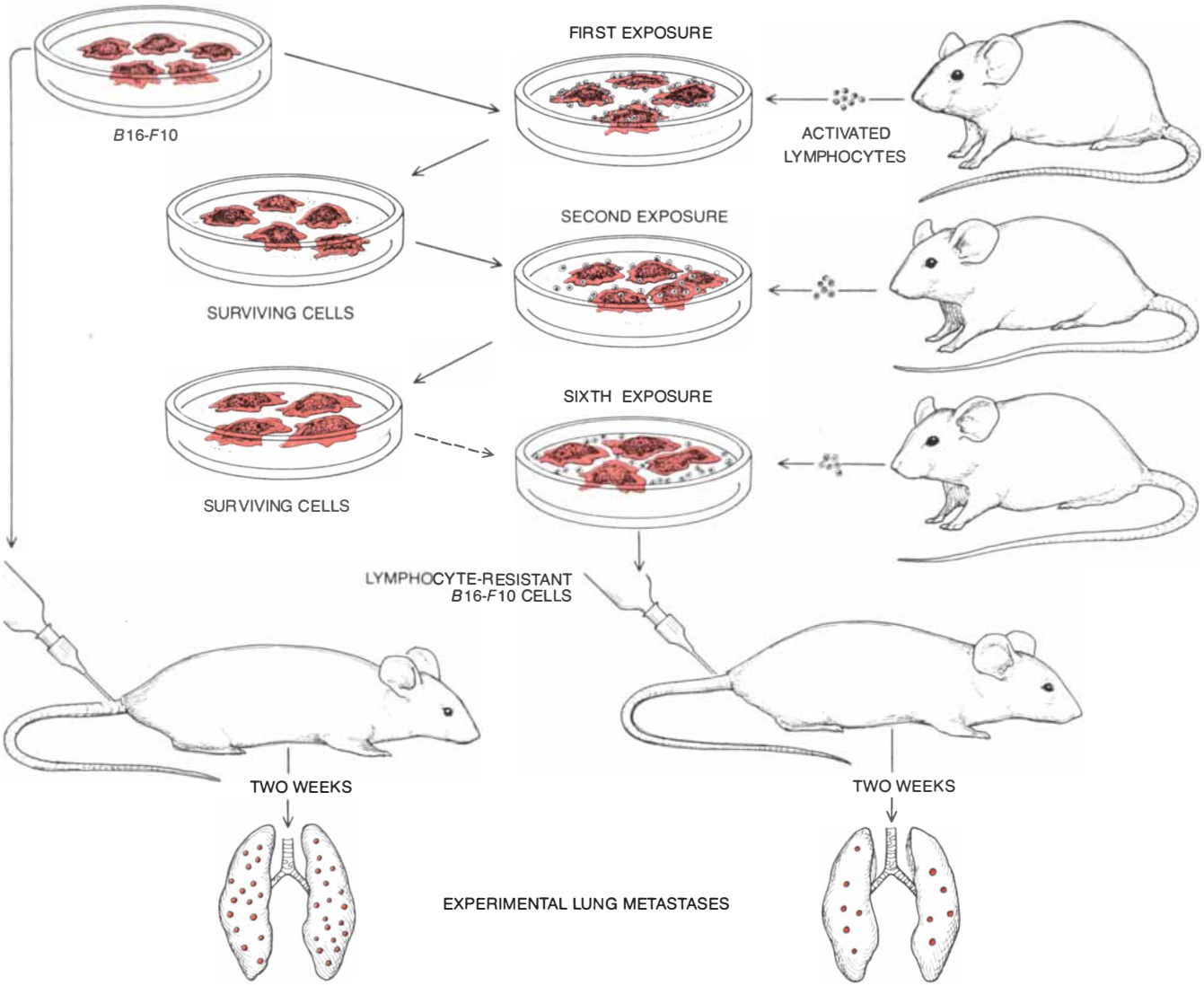


single-cell clone was grown and the cloned cells were injected into other groups of mice. The numbers of metastases produced by the uncloned cells and by each of the clones were very different, indicating that the original population (*1*) must have been highly heterogeneous. If it had been homogeneous, the numbers of metastases produced by uncloned cells and by each of the clones would have been similar.

cluding some done in Feldman's and in Baldwin's laboratories and in Peter Alexander's laboratory at the Chester Beatty Research Institute, lymphocyte-mediated immune responses have been effective in preventing metastatic spread or in eliminating established micrometastases. The fact that one cannot easily predict whether host immunity will enhance or inhibit tumor spread means that immunity and metastasis are not related in a simple way. The experimental findings are reinforced by several clinical observations: procedures designed to augment a cancer patient's immune status have inhibited metastasis in some cases and stimulated metastasis in others. Chemotherapy or immunotherapy for cancer may, in addition to killing some cancer cells, tend to select subpopulations of resistant cells that are highly metastatic. Whether the therapeutic effect outweighs the undesirable selection effect is a matter of judgment in each case.

Certainly the design of effective therapies should take into account the fact that only a few highly metastatic cells in a tumor population may ultimately do the most damage. Recent experiments in my laboratory by Reuben Lotan show that each selected *B*16 melanoma line or unselected *B*16 clone responds differently to drugs administered in chemotherapy. The heterogeneity of tumor populations may be responsible for some clinical and experimental failures of chemotherapeutic drugs if resistant subpopulations survive the drug treatment. A drug may be chosen because it inhibits the growth of a primary tumor or of its tissue-culture counterpart and yet may fail to halt the growth of some tumor-cell subpopulations that are present in micrometastases.

Future efforts will undoubtedly be focused on understanding the biology of tumor progression and the metastatic process. Studies of tumor-cell lines that metastasize preferentially to particular organs may elucidate the mechanisms controlling the homing of those cells and their implantation and growth in a chosen environment; studies of cell lines that are resistant to certain drugs or to the immune system should make it possible to learn why some tumors defeat attempts at therapy. Learning more about metastasis and tumor progression in animal tumor systems should lead to the design of more rational therapies for the management of cancer.



EFFECT OF LYMPHOCYTES ON METASTASIS was studied by Fidler. He exposed *B*16-*F*10 melanoma cells to activated lymphocytes, that is, to lymphocytes from the spleen and lymph nodes of mice that were immune to *B*16 tumor cells. Some of the *B*16-*F*10 cells were killed; the survivors were again subjected to the activated lymphocytes. After the process had been repeated six times the remaining *B*16-*F*10 cells were completely resistant to killing by these lymphocytes. Yet when the resistant cells were injected into mice, they produced many fewer lung metastases than untreated *B*16-*F*10 cells. Resistance to lymphocytes did not help cells metastasize. Rather it reduced their metastatic potential by eliminating lymphocyte–tumor-cell interactions that are important for arrest and survival.

Food, clothing, and shelter

are certainly important, but until less than a generation ago there was little appreciation of mankind's basic need for abundant, swiftly and efficiently produced, handsomely legible copies of whatever has previously been committed to paper.

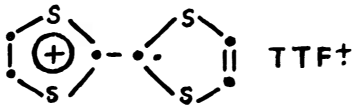
Photoconductive materials meet that need. They meet it so well that mankind—all but a precious few of us—can forget that that's how it's done, just as they needn't comprehend the electron transfers that make photosynthesis work to fuel them at dinnertime.

Kodak Ektaprint copier-duplicators have proved themselves in the marketplace by using organic photoconductors. This makes us want to know more

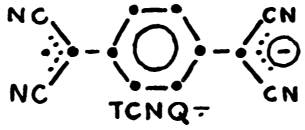
about such materials. Jerome H. Perlstein *(left above)* was an inorganic chemist until seized with fervor for the notion that with organics the possibilities for tailoring photoconductivity for other uses are limitless and easier.

We are one of several high-technology companies who feel a need to know more about the chemical physics of organic photoconductors before getting down to commercial nitty-gritty, our copier success notwithstanding. Organic photoconductors and organic dark conductors that could be linked as polymers and cast from a solvent would suit our talents nicely.

When he was still a professor, Dr. Perlstein participated in the construction of tetrathiafulvalene (TTF). As a cation it complexes notably with



anionic tetracyanoquinodimethane (TCNQ) to form a conductive salt. The



planar radical ions, each with an unpaired electron, stack up so that electrons run up and down the TCNQ stack and holes up and down the TTF stack. So much research has been lavished by other labs on TTF/TCNQ that Perlstein and Larry Isett *(joining Kodak Research Laboratories colleague Perlstein in posing symbolically with the photoconductive belt from a Kodak copier)* have been seeking their thrills elsewhere. Dr. Isett has come up with *bis*(tetrathiotetracene)triiodide, a new organic metal with a conductivity in the good direction at room temperature no worse than about 1/1000 that of copper. (Electrical and magnetic details in *Phys. Rev. B* 18:439 [1 July '78].) For a copy of Perlstein's tutorial *"Organic Metals"—The Intermolecular Migration of Aromaticity* write Dept. 55-W, Kodak, Rochester, N.Y. 14650. If you would rather have a rational solution to the problem depicted below, use that same address.

Kodak

"That's the group for the stockholders' guided tour of Advanced Technology... We're just waiting in line to use the copier."

SCIENCE AND THE CITIZEN

The Literate U.S.

Literacy in the U.S., according to the current cliché, is in a bad way. It is generally accepted that people read and write less and do so less well, that young people are even worse and that perhaps it does not matter, since the electronic media are rapidly making the written word irrelevant. John R. Bormuth of the University of Chicago denies the truth of these popular notions. In an article in the journal *Visible Language* he argues that there is no good evidence that the level of reading and writing skills is lower than it used to be and is declining, and that on the contrary a large and growing proportion of the population have attained a high level of literacy and the volume and economic value of written communication have been increasing.

Bormuth defines literacy as the exchange of information through the written word. He thus views literacy "not solely as an aesthetic endeavor, as a problem of educational concern, as a matter of social equity or as a necessity of democratic government but also and primarily as an *economic* activity," which he undertakes to quantify. Although "tests of reading or writing skills... measure only an intermediate product of a component activity of the literacy program rather than measuring the end product, literacy itself," Bormuth begins by disparaging recent evaluations of the population's ability to read and write. He maintains that actually "there are no reports of repeated observations of [students'] actual literacy... and often poorly designed studies reporting comparisons of achievement-test scores over time." Whereas "it seems fairly plausible" that a long-term increase in reading ability was indeed slowed or even reversed after 1965, the studies necessarily do not include the scores of school dropouts. The dropout rate declined sharply after 1965, so that "larger and larger numbers of low-ability students stayed in school, where their ability would be tested."

Studies of adults' reading ability, according to Bormuth, fail to justify the choice of the reading task presented and of the criterion selected to define illiteracy. He finds the "astonishingly large" percentages of college graduates and of professional and managerial workers classified as illiterate by several studies to be "ample grounds for impugning" the reliability of those studies. And he points out that the testing agencies and corporations conducting the tests had "vested interests in leading the public to believe that the adult illiteracy rate was very high." As for the general complaint of college teachers and employers that young people cannot write, Bormuth answers that both professors and supervisors of white-collar workers are being exposed to a broader segment of the population than they used to be.

As evidence of an "increasing level of literacy" in "a large and growing fraction of the population" in the past 25 years or so, Bormuth cites such statistics as these: The number of items checked out of libraries rose from 2.5 per person to 4.4; white-collar workers (whose jobs depend on reading and writing) rose from 36 percent of the work force to more than 50 percent; estimated time spent reading on the job rose from 104 minutes per person per day to 141 minutes; the number of books and pamphlets purchased rose from 6.8 to 8.6 per person per year; although newspaper circulation increased less than the number of households, newspapers now have more pages, so that the volume of newspaper material purchased increased from 38 to 52 pages per household per day; 235 pieces of mail (air, priority and third class) per person were mailed in 1950 as against 349 in 1975; in addition to the well-known increase in high school and college attendance, there has been an increase in the number of days in the school year and therefore presumably in the amount of literacy instruction per student.

Bormuth calculates that total annual U.S. expenditures for "literacy activities" (publication and distribution of materials, instruction in literacy skills, and salaries and wages for the time spent reading and writing at work) rose from $138 billion in 1947 to $407 billion in 1972 (in constant dollars), from $961 per person to $1,949 and from 20.8 percent of the gross national product to 26.4 percent. Is literacy worth the cost? Bormuth compares the estimated amounts paid to people for reading and writing on the job (the benefits) with the estimated amounts spent to produce and distribute instruction materials and provide instruction in literacy skills (the costs) and concludes that the benefits have consistently been between five and six times as large as the costs.

Moreover, literacy "is one of our nation's most important economic activities," provided it is not "trivialized" by "dwelling primarily and even exclusively on its aesthetic, literary and intellectual uses." In 1972 total expenditures for literacy amounted to more than a fourth of the G.N.P. The amount was twice the size of the defense budget and 5.8 times as large as the amount spent for buying new and used cars. "Literacy activities" consumed about 29 percent of the average worker's time on the job and about 17 percent of the average adult's waking hours. "Knowledge is our most important resource," Bormuth concludes. The advance of technology and the increased complexity of social organization make the communication of information more crucial. "Although we have developed many other media for communicating some of this information, the written word has borne and continues to bear a large fraction of the load. The amount, and perhaps the fraction, has been growing rapidly and steadily."

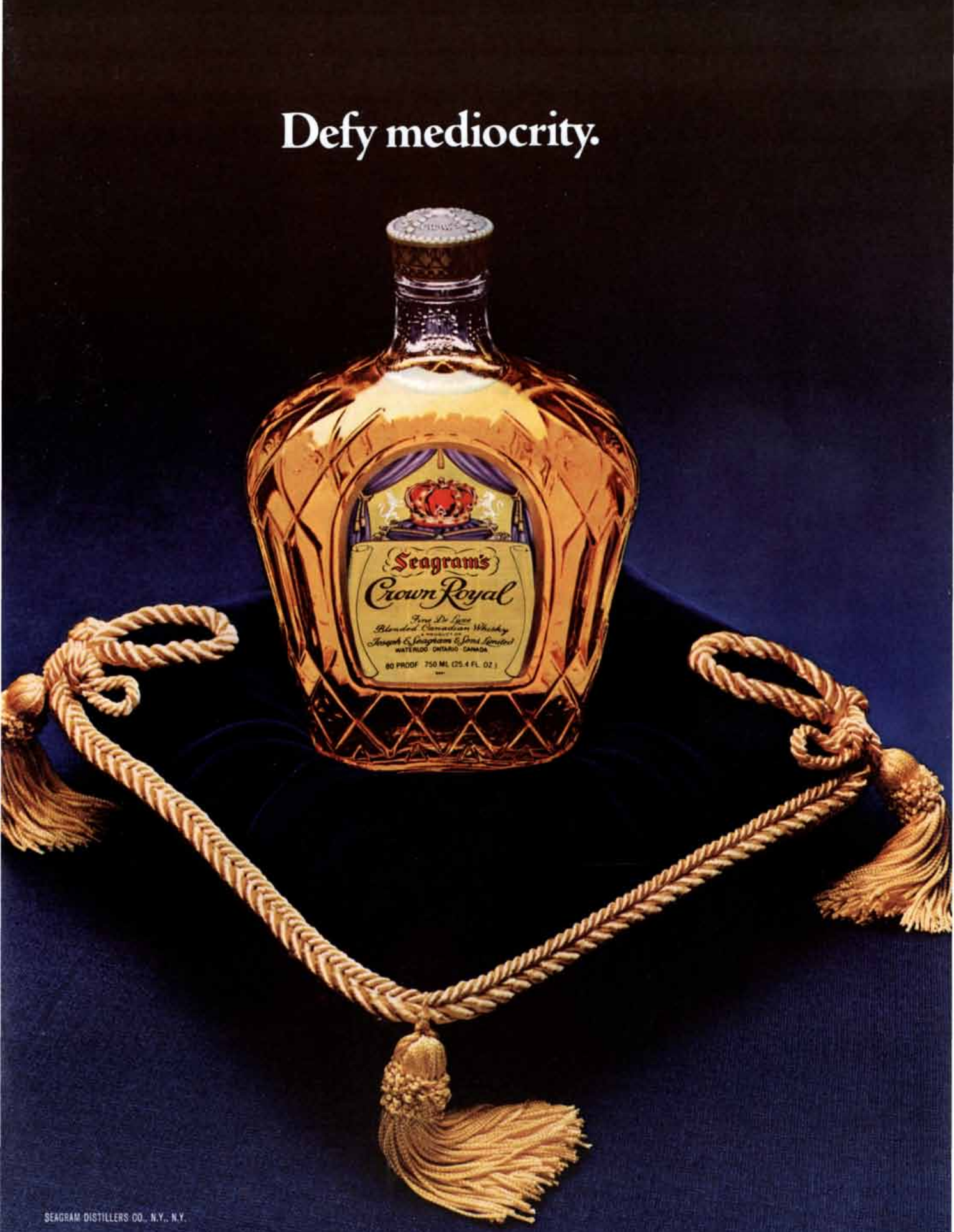
The Thirty-Five Steps

Ever since Friedrich Wöhler achieved the first synthesis of an organic substance, urea, in 1828, chemists have tested their skills by trying to duplicate in the laboratory a succession of increasingly exotic compounds obtained from living organisms. Once a substance has been extracted and its structure established the organic chemist can call on the vast arsenal of techniques developed since Wöhler's day to assemble the three-dimensional structure the living cell elaborates with the enormously complex but highly efficient molecular tools called enzymes.

At any given time there are usually a dozen or so natural products of biological or pharmaceutical interest that challenge the organic chemist's ingenuity because they resist duplication in the laboratory by all previously known synthetic approaches. In the past third of a century such challenges have been presented by penicillin, cortisone, morphine, chlorophyll, vitamins A and B-12, reserpine (the snakeroot derivative and the first of the tranquilizers), insect juvenile hormones, colchicine (the plant product that inhibits cell division), erythronolide B (the precursor of the erythromycin family of antibiotics) and the prostaglandins (animal hormones with a wide variety of subtle effects). In each case laboratory synthesis was achieved, but only after lengthy experimental work.

Curiously the complexity of a particular synthesis is not readily apparent to the uninitiated. A compound's superficial resemblance to a more familiar molecule can be deceptive. There is no test of the difficulty of a synthesis except to observe how many chemists consider the problem worth tackling and how long the solution takes. On the basis of such criteria the recent synthesis of gib-

Defy mediocrity.



SEAGRAM DISTILLERS CO., N.Y., N.Y.

Physician, did you miss any of these significant developments in medical science?

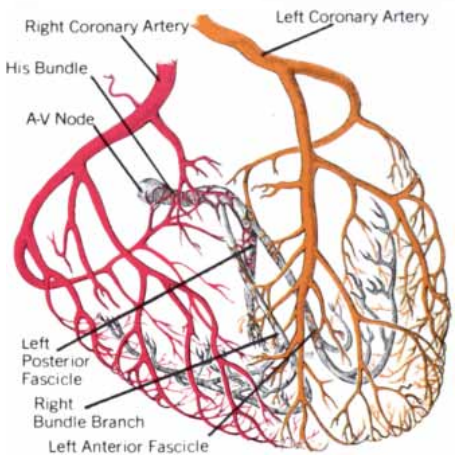
• The lidocaine level in the blood may fall below the therapeutic range when the drug is given as a single bolus followed by a continuous infusion. To avoid a "therapeutic hiatus," a second bolus should be given 20 to 30 minutes after the first bolus...

• Evidence is mounting that there is widespread over-diagnosis of pulmonary embolism at a time when heparin, in therapeutic doses, has become an important cause of drug-related death in the relatively well, hospitalized patient...

• The widely used street drug, phencyclidine, appears to cause persisting and possibly permanent brain dysfunction after repetitive use...

• There is alarming incidence of acute leukemia which may reach five percent in patients who appear to have been successfully treated with chemotherapy or radiation therapy for Hodgkin's disease...

Every day medical science evolves in ways that affect patient management and therefore patient well-being. But with more than 2,000 journals published, innumerable meetings held and hundreds of courses given, it is all too easy for developments that compel change in diagnosis and treatment to get by even the most conscientious reader and meeting-goer. And since such discoveries are today's news, they will not appear in the standard texts for years.



Branches of the right and left coronary arteries supply blood to the AV node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

We have a solution to this problem: It is called SCIENTIFIC AMERICAN *Medicine*. It makes such developments accessible to you when and where you need them, month after month.

Here is how SCIENTIFIC AMERICAN *Medicine* works. When you subscribe, you receive a two-volume, 2,000 page text, organized according to the familiar subspecialties so that you know where to find the answers to your questions.

The authors of SCIENTIFIC AMERICAN *Medicine* are leading authorities from the faculties of the Harvard Medical School and the Stanford University School of Medicine. These seasoned scholar-practitioners, close to the creation of new medical knowledge, write for you the practitioner, not the student. Therefore, at the time you receive your copy of SCIENTIFIC AMERICAN *Medicine*, you will hold in your hands the world's most advanced text of medicine. As long as you continue to subscribe it will stay that way.

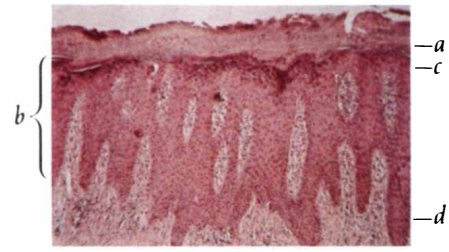
Each month you will receive several updated chapters. The new chapters which replace the old chapters contain reports of the significant advances in medical science that affect patient management. The index will also be updated each month. Furthermore, updated material will be summarized in a monthly bulletin accompanying the updated chapters so that you can quickly scan the information.

As the updated material is incorporated into the text, older more familiar material is deleted. Thus the amount of useful knowledge increases but the number of pages does not.

A Continuing Medical Education Program

For subscribers who wish it, SCIENTIFIC AMERICAN *Medicine* offers at no extra charge a continuing medical education program sponsored by the Stanford University School of Medicine's Office of Postgraduate Medical Education and administered by the Educational Testing Service of Princeton, New Jersey. The program consists of eight patient management problems per year, each written by a member of a Panel of Examiners drawn from leading clinical institutions. By completing all eight PMP s in a 12-month period the subscriber can earn 32 Category 1 credits per year, from the Stanford University School of Medicine, toward the AMA's Physician's Recognition Award.

This complete, concise medical knowledge service is yours for the drop of the accompanying card. Please fill it out and mail it to us. We think you will find that SCIENTIFIC AMERICAN *Medicine* is your answer to the challenge of bringing all that modern medicine has to offer to your patients.



Histopathologic changes in psoriasis.

Extensive psoriasis (top) cleared completely (bottom) after four weeks of treatment with anthralin.

THE DISTINGUISHED AUTHORS, AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN *MEDICINE*.

1. Cardiovascular Medicine

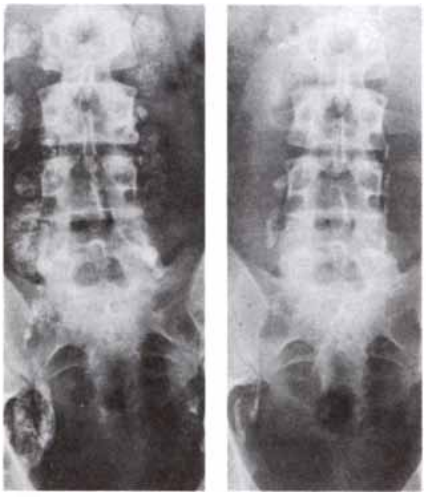
Edgar Haber, M.D., Professor of Medicine, Harvard Medical School, Physician and Chief of Cardiology, Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Professor of Medicine, Stanford University School of Medicine

Roman W. DeSanctis, M.D., Professor of Medicine, Harvard Medical School, and Physician and Director, Coronary Care Unit, Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Associate Professor of Medicine, Harvard Medical School, and Assistant Physician and Associate Director, Coronary Care Unit, Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Instructor in



Extensive retroperitoneal Hodgkin's disease before and after MOPP chemotherapy.

Medicine, Harvard Medical School, and Chief, Hypertension Unit, Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Professor and Chairman, Department of Dermatology, Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Professor of Medicine and Dean for Students and Alumni, Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Professor of Medicine and Head, Division of Gastroenterology, Stanford University School of Medicine

5. Hematology

Stanley L. Schrier, M.D., Professor of Medicine and Head, Division of Hematology, Stanford University School of Medicine

6. Immunology

John David, M.D., Professor of Medicine, Harvard Medical School, and Physician, Robert B. Brigham Hospital

7. Infectious Disease

Thomas C. Merigan, M.D., Professor of Medicine and Head, Division of Infectious Diseases, Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Professor of Medicine, Harvard Medical School, and Chief, Infectious Disease Unit, Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Assistant Professor of Medicine, Harvard Medical School, and Hospital Epidemiologist and Associate Physician, Massachusetts General Hospital.

Adolf W. Karchmer, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Associate Physician, Massachusetts General Hospital

Robert H. Rubin, M.D., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

8. Intensive And Emergency Care

Edward Rubenstein, M.D., F.A.C.P., Associate Dean of Postgraduate Medical Education, and Professor of Clinical Medicine, Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Professor of Medicine, Harvard Medical School, Director of Research, Howard Hughes Medical Institute, and Physician, Peter Bent Brigham Hospital

10. Nephrology

Roy H. Maffly, M.D., Professor of Medicine, Stanford University School of Medicine, and Chief, Renal Service, Palo Alto Veterans Administration Hospital

11. Neurology

Robert W.P. Cutler, M.D., Professor of Neurology, Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Professor of Medicine and Radiology, and Chief, Division of Oncology, Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Associate Professor of Psychiatry, Harvard Medical School, and Chief, Psychiatric Consultation—Liaison Service, Massachusetts General Hospital

14. Respiratory Medicine

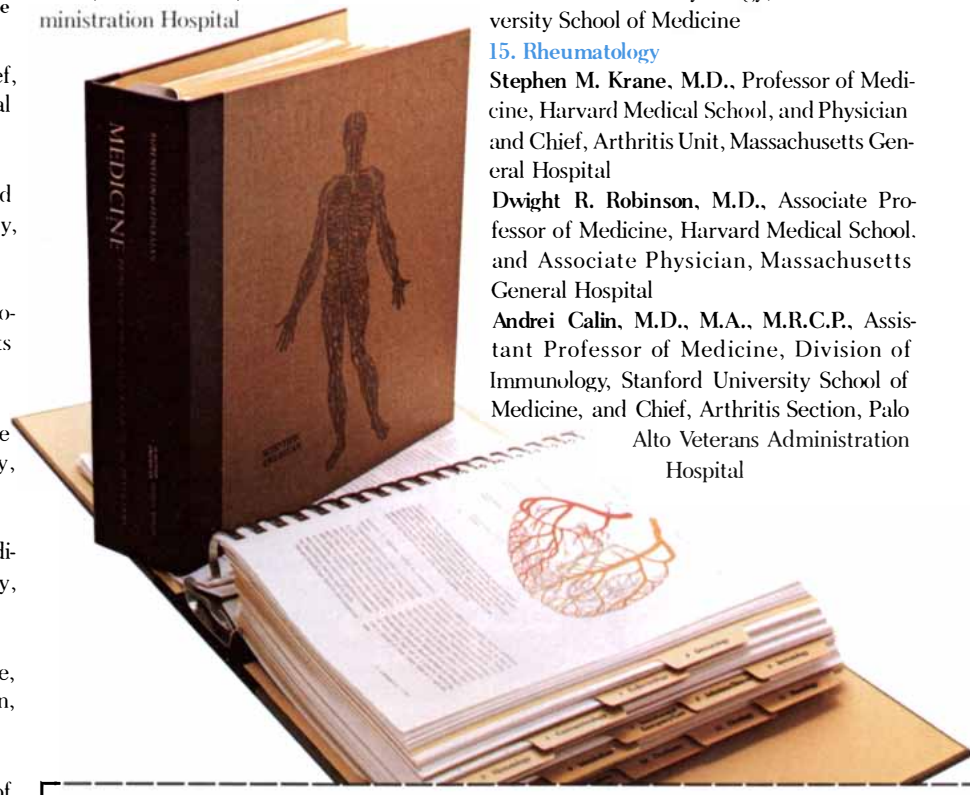
Eugene D. Robin, M.D., F.A.C.P., Professor of Medicine and Physiology, Stanford University School of Medicine

15. Rheumatology

Stephen M. Krane, M.D., Professor of Medicine, Harvard Medical School, and Physician and Chief, Arthritis Unit, Massachusetts General Hospital

Dwight R. Robinson, M.D., Associate Professor of Medicine, Harvard Medical School. and Associate Physician, Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Assistant Professor of Medicine, Division of Immunology, Stanford University School of Medicine, and Chief, Arthritis Section, Palo Alto Veterans Administration Hospital



SCIENTIFIC AMERICAN MEDICINE

Post Office Box 5911, New York, N.Y. 10017

Gentlemen: Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement, and update it regularly by sending me new monthly subsections. As part of this service, I understand that the annual subscription of $150 for SCIENTIFIC AMERICAN *Medicine* is tax-deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a complete refund. **SA3-79**

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*.
☐ I shall also enroll in the CME Program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $150
☐ Please bill me

Name_____

MD Specialty_____

Address_____

City_____ State_____ Zip_____

Subscribers outside of the U.S., its possessions, and Canada will be charged extra for shipping and handling; updates will be sent by surface routes unless air mail delivery is requested.

berellic acid by Elias J. Corey and his co-workers at Harvard University has been widely applauded.

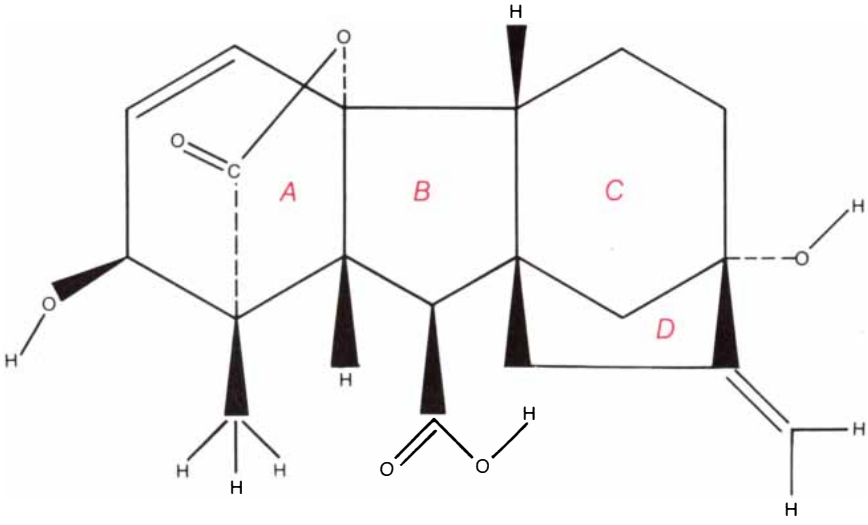
Gibberellic acid, one of the family of substances known as gibberellins, is a natural hormone that stimulates the growth and maturation of plants. It was first recognized about 40 years ago in Japan as the product of a fungus, *Gibberella fujikuroi,* responsible for a disease of rice plants that causes the stalks to grow tall and spindly. Since the 1950's the gibberellins have been manufactured commercially from the fungus by fermentation. Today the gibberellins' biggest use is in increasing the size of grapes and stimulating the growth of sugarcane.

After a false start in which the three-dimensional position of one hydrogen atom was mislocated the structure of gibberellic acid was established in the early 1960's by X-ray crystallographers at Imperial Chemical Industries Limited and at Harvard. In the next 15 years the synthesis of gibberellic acid was the subject of about 150 papers from some 25 different laboratories in the U.S., Britain, Japan, Israel and Australia. None reported success.

The structure of gibberellic acid looks innocent enough (*see illustration below*). The three-ring system of the molecule, consisting of a five-member ring sandwiched between two six-member rings, is found in many natural products and was first synthesized many years ago. Unfortunately the chemist cannot simply tack atoms or appendages onto a familiar structure at will, for example the lactone (O–CO) bridge across the first (*A*) ring in gibberellic acid or the two-carbon unit that forms a fourth (*D*) ring below the third (*C*) ring. The lactone and the highly reactive double bond on the *D* ring make the molecule so unstable that it comes apart rapidly even in neutral water.

Corey, who has introduced a succession of ingenious procedures into organic synthesis, began work on gibberellic acid in 1963 with one graduate student. When the original plan of attack failed at one crucial point, Corey turned his attention to other problems, notably the syntheses of insect juvenile hormones and prostaglandins, both of which were achieved. By 1971 he had devised a new strategy for the synthesis of gibberellic acid and began work with another graduate student, Rick L. Danheiser, who remained for the next six years as Corey's chief collaborator.

In brief, the synthesis proceeds from right to left, as the molecule is conventionally drawn. The synthesis begins with the familiar benzene derivative catechol (whose known synthesis did not have to be repeated), which provides the *C* ring. To this is added the *B* ring, initially in the form of a six-member ring but later collapsed to a five-member structure. The *D* ring is constructed stepwise below the *C* ring, and the *A* ring is similarly constructed to the left of the *B* ring. A lactone bridge is then erected across the *A* ring. One of the most difficult problems, reminiscent of a similar key step in the synthesis of cortisone, was the placement of the OH group at the juncture of the *C* and the *D* ring. The entire synthesis requires about 35 steps, many of them involving novel catalysts and reaction conditions. Although the laboratory synthesis does not seem to have commercial utility, it will make possible the creation of modified forms of gibberellin whose effect on plant cells may throw light on the mode of action, now almost completely unknown, of the normal substance. In addition to Danheiser, Corey's co-workers included Srinivasan Chandrasekaran, Gary E. Keck, B. Gopalan, Samuel D. Larsen, Patrice Siret and Jean-Louis Gras.



STRUCTURE OF GIBBERELLIC ACID, synthesized by Elias J. Corey and his co-workers, basically consists of four rings. Wedges are bonds that project up from plane of the molecule.

Down on the Fish Farm

Aquaculture, or fish farming, is widely considered to be a promising way of augmenting the world's supply of animal protein. Since it is not yet pursued on a large scale, however, much remains to be learned about what can be accomplished with it. The International Development Research Centre (IDRC), a public corporation established by the Canadian parliament to adapt science and technology to the needs of developing countries, has supported a number of research projects in aquaculture. Summarizing them in a booklet titled "Fish Farming," the IDRC concludes: "With proper management aquaculture has great potential for increased production—sufficient to meet the demand for fish protein in many developing countries." Moreover, "much of the technology is relatively simple, inexpensive and consumes a minimal amount of energy."

The IDRC has given priority in its program to the improvement of breeding stocks and the increase of yields. In breeding the critical problem is that many fishes do not reproduce well in captivity. The reason is hormonal. In several projects supported by the IDRC such fishes did breed successfully after they were injected with gonadotropic hormone that was extracted from the pituitary glands of Pacific salmon being processed by commercial fish packers. The species induced to breed in this way include Chinese carp in Malaysia and milkfish in the Philippines. Investigators in Asia are now looking for sources of gonadotropin that can be obtained more cheaply and are more species-specific.

The chief approach to the increase of yields has been through experiments in polyculture, which means stocking a pond with several compatible fish species of different feeding habits. When this approach is followed correctly, the yield is far greater than it is when the pond is stocked with an equal number of fish of a single species. Work in India has shown that a combination of three local carp species and three species from other regions produced yields 10 or more times greater than those obtained from monoculture systems.

The IDRC also supports several projects aimed at increasing the production of mollusks, particularly oysters and mussels. They "are in many ways ideal animals for aquaculture," since they do not require pens (being sedentary), have simple nutritional requirements and can be stocked at remarkably high densities with no apparent ill effects. In Sierra

DP Dialogue

Notes and observations from the IBM Data Processing Division that may prove of interest to the engineering community

Prototype model of ITT Gilfillan PAR-80 phased array radar, which guides an aircraft through 20 miles of approach to a precise touchdown. Computer simulation played a major role in its design.

Simulators Search Out Optimal Radar Designs for Gilfillan

When a search radar finds a target in the sky, a "signature" is hidden in the reflected pulse: a characteristic mix of signal phases and polarities from which the aircraft can be identified as, for example, a fighter, bomber, or commercial airliner. Systems made by ITT Gilfillan, a division of International Telephone and Telegraph Corporation, include digital filters that extract this signature from the faint microwave pulses reflected back into the radar antenna.

To design these systems, ITT Gilfillan simulates several radar functions in an IBM System/370 Model 148. "Our engineers wrote the simulator programs themselves in APL and PL/I," says Eric Berg, engineering group supervisor. "They enter parameters and run them from online computer terminals in the engineering department."

Geometric Model

In one computer model engineers can enter a geometric representation of a target. The radar return signal is simulated from an approximation of the energy that would be scattered by an ensemble of geometric shapes. A complete signal is generated by rotating the geometric representation over the angle space normally flown by an aircraft. It is difficult to distinguish the synthetic signal from measurements made on an actual target.

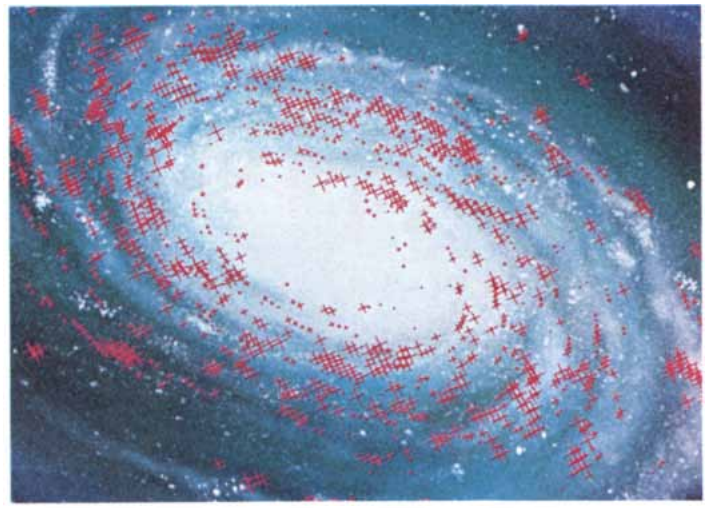
These simulated synthetic signals provide the engineer with the tool whereby he can design the optimum radar transmitter and receiver. Convergence to the optimum design is hastened because only the designs proven through simulation are tested in the real radar.

Computer simulation also helped ITT Gilfillan find the optimal scanning program used in the PAR-80. The sweeping motion of a radar beam in space is often more complex than the familiar side-to-side antenna movement. It may scan mechanically according to a stored program, or electronically by cycling phase relationships in the antenna. In some systems, frequency and power are modulated during the scan cycle.

"By testing scan programs against typical targets in a computer simulation, we can measure coverage, tracking accuracy, and detectability of the target," Berg adds. "In a system for landing control, for example, it's important to prevent momentary loss of an aircraft signal during its approach. Here the model finds holes in coverage of the glide path. In one recent design we simulated 66 different scan programs within a few days. The computer plotted the scan pattern of each and the systems engineer picked the one that gave the best coverage of the approaching aircraft.

"In a short period of time, a designer can make repeated trial runs, each guided by the results of the previous one. Interactive computing is a uniquely powerful tool in the search for an optimal radar system design."

Probing the Mystery of Spiral Galaxies



A computer plot is superimposed on an artist's rendering of the spiral galaxy known as M81. The computer model, developed at the IBM Thomas J. Watson Research Center, produces spiral arms similar to those of real galaxies.

Galaxies are huge aggregations of stars, gas and interstellar dust, thousands of light-years in diameter, and with a total mass tens or hundreds of billions of times that of the sun. The majority are disk shaped and exhibit a spiral structure.

A computer model developed by H. Gerola and P. E. Seiden at IBM's Thomas J. Watson Research Center, Yorktown Heights, New York, lends powerful support to a new theory of spiral structure. Observations indicate that a galaxy undergoes "differential rotation," explains Dr. Gerola, Research Staff Member in the Center's Department of Physics. "It turns faster at its outer edge than nearer the center, so that there is shear – a pulling apart of adjacent points – within the disk. Furthermore, new evidence provided recently by radio-astronomy suggests that new stars are constantly being formed in the galaxy by chain reactions; when a high-mass star explodes into a supernova, the shock wave compresses the surrounding gas, triggering the nucleation of new stars, some of which become supernovae."

Gerola and Seiden connected these two facts through computer simulation of star formation by chain reaction in a differentially rotating galaxy. The key was recognizing that the star formation process was "stochastic." That is, new massive stars are not necessarily created adjacent to a supernova; rather, a probability exists for their formation.

Gerola and Seiden constructed a two-dimensional model of a galaxy, divided into rings mathematically rotated by the computer according to the observed differential motion. In turn, the rings were further subdivided to cover the whole galaxy with "cells" of equal area. The researchers then instructed the IBM System/370 computer to populate about one percent of the cells, randomly selected, with bright new stars. Each was assigned a finite probability at successive intervals of creating a similar star in an adjoining cell, simulating the consequence of a supernova explosion.

"When we introduced these effects," Seiden says, "the model immediately began to produce large-scale spiral features.

"Although individual stars are constantly being formed and disappearing in explosions, they become grouped in 'bunches' which are gradually stretched into spiral arms by the differential rotation of the galaxy.

Simulated galaxies generated by the Gerola-Seiden model develop spirals surprisingly similar to those of real galaxies, in density, pitch angle, and quasirigidity of rotation. While individual elements do not persist, they continuously regenerate so that the model exhibits a well-defined spiral structure over the entire lifetime of the galaxy.

Gigi Penn and Wausau Homes, Inc., which makes prefabricated houses, found each other through the Wisconsin Job Service.

Matching the Jobless to the Job in Rural Wisconsin

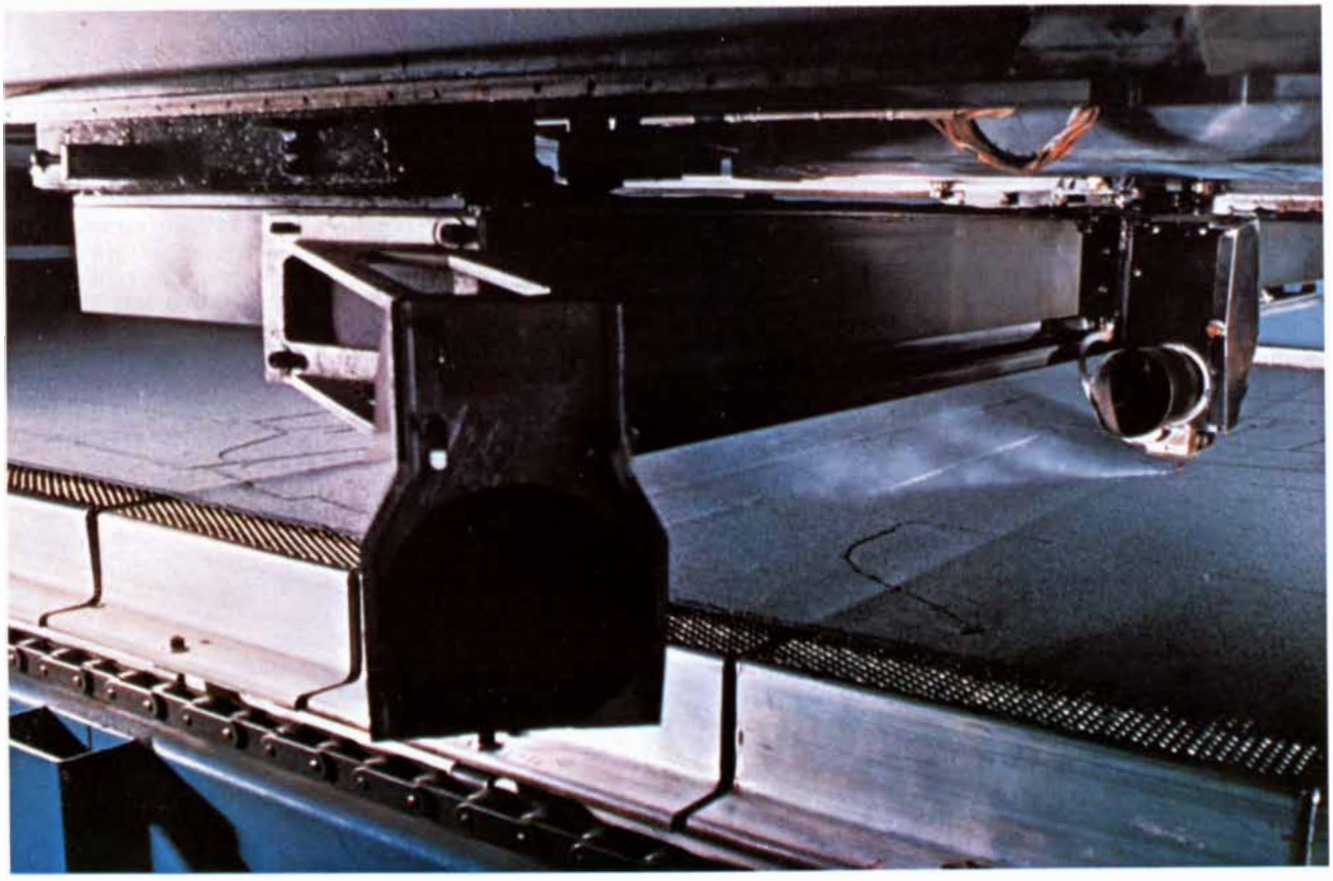
Unemployment is usually thought of as an urban problem, but there are people out of work in rural areas, too. Now computer-assisted job matching is helping rural Wisconsin residents find work.

This program is the first of its kind anywhere in the country outside of a major city. It matches the basic characteristics of a person – education and experience – with those of job openings, explains Hugh Kelly, a bureau director with the Wisconsin Department of Industry, Labor and Human Relations.

A state government computer in Madison, the capital, is online to state-operated Job Service Offices in Wausau and Merrill. When an employer places a job order in either office, the IBM System/370 reviews the entire file of applicants to identify those best suited to the opening. It provides fast attention to employers when they have a vacancy and lets them see the most qualified applicants available. The system also helps applicants by determining their unemployment benefits.

The program has implications for the nationwide systems now being implemented in 25 states, with additional states being added as they are funded. Since the Wisconsin Job Service is part of the national network of 2,800 offices, the project has a huge potential in helping tap the reservoir of unemployed who want to find work.

And – not least in importance – it will help cut down on paperwork, freeing professional people to work directly with applicants and employers.



High-speed laser cutter slices through fabric for men's suits. The beam is directed into the cloth by a system of mirrors on the moving arm.

Laser Tailors Men's Suits for Richman Brothers

Responding to information recorded on a magnetic tape by a computer, the mechanical arm flies over the surface of the cloth at velocities which reach 36 inches per second. The blue-white light from its end makes a fine, surgically clean cut as it moves. The beam, from a laser, is directed into the cloth by a mirror system attached to the moving arm. The cloth is destined to become a man's suit.

J. O. Burton is manager of Manufacturing Research and Development for Richman Brothers, a major men's clothing manufacturer and retailer headquartered in Cleveland. "Through the use of the computer and a laser cutter, we are saving 10 to 15 percent of cloth which was formerly wasted by manual planning and cutting techniques," Burton says. "In addition, we receive consistent accuracy on all cut parts which in the end means we produce a better tailored garment for our customer."

Use of the IBM System/370 Model 158 in manufacturing begins when cloth is delivered to Richman Brothers. It is first run through processing machines on which sensors – aided by an observer at an online terminal – note any blemishes or flaws. Other sensors report the position of each flaw. This information will be stored in the computer to provide instructions for automatic cutting around these damaged areas.

Another integral step in the manufacturing process occurs when the computer "grades" the designer pattern, usually a size 40 regular, to produce patterns for the full range of sizes. Working from the master pattern which has been entered into the computer through a digitizer, each component is scaled up and down the entire size range. This is not a simple process, Burton points out; different elements of the pattern expand and contract in different, complex ways from size to size.

These components are then displayed on an IBM graphic terminal where an operator using a light pen makes the final placement of components in the cutting pattern in order to obtain maximum utilization of the material to be cut. This is the final step that provides the information for the laser cutter.

"We are working our way toward a total system concept," he continues, "using the data base to keep track of fabric, patterns, orders, and the progress of work through the plant. Very shortly the computer will automate the matching of patterns so that plaids and stripes will extend evenly across garment seams.

"These jobs, which were formerly done manually, required countless years of experience and skill. As our knowledgeable people retire, these skills are being lost. Fortunately, with the aid of the computer we will continue to turn out finely tailored suits and be able to reduce our manufacturing costs in the process."

DP Engineering Dialogue is designed to provide you with useful information about data processing applications, concepts and techniques. For more information about IBM products or services, contact your local IBM branch office, or write Editor, DP Dialogue, IBM Data Processing Division, White Plains, N.Y. 10604.

IBM®
Data Processing Division

Leone an oyster-culture project supported by the IDRC yielded in from six to nine months oysters of a size that could be attained only in three or four years in the cooler waters off the east coast of Canada and the U.S. In coastal waters off Singapore experiments in mussel culture have produced yields as high as 250 kilograms per square meter. The IDRC notes that such yields could theoretically produce as much as 100 tons of protein from a one-hectare area of sea surface, whereas one hectare of land planted with soybeans will produce perhaps one ton of protein.

Wheels within Wheels

In describing a river there is no need to specify the state of every water molecule. The movements of individual molecules can be ignored because they have no influence on hydrodynamic currents and waves. There are conditions, however, under which microscopic events can have macroscopic consequences. For water those conditions are a temperature near 647 degrees Kelvin and a pressure near 217 atmospheres. That combination of temperature and pressure defines the critical point of water, where the liquid and the vapor reach the same density and the distinction between the two phases disappears. Near the critical point random fluctuations in density, which are always present but are ordinarily confined to aggregates of a few molecules, grow very large. The water turns milky because the fluctuations scatter light. Precisely at the critical point the scale of the fluctuations becomes infinite, and disturbing a single molecule can change the phase of the entire volume. Any description of water under these conditions would have to account for its structure at all possible scales, from the scale of single molecules to that of the bulk fluid. Formulating such a description has proved to be an intractable problem.

A substantial number of the outstanding problems in physics (and at least a few in other fields) have long resisted solution for the same reason. A closely analogous problem is the onset of spontaneous magnetization as a ferromagnetic material is cooled below the critical point called the Curie temperature. Well above this temperature the spins of unpaired electrons, which are the fundamental magnets of the system, are randomly oriented. As the Curie temperature is approached the spins begin to line up, first at small scales involving only neighboring atoms, then at larger and larger scales. At the Curie temperature, changing the orientation of a single spin could alter the direction of the magnetization eventually assumed by the material as a whole.

Related problems include transitions to superconducting and superfluid states, transitions between ordered and disordered phases in alloys, the development of turbulence in fluids, the conformation of long-chain polymers, the percolation of a liquid through a solid matrix and the interactions of elementary particles in a quantum field theory. What makes all these phenomena difficult to explain theoretically is the need to deal simultaneously with many scales of length or energy. In the past several years a new approach to these problems, called the renormalization group, has begun to make a little progress. The essence of the technique is to treat each scale of length in turn, starting with the smallest.

The mathematical procedure called renormalization was invented in the 1940's as a step in the formulation of quantum electrodynamics, the theory that describes the interactions of an electrically charged particle with an electromagnetic field. The difficulty in that case was an infinite number of energy scales encountered in calculating the mass or charge of an electron. In the 1950's it became apparent that the aims of the renormalization procedure can be achieved by a large family of mathematical transformations; these make up the renormalization group. The more recent problem-solving methods based on the renormalization group differ significantly in mathematical procedure and in physical point of view; indeed, the two methods are different enough for them to have been facetiously labeled "ancient" and "modern." The modern method was introduced in 1971 by Kenneth G. Wilson of Cornell University, who based his work in part on earlier ideas of Leo P. Kadanoff of the University of Chicago.

To explain how the renormalization group works, consider again a ferromagnetic material near the Curie temperature. Examined at a microscopic scale the magnet is a lattice of atoms, each atom having associated with it a spin vector that can point in any direction. The lattice is assumed to be of infinite extent, which at the atomic scale is not an unreasonable approximation. The first step in the analysis of this system is to divide the lattice into blocks of a few atoms each. Next all the spins in a block are replaced by an average spin calculated for the entire block; in this way the number of spins per unit volume is reduced. The final step is to rescale each block to the dimensions of the original lattice and adjust the magnitude of the spin. Through this series of operations a new lattice of atoms is defined, one that reflects the distribution of spins on a slightly larger scale.

If meaningful results are to be obtained, these operations must be repeated many times. (Renormalization-group methods make prodigious demands on computer resources.) During the first several iterations the configuration of the spins may vary wildly from generation to generation. As the computations continue, however, the variations are damped out, and eventually the configuration becomes stable; the system is said to have reached a fixed point. The entire procedure is then repeated with slightly different initial conditions. From the nature of the fixed point and from the way the system approaches it, properties of the magnet at its critical temperature can be determined.

The program for describing a lattice at progressively larger scales is defined mathematically, but it can be given an intuitive justification. The strategy leads to a correct description, Wilson has pointed out, because it relates events at adjacent length scales, which are closely coupled in nature. A momentary fluctuation at the atomic scale of one or two angstrom units (one angstrom unit is equal to 10^{-8} centimeter) can strongly influence fluctuations with a characteristic length of from two to four angstroms, and they affect fluctuations of from four to eight angstroms, and so on.

Communication through a cascade of length scales accounts for one of the most striking properties of critical phenomena: extreme sensitivity to small perturbations. In a ferromagnet near the Curie point a slight change in temperature cannot directly alter the direction of more than a few spins, but when those few changes are propagated through the cascade of length scales, they can give rise to macroscopic effects.

If some distinctions are amplified in this way, it is equally remarkable that others are strongly attenuated. For example, two ferromagnetic materials can have quite different atomic structures, but as the length scale increases, the effects of structural differences gradually disappear. At the Curie temperature the macroscopic behavior of the two ferromagnets is identical. In a larger context the modern theory of the renormalization group seems to confirm a somewhat earlier hypothesis called critical-point universality, which states that any system near a critical point, no matter what its detailed composition, must fall into one of only a few possible categories of critical behavior. The applicable category is determined by the number of dimensions needed to describe the system and by its symmetry properties. Systems that outwardly seem quite dissimilar can be members of the same category and hence should be formally identical. For example, a graph of the density of water near the critical point and a graph of the magnetization of nickel near the

IN A BMW, THE DRIVER IS NOT ASKED TO COMPENSATE FOR THE SHORTCOMINGS OF THE AUTOMOBILE.

The extraordinary road-holding ability of the BMW 320i is graphically illustrated, above, travelling at a high speed through turn #6 at the Sears Point International race course. Note: as the BMW goes into the turn, the unique angling of the MacPherson front struts reduces the lean of the inside front wheel, while the outside front wheel remains vertical; inside and outside rear wheels remain vertical due to semi-trailing arm design in the rear.

Like all honeymoons, the first few months of new car ownership are generally a period of adjustment.

A time when man accommodates himself to the idiosyncracies of his machine—those little quirks of engineering that make the act of driving something to be endured rather than something to be enjoyed.

Perhaps because of our long involvement in motor racing—where such an antagonistic relationship would be intolerable—we at the Bavarian Motor Works in Munich, Germany, take a wholly different approach to building automobiles.

An approach that begins with the seemingly unusual engineering concept that, above all, a car must function as one with its driver.

YOU DRIVE A BMW.
IT DOES NOT DRIVE YOU.

When you drive the BMW 320i for the first time, you will experience a curious sensation of being part of the car itself—an exhilarating feeling of control.

When you press the accelerator, the two-liter, K-Jetronic, fuel-injected engine responds without lag.

Its four-speed transmission (automatic is available) slips precisely into each gear.

The suspension—independent on all four wheels—is quick and clean through the corners.

Its rack and pinion steering is sharp and accurate.

Careful study has been made of the critical relationship between seat location, visual position, steering wheel angle, pedals and controls.

Instruments are clearly visible and illuminated at night by an optically beneficial orange light; controls are readily accessible.

Pedal direction and pedal pressure have been carefully balanced to reduce fatigue and facilitate effortless gear changing.

All told, the only major adjustment you will have to make as the new owner of a BMW 320i is that you will have no adjustments to make at all.

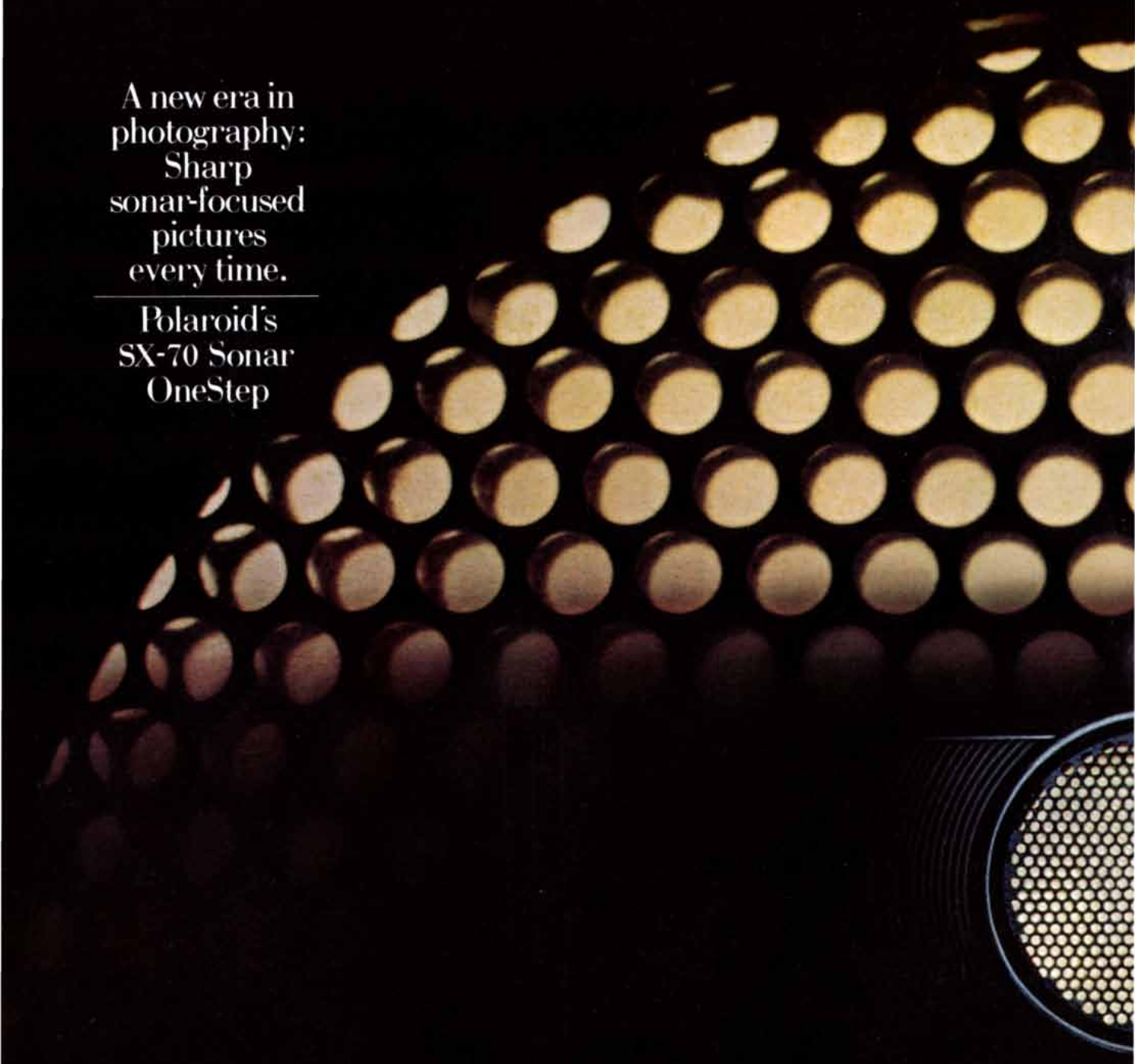
If the thought of owning such a car intrigues you, call us anytime, toll-free, at 800-243-6000 (Conn. 1-800-882-6500) and we'll arrange a thorough test drive for you at your convenience.

THE ULTIMATE DRIVING MACHINE.

Bavarian Motor Works, Munich, Germany.

A new era in photography: Sharp sonar-focused pictures every time.

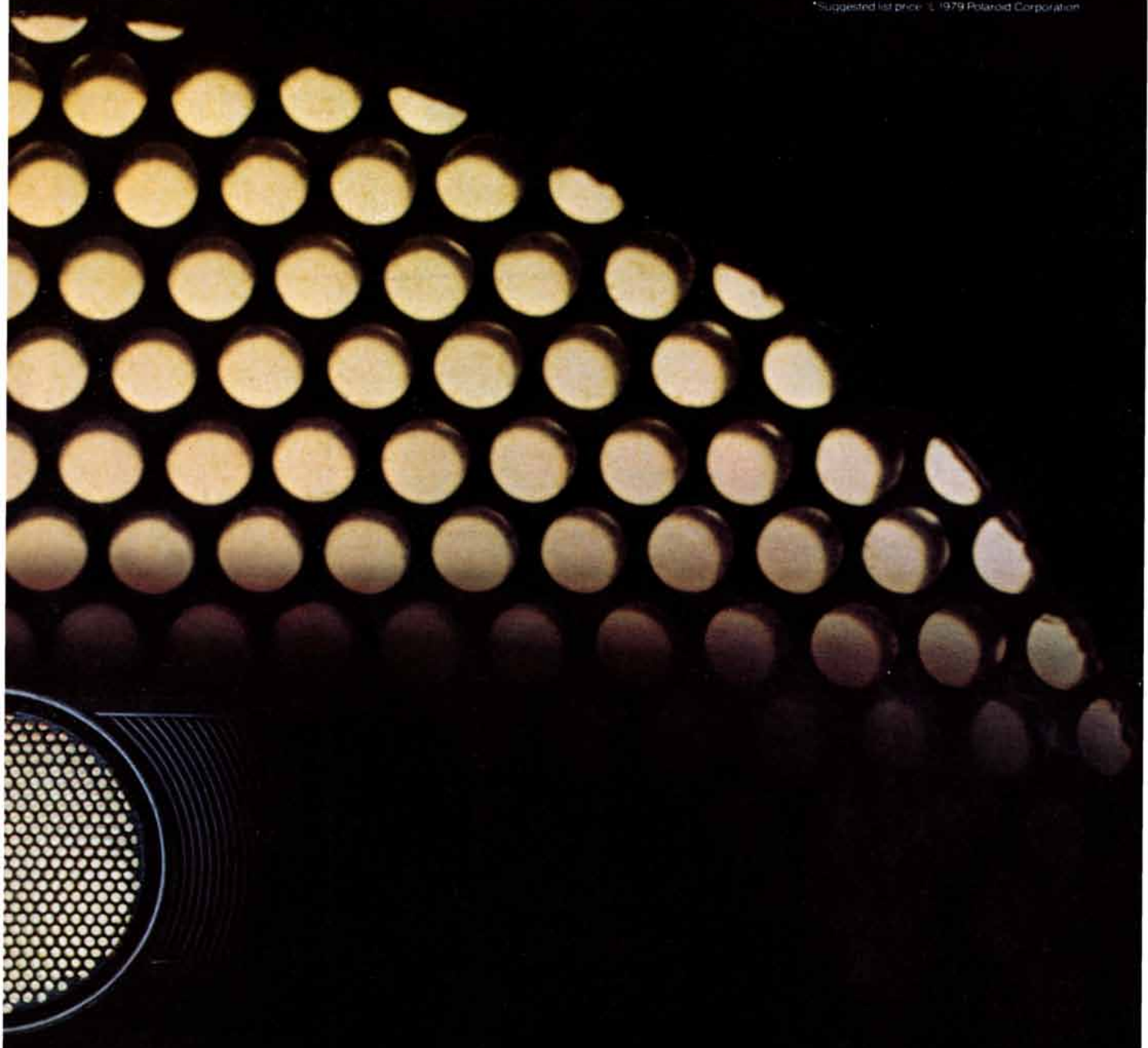
Polaroid's SX-70 Sonar OneStep



Now the world's finest instant camera gives you sharp, precisely-focused pictures every time – at the touch of one button.

Polaroid's SX-70 Sonar OneStep camera automatically focuses itself – even in total darkness – with sonar.

You simply press one button. Sonar automatically measures the exact distance to your subject. The 4-element glass lens rotates to the precise focus – from as close as 10.4 inches to infinity. All within a split second. And simultaneously, the single-lens reflex viewfinder shows you through the lens what your picture will be.



In fact. sonar can change focus so quickly. you can shoot as fast as every 1.5 seconds.

Polaroid's SX-70 Sonar OneStep camera sets all exposures automatically. too. Including automatic time exposures up to 14 seconds. And flash pictures up to 20 feet. Its built-in motor even hands you the developing picture.

There are two Polaroid* Sonar OneStep Land cameras: the folding SX-70 Sonar OneStep (described here). in genuine leather and brushed chrome finish. And the Pronto* Sonar OneStep. at $99.95*. No other cameras make precision photography so automatic.

Curie temperature are interchangeable.

Although the potential power of the renormalization group is widely acknowledged, its accomplishments so far are modest. The critical behavior of fluids and of ferromagnets has been worked out, but only for model systems, not for real substances. Although the model systems are greatly simplified versions of the real problem, in some instances even they had not been solved by any other method. Recent experimental studies of fluids near a critical point tend to confirm the results of the renormalization-group analysis.

The greatest success of the renormalization group is a precise solution of a problem in solid-state physics called the Kondo effect. The effect is observed when a nonmagnetic metal such as copper is contaminated with traces of a magnetic metal such as iron; at low temperature, anomalies appear in the electrical properties of the impure metal. The problem of explaining these anomalies inspired considerable work during the 1960's, perhaps not so much because the problem is intrinsically important as because it was considered an annoyance: a solution seemed tantalizingly close at hand. Actually the Kondo effect conceals a difficult problem of many energy scales. It is necessary to evaluate interactions of the magnetic impurities with the conduction electrons of the metal, which can take on any energy in a wide and essentially continuous band. A solution based on renormalization-group methods was finally presented in 1974 by Wilson.

A problem of compelling interest, on which Wilson and others have expended substantial effort, is the description of interactions between quarks, which are the presumed constituents of protons, neutrons and many other fundamental particles. A provisional theory of these interactions has already been devised; it is modeled on quantum electrodynamics and is called quantum chromodynamics. The theory is a useful predictive tool, however, only under a limited set of conditions, and there are many important properties of quarks and of the particles they compose that cannot yet be described with it. The difficulty again is one of treating many scales of energy or length. When quarks are examined at high energy and short distances (deep within the proton, say), they seem to be only weakly bound together; at low energy and long range, however, they cohere so strongly that it seems they cannot be separated at all. Quantum chromodynamics successfully accounts for the short-range behavior of the quarks, but no method has yet been found for calculating the long-range interaction.

The renormalization group is an obvious candidate method for extending quantum chromodynamics to all scales of length and energy. So far numerical results have not been obtained, but a hypothesis has been suggested. In quantum electrodynamics the separation of electric charges (the process called ionization) might be regarded as a phase transition, governed by a fixed point in an abstract space. For quarks the corresponding fixed point may not exist, or it may lie in a region accessible only with infinite energy. In either case the quarks would be confined permanently to a condensed phase, like water that can never boil no matter how much it is heated.

The Acquittal of Ptolemy

In October, 1977, this department carried an account of arguments advanced to show that Claudius Ptolemy, the Greek astronomer who flourished in Alexandria in the second century, was a fraud. The case was made by Robert R. Newton of the Applied Physics Labora-

tory of Johns Hopkins University in a book titled *The Crime of Claudius Ptolemy* (published by the Johns Hopkins University Press). Since that time historians of astronomy have studied the book and several have concluded that the charge of fraud is groundless.

Newton's accusation is scarcely a trivial one. Ptolemy is considered the greatest astronomer of antiquity. His '*E Mathematike Syntaxis* (*Mathematical Treatise*), renamed *Almagest* (*The Greatest*) by ninth-century Arab scholars, presents a comprehensive theory of planetary motions that was universally accepted for 1,400 years. On the basis of this theory Ptolemy developed an elaborate mathematical system for predicting the future positions of the planets. In the *Syntaxis* Ptolemy also compiled the most extensive star catalogue of antiquity. The *Syntaxis* contains numerous references to Ptolemy's astronomical forebears, particularly Hipparchus (who compiled the first star catalogue), and in many instances Ptolemy's writings provide the only source of information historians currently have about early Greek astronomy and Babylonian chronology. For example, they include the only surviving list of the length of the reigns of Babylonian kings.

In *The Crime of Claudius Ptolemy* Newton charges the astronomer with the systematic fabrication of the data that underlay his theory of planetary motions. That makes Ptolemy, writes Newton, "the most successful fraud in the history of science" and the author of a book that "has done more damage to astronomy than any other work ever written." The foundation of Newton's case is statistical analyses intended to demonstrate that the accuracy of certain astronomical observations Ptolemy says he made is so great that in some cases the odds are a billion to one against his having made them with the instruments he describes. On the other hand, where Ptolemy's observations are inaccurate according to modern theory Newton contends that he should have been able to make more accurate ones with the instruments he says he used. Here too Newton calculates the odds against Ptolemy's having made such errors; finding odds as great as 10^{92} to 1, Newton concludes that the observations are fraudulent.

In an article to appear in *The American Scholar* Noel M. Swerdlow of the University of Chicago maintains that Newton's statistical analysis is invalid. To calculate the probability of a given event's occurring a certain number of times, Newton often relies on what is called the product rule: the probability of a single occurrence is multiplied by the number of trials. For example, the probability of rolling a one with a die is

one in six, or 1/6; the probability of rolling two successive ones is 1/6 times 1/6, or 1/36, and the probability of rolling three successive ones is 1/6 times 1/6 times 1/6, or 1/216.

To apply the product rule it is necessary to know the probability of a single occurrence and the total number of trials. Moreover, the rule works only when the occurrences are independent of one another. In other words, for the product rule to work, whether or not an event takes place on a certain trial cannot affect the probability of the event's taking place on any other trial. The preceding condition holds for the die: if a one shows up on a given roll, that does not affect the probability of a one coming up on any other roll.

According to Swerdlow, the product rule does not hold with ancient astronomical observations such as Ptolemy's because none of the requisite conditions for applying the rule are met. Newton simply has no way of determining the probability of any one of Ptolemy's observations' having a particular value, and he has no way of knowing the total number of observations from which Ptolemy selected those he reported. Furthermore, Newton cannot know whether or not the observations are independent of one another, as the product rule requires. Thus, concludes Swerdlow, Newton misapplies statistical methods when he first assumes an initial probability of one in 10, or 1/10, that a lunar-eclipse observation is not fraudulent and then multiplies it up to $1/10^{12}$ in order to calculate the probability that 12 lunar-eclipse observations could be anything but fraudulent.

Ptolemy's observations confirm his theoretical calculations with such accuracy that from the viewpoint of modern science they suggest fabrication. Victor E. Thoren of Indiana University and Owen J. Gingerich of Harvard University have independently pointed out that this kind of precision is completely understandable from the viewpoint of ancient science. The people who did astronomy in Ptolemy's time were mathematicians, who concerned themselves with proof, rigor, logic and consistency rather than with observational accuracy. It was the established ethic of ancient science to report only those observations that best confirmed theory and to disregard the rest. This custom accounts for the close agreement between Ptolemy's observations and his theoretical work. It was not until the development of statistical and probabilistic methods at the end of the 18th century that natural philosophers started to report all random observations, because only then did they have the techniques for interpreting large bodies of imprecise data by calculating means, medians,

standard deviations and the like. In short, according to Swerdlow, Thoren and Gingerich, Newton's case against Ptolemy collapses because it is based on faulty statistical analysis and a disregard of the methods of early astronomy.

The Eskimo and the Indian

The Eskimo have generally been viewed as the last Asiatic immigrants to the pre-Columbian New World, late coastwise voyagers rather than early hikers across the then dry Bering Strait. Two Canadian physical anthropologists have now presented genetic and morphological evidence indicating that it is otherwise. They link the Eskimo with Indians of the Athapaskan-language group, such as the Haida and Tlingit of the Pacific Northwest and the Navajo and the Apache of the Southwest.

Writing in *Current Anthropology,* Emöke J. E. Szathmary of McMaster University and Nancy S. Ossenberg of Queen's University report their analysis of genetic-marker data acquired from 12 populations, including Aleutian Islanders, Asiatic Eskimo, New World Eskimo and seven North American Indian tribes. When the two investigators constructed "tree" diagrams indicating the comparative genetic distance between the 12 populations, the New World Eskimo and Indians of the Athapaskan-language group were the most closely related. The Asiatic Eskimo and the Aleuts, although closely related to each other, were about equally remote from the New World Eskimo and the seven Indian tribes. When a similar tree of cranial measurements was constructed, based on a sample of more than 800 skulls, the Asiatic Eskimo again proved to be remote from both the New World Eskimo and the North American Indians. The Aleutian Islanders, however, showed cranial features closer to the Athapaskan-language group of Indians than to either the Asiatic or the New World Eskimo.

Szathmary and Ossenberg suggest that the Indians and the Eskimo crossed to the New World together via the Bering land bridge some 10,000 years ago. They may then have been a single population or have begun to differentiate to some degree. The major separation, however, would have begun with the onset of climatic changes some 8,000 years ago, when the tundra of the Alaskan interior was gradually replaced by forest. The Eskimo withdrew to the coast; the Indians adapted to the forest. This geographical barrier to gene flow, together with evolving cultural barriers, is responsible for the modest genetic and language differences between today's Eskimo and Indian.

The Spectrum of Atomic Hydrogen

For almost a century light emitted by the simplest of atoms has been the chief experimental basis for theories of the structure of matter. Exploration of the hydrogen spectrum continues, now aided by lasers

by Theodor W. Hänsch, Arthur L. Schawlow and George W. Series

The spectrum of the hydrogen atom has proved to be the Rosetta stone of modern physics: once this pattern of lines had been deciphered much else could also be understood. Most notably, it was largely the effort to explain the spectrum of light emitted by the hydrogen atom that inspired the laws of quantum mechanics. Those laws have since been found to apply not only to the hydrogen atom but also to other atoms, to molecules and to matter in bulk. They are the ultimate foundation of modern chemistry, of solid-state physics and even of applied sciences such as electronics.

The central position of the hydrogen atom in the history of 20th-century physics might seem to suggest that the spectrum has long been known in all its details. That is not so. Only in the past few years have some of the subtler features of the spectrum been resolved, and many others have not yet been observed directly. Measuring the positions of the spectral lines remains today a significant test of the predictions of the quantum theory.

The most recent advances in the analysis of the spectrum can be attributed to a new tool of spectroscopy: the laser. Because the finest details of the spectrum are closely spaced they can be distinguished only by light that is highly monochromatic, or confined to a narrow range of wavelengths. The laser is a source of such light. As a result of some ingenious laboratory technology it has become a spectroscopic instrument of unprecedented resolution.

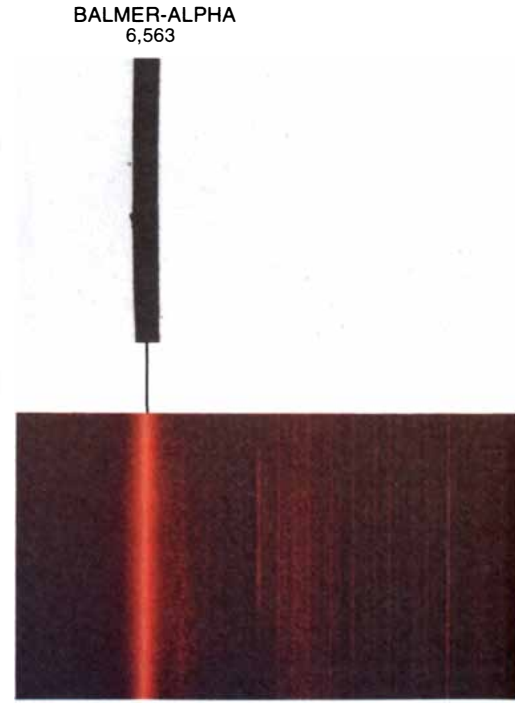
The Line Spectrum

When light from a hot filament is dispersed according to wavelength by a prism or a diffraction grating, the result is a continuous fan of colors, but the spectrum from a pure, rarefied gas of atoms or molecules consists of discrete lines. If the spectrum is recorded from light emitted by the gas, it appears as a sequence of bright lines against a dark background. If light is passed through the gas, discrete wavelengths are absorbed, giving rise to dark lines on a bright background.

Hydrogen is the simplest of atoms, being made up of a single electron and a nucleus that consists of a single proton, and so it can be expected to have the simplest spectrum. The spectrum is not, however, an easy one to record. The most prominent line was detected in 1853 by Anders Jonas Ångström. (The common unit for measuring wavelengths of light is named for Ångström; one angstrom unit is equal to 10^{-8} centimeter.) In the next two decades three more lines were observed, but the first extended series of atomic-hydrogen lines was found not in the laboratory but in the spectra of stars. In 1881, working with the first photographs of stellar spectra, Sir William Huggins identified 10 lines as being emissions of atomic hydrogen.

It may seem surprising that lines of the hydrogen spectrum were seen in astronomical observations before they were seen in terrestrial experiments. The difficulty in measuring the spectrum in the laboratory is not in detecting the lines but in preparing pure atomic hydrogen. Ordinary hydrogen gas consists of diatomic molecules (H_2), which have a spectrum that is much more complicated than the spectrum of the isolated hydrogen atom. Splitting the molecules requires more energy than can be supplied by most thermal excitations, such as a flame. One device that furnishes the energy very effectively is the gas-discharge tube, where electrons accelerated by an applied voltage dissociate the molecules. Over the years means have been discovered for enhancing the atomic spectrum and suppressing the molecular one.

The spectral lines detected by Huggins range in wavelength from the red portion of the visible spectrum to the near ultraviolet. The first two lines are quite far apart, but the subsequent ones come at smoothly decreasing intervals, and those at the shortest wavelengths are bunched closely together. In 1885, working from the astronomical measurements, Johann Jakob Balmer found that he could account for the positions of all the known lines by applying a simple empirical formula. The entire set of lines has since come to be known as the Balmer series. Another group of lines, the Lyman series, lies in the far ultraviolet, and there are other series at longer wavelengths. Within each series the individual lines are designated by Greek letters, starting with the line of longest wavelength. Thus the bright red line first seen by Ångström is the Balmer-alpha line, the next is the Balmer-beta line and so on. Because of the prominence of the Balmer-alpha line it is sometimes called



PATTERN OF DISTINCT LINES is formed when the light emitted by hydrogen atoms is dispersed according to wavelength. The lines are those of the Balmer series, and their approximate wavelengths are given in angstrom units (one angstrom is equal to 10^{-8} centimeter). The color photograph was made by Jon Brenneis and one of the authors (Schawlow)

simply the hydrogen-alpha, or H-alpha, line.

In 1889 Johannes Rydberg discovered that the line spectra of many elements, most notably those of the alkali metals, could be fitted by a single empirical formula. A later version of the formula, which is equivalent to Rydberg's equation but more explicit, has the form

$$\frac{1}{\lambda} = R\left[\frac{1}{(m+b)^2} - \frac{1}{(n+c)^2}\right].$$

Here λ (the Greek letter lambda) is the wavelength of a particular line in an atomic spectrum, m and n are numbers that take on successive integer values (1, 2, 3 and so on) and R, b and c are constants. The values of b and c depend on what series of lines is being measured, but R is the same for all lines of all the elements. R is now called the Rydberg constant; if the wavelength is expressed in meters, R has the value 1.097×10^7 and the dimensions of reciprocal meters. After Rydberg had checked his formula against a number of spectra he wrote: "I had just finished testing various forms of the function when I heard of Mr. Balmer's communication on the spectral lines of atomic hydrogen. I was delighted to find that his formula is a special case of mine, with the same value of R, and with $c = 0$." Hence for the hydrogen atom the equation has a simpler form:

$$\frac{1}{\lambda} = R\left[\frac{1}{m^2} - \frac{1}{n^2}\right].$$

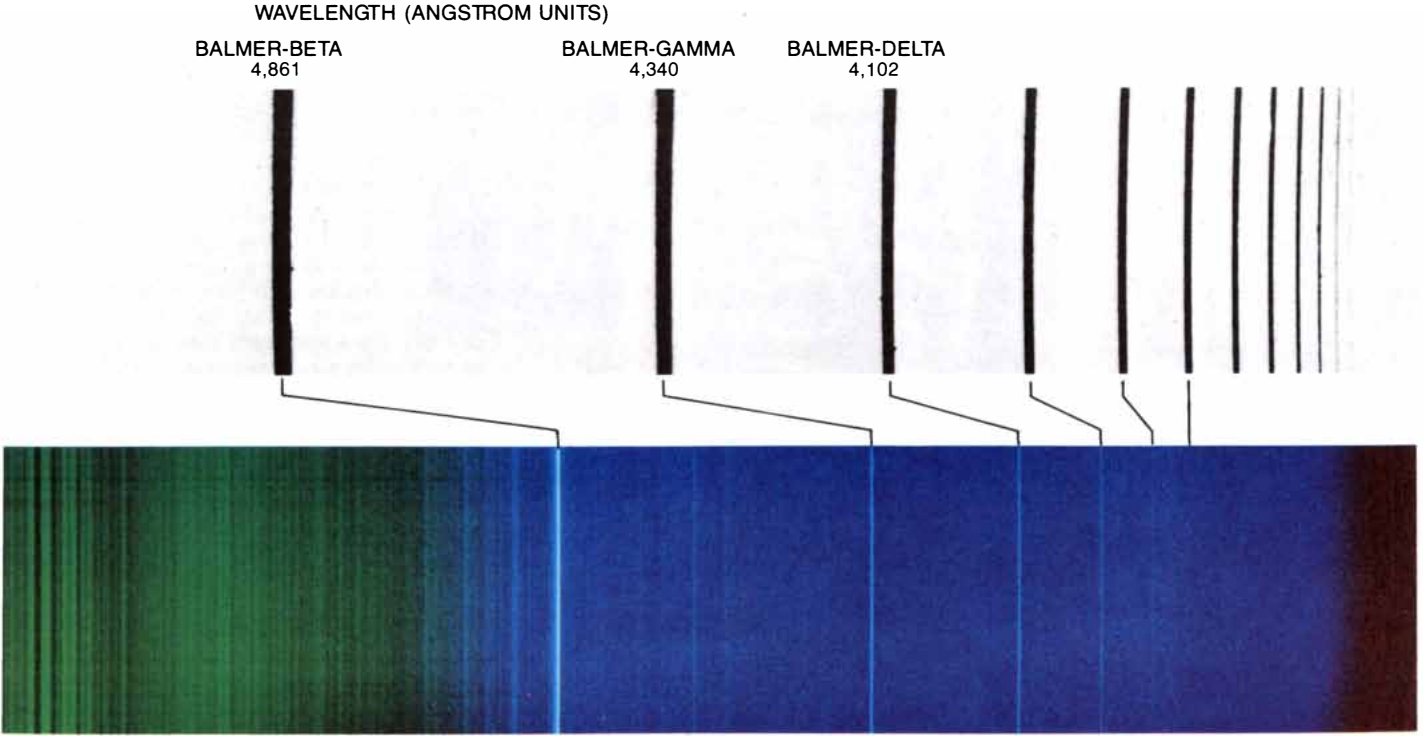
When the appropriate values of m and n are substituted in this formula, it yields the wavelengths of all the lines in the hydrogen spectrum.

Atomic Structure

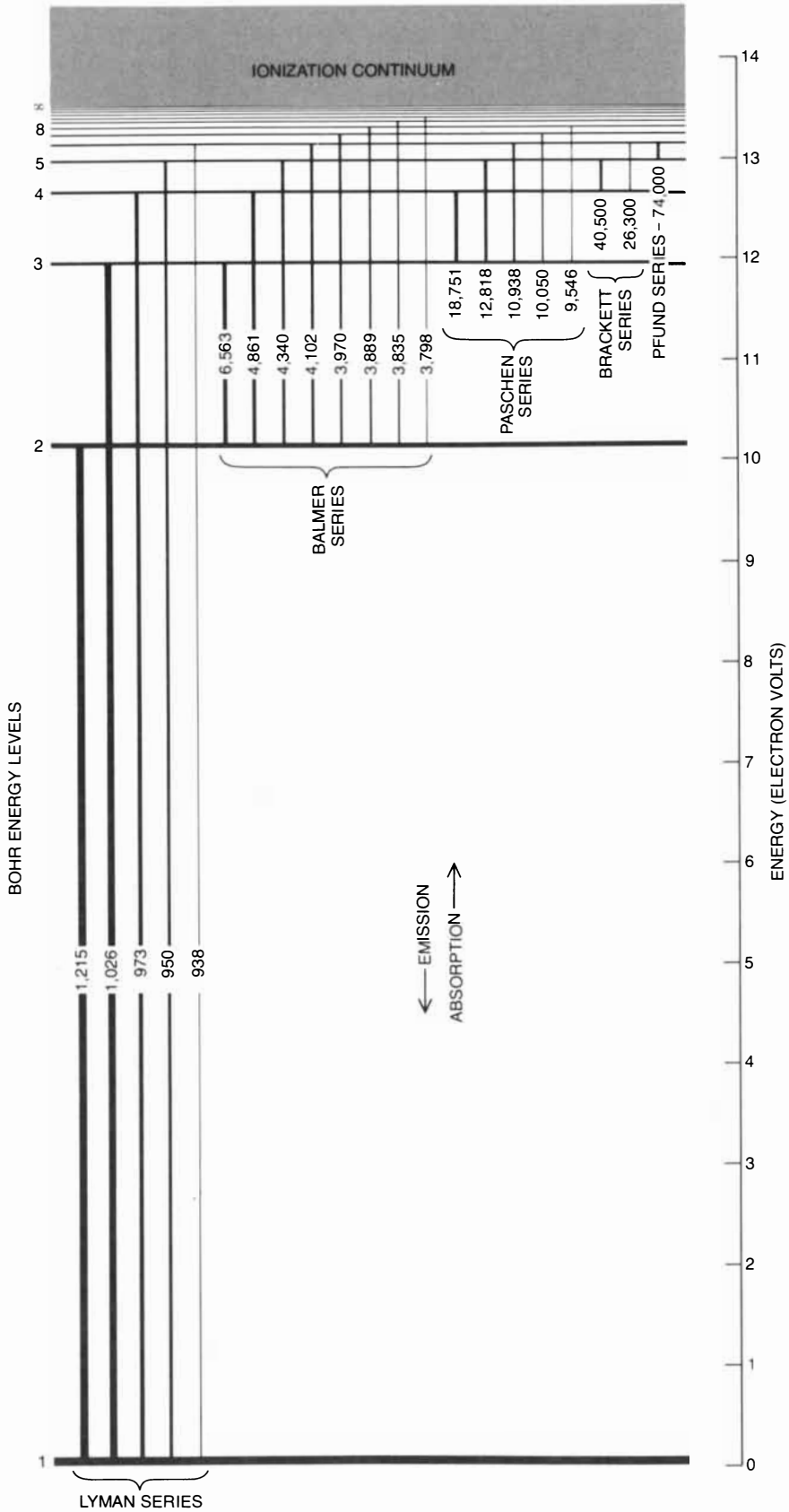
In 1912 Niels Bohr spent most of the year at the University of Manchester working in the laboratory of Ernest Rutherford, who had just made a vital contribution to the understanding of atomic structure. By then the Balmer and the Rydberg formulas were well known to spectroscopists, and they had been thoroughly tested, but Bohr had never heard of either of them. When he was told of Balmer's equation, he was able to derive it almost immediately from the properties of the hydrogen atom. Abraham Pais and T. D. Lee have reported that Bohr was asked late in his life how he could have escaped knowing of the Balmer formula. He explained that in those days most physicists regarded atomic spectra as being so complicated they were hardly a part of fundamental physics. They were thought of, he said, as being like the notes of a piano, whose tones depend in a complicated way on the structure of the instrument as a whole.

The reference to musical notes was not an idle one. In the 19th century many investigators had speculated that the light emitted by atoms might be understood by some analogy to the modes of vibration of a solid body. The presence of multiple lines in each series could then be explained by assuming that the first line was the fundamental mode and the rest were overtones, or harmonics. No optical spectrum could be resolved into a fundamental frequency and its overtones, however, and so more direct examination of the atom itself was undertaken.

Rutherford's contribution was the demonstration that the atom is not a solid body but rather consists of a small, dense, positively charged nucleus and the requisite number of negatively charged electrons to form a stable and electrically neutral system. Bohr undertook to explain the spectra of atoms in the context of this model. In doing so he risked the ridicule of his contemporaries by hypothesizing that within the atom the established laws of physics do not apply. Those laws predicted that any bound electron would radiate away all its energy and fall into the nucleus. Bohr



in the latter's laboratory at Stanford University. The black-and-white spectrum was recorded in 1927 by Gerhard Herzberg, who was then at the Darmstadt Technical University in Germany. The intervals between the lines are different in the two photographs because Brenneis and Schawlow dispersed the light with a diffraction grating, whereas Herzberg employed a prism; the prism introduces a distortion because the refractive index of glass is a nonlinear function of wavelength. In both cases the source of the spectrum was a gas-discharge tube in which hydrogen molecules (H_2) are dissociated into atoms by an electric potential. In the color photograph the yellow and green bands and the continuous blue background at the right are not part of the atomic-hydrogen spectrum but result from molecular contaminants in the discharge tube. Herzberg eliminated such contamination by cooling the tube in liquid nitrogen. Lines to the right of Balmer-delta, which appear white or light blue in the color photograph, are actually in the ultraviolet region of the electromagnetic spectrum.



TRANSITIONS BETWEEN STATES of the hydrogen atom give rise to the line spectrum. Each state is designated by an integer, called the principal quantum number of the corresponding Bohr energy level. When an atom changes from one state to another, the difference in energy appears as a quantum of radiation. The energy of the quantum is directly proportional to the frequency of the radiation and inversely proportional to the wavelength. Absorption of radiation stimulates a transition to a state of higher energy; an atom falling to a state of lower energy emits radiation. The spectrum is organized into series of lines that share a lower level. Wavelengths are given in angstroms; the relative intensity of the lines is indicated by thickness.

proposed that an atom could exist in a state where the energy of the electron was not dissipated: a stationary state. Light is emitted or absorbed, he suggested, only when an electron moves from one stationary state to another. The frequency of the light is determined by the difference in energy between the two states, according to the quantum principle introduced by Max Planck. The frequency is found by dividing the energy difference by Planck's constant, which is designated h.

In Bohr's model the Balmer and the Rydberg formulas have an obvious physical interpretation. The expression $(1/m^2 - 1/n^2)$ is proportional to the difference in energy between two states of the atom; the integers m and n label the states themselves. Bohr did more than just explain this empirical equation, however; he went on to evaluate the Rydberg constant in terms of three fundamental quantities, the electric charge of the electron, e, the mass of the electron, m, and Planck's constant, h. If for the sake of simplicity the nucleus of an atom is assumed to be infinitely massive, then the Rydberg constant is given by the equation

$$R = \frac{2\pi^2 me^4}{h^2} .$$

Later refinements have complicated Rydberg's empirical formula for the wavelengths of spectral lines, and so the Rydberg constant is now defined as this combination of m, e and h.

Bohr imagined that the electrons in an atom follow circular orbits. In calculating the energy of the stationary states he began with the assumption that for orbits of very large diameter the frequency of the emitted light should correspond to the frequency of the orbiting electron, which was a prediction of classical physics. This "correspondence principle" led to the intriguing conclusion that for a stationary state designated by the integer n the angular momentum of the orbiting electron is equal to $nh/2\pi$. Hence the angular momentum, like the energy, can change only in discrete steps. This "quantization" of the angular momentum is of fundamental significance: without it the atom could radiate at any frequency and the state would no longer be stationary.

Quantum Mechanics

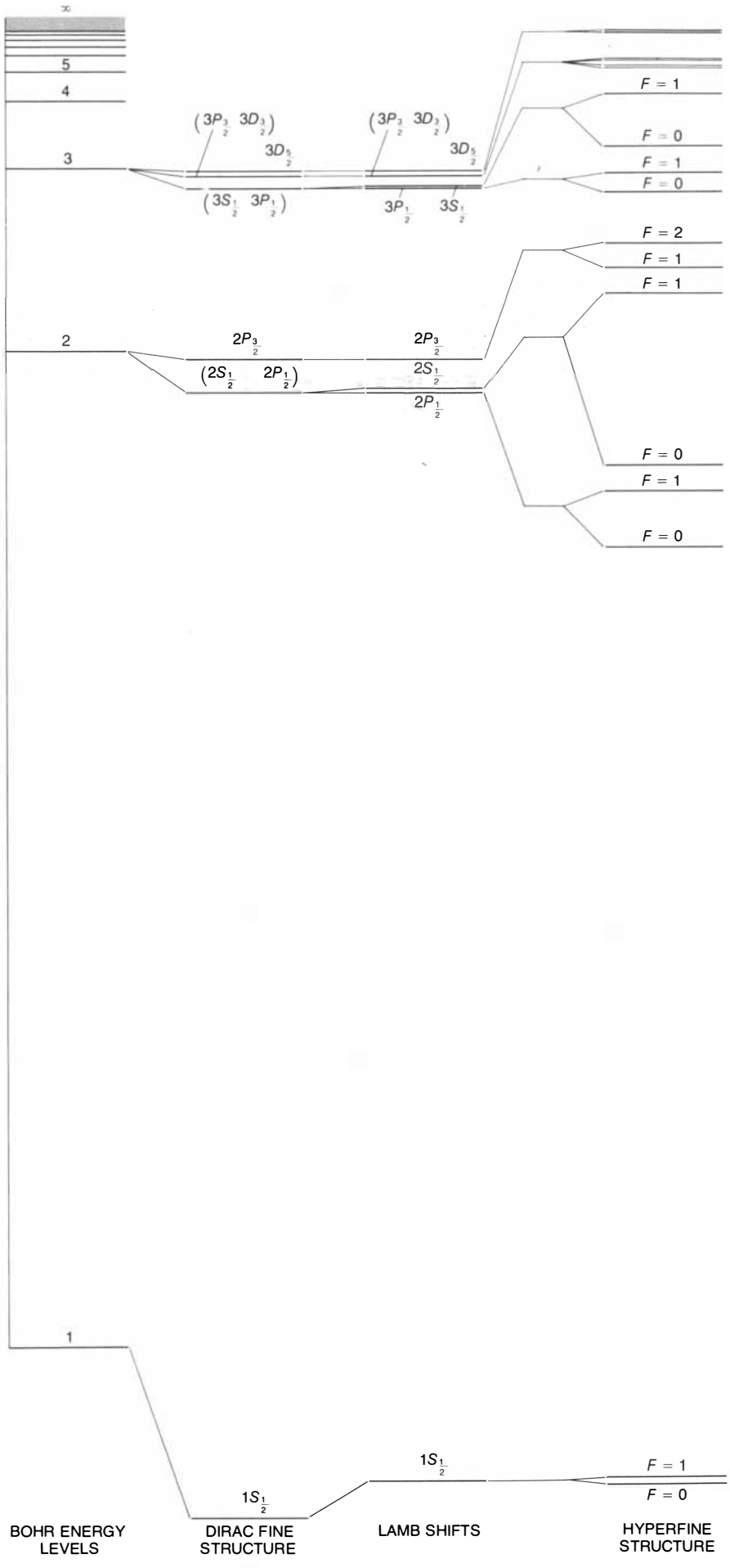
The Bohr atom was a great conceptual advance over earlier theories, but it soon proved inadequate to explain the observed features of atomic spectra, even those of the simplest spectrum, that of hydrogen. Indeed, one detail of the hydrogen spectrum that the Bohr theory could not account for had been known for 20 years when the model was constructed. In 1892 A. A. Michelson had employed the interferometer he had

invented to examine the shapes of individual spectral lines, that is, their variation in intensity as a function of wavelength. He had found that the Balmer-alpha line is not a single line at all but rather has two components, separated in wavelength by .14 angstrom.

Bohr suggested that the splitting of the Balmer-alpha line might be explained if the electron's orbit in the hydrogen atom is not circular but elliptical. A correction to the orbital motion required by the special theory of relativity would then split each stationary state into a group of states. A more comprehensive discussion of this idea was subsequently presented by Arnold Sommerfeld, who deduced the observed splitting from the calculated properties of the orbits. In his calculation he introduced an important dimensionless number called the fine-structure constant, equal to $2\pi e^2/hc$. The numerical value of this constant, about 1/137, has been a source of speculation for physicists ever since.

A more satisfactory treatment of the line splitting was not possible until quantum mechanics was introduced in the 1920's. Two formulations of quantum mechanics were devised, one by Erwin Schrödinger and the other by Werner Heisenberg and Max Born; they were soon found to be mathematically equivalent. An essential idea in both formulations was that the motion of an electron cannot be defined precisely but must be described in terms of probabilities. Both theories predicted the Bohr energy levels of the hydrogen atom, and hence the same wavelengths of the spectral lines, and the predictions could be extended to more complicated atomic systems. Again relativistic corrections had to be introduced in order to explain the splitting of the hydrogen lines.

By the 1920's, however, better measurements of the shape of lines in the hydrogen spectrum were available, and it soon became plain that a further adjustment to the theory was needed. The adjustment was made when it was rec-



SPLITTING OF ENERGY LEVELS results mainly from relativistic and magnetic interactions related to angular momenta in the atom. The single electron of a hydrogen atom can have both orbital and spin angular momentum, and the various possible combinations of these quantities generally have different energies. As a result the Bohr energy levels are split into components called the fine structure of the levels. A subtle effect called the Lamb shift displaces certain of the fine-structure components, creating additional splittings. Finally, the interaction of the electron's magnetic moment with that of the nucleus gives rise to a hyperfine structure. The fine structure and the Lamb shifts are shown here at a larger scale than the Bohr levels, and the hyperfine splittings of the second and third Bohr levels are shown at still greater magnification.

NO RUM REFLECTS PUERTO RICO LIKE RONRICO.

Puerto Rico is the Rum Island, the world's foremost rum-producing region. And Ronrico is *the* rum—authentic Puerto Rican rum since 1860. Ronrico's smooth, light taste has been the pride of six generations of Puerto Rican rum masters. One sip will tell you why.

RONRICO: AUTHENTIC RUM OF PUERTO RICO.

General Wine & Spirits Co., N.Y.C.

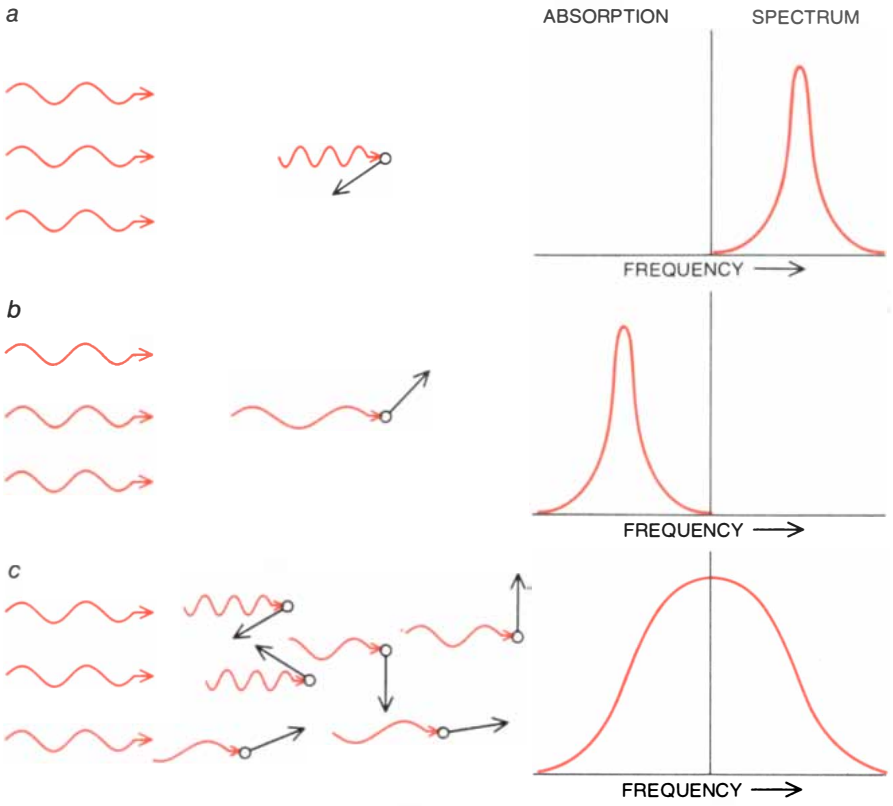
ognized that the electron can acquire angular momentum not only by orbiting the nucleus but also by spinning on its axis. The concept of electron spin was introduced by George Uhlenbeck and Samuel A. Goudsmit to explain measurements of the spectra of alkali-metal atoms. By including electron spin in the theory it was possible to predict with reasonable accuracy not only the wavelengths of the spectral lines but also their intensities; that had not been true for the Bohr-Sommerfeld theory. In 1928 the concept of spin for the electron—and also for the proton and the neutron—was given a more secure foundation when P. A. M. Dirac found a new way of constructing a theory of quantum mechanics that is also consistent with the special theory of relativity. In the Dirac theory electron spin appears as a natural consequence of the basic equations rather than as a special postulate. The magnitude of the spin angular momentum is $1/2\ h/2\pi$.

Fine Structure

The relativistic quantum mechanics of Dirac predicts the structure of the hydrogen spectrum in great detail. As in earlier theories each basic energy level is designated by an integer, n, which is called the principal quantum number. For the lowest energy level (the ground state of the atom) n is equal to 1, for the first excited state n is equal to 2 and so on. These integers are the same ones that appear in Rydberg's formula for the spectral lines.

The basic lines of the spectrum—ignoring for the moment any splitting into finer components—are generated by transitions between states that have different principal quantum numbers. The Balmer-alpha line, for example, results from transitions between the states $n = 2$ and $n = 3$. If a hydrogen atom falls from $n = 3$ to $n = 2$, it emits light at the wavelength of the Balmer-alpha line; if the atom then absorbs a quantum of light at the same wavelength, it makes the opposite transition, from $n = 2$ to $n = 3$. All the lines in a series share a common lower state. The transitions that give rise to the Balmer series connect states with principal quantum numbers of 2 and 3, 2 and 4, 2 and 5 and so on. In the Lyman series all the transitions begin at the state $n = 1$.

The basic lines of the hydrogen spectrum reflect the dominant interaction in the hydrogen atom: the electrostatic attraction between the nucleus and the electron. The splitting of each line reflects subsidiary interactions, relativistic effects and magnetic effects. The magnetic effects come about because the spinning electron, acting as a magnet, is subjected to a magnetic field attributable to the effective motion of the charged nucleus around the electron. This splitting is called the fine structure of the spectrum. Because the nucleus itself also has spin and acts as a magnet, it interacts with the magnetic field arising from the motion of the electron's charge and also from the proximity of the electron's magnetic field. These magnetic interactions are on a still smaller scale, and the corresponding splittings of the spectral lines are called the hyperfine structure.



DOPPLER BROADENING obscures many details of the hydrogen spectrum by smearing each narrow component over a band of wavelengths. Because the atoms of a gas have random velocities, at any given moment some are moving toward a source of light and others are moving away from it. For an atom moving toward the source (*a*) the light appears to have a higher frequency (or shorter wavelength) than it does in the laboratory frame of reference. For an atom moving away (*b*) the frequency is shifted down. As a result of these Doppler effects the measured frequency of a spectral component is different for atoms that have different velocities. The absorption profile for all the atoms in the gas (*c*) is not a sharp peak but a broad curve.

A system of classification has been developed for the fine and hyperfine energy levels, which specifies the magnitudes of the various angular momenta (orbital, electron-spin, nuclear-spin and total) in a stationary state. The principal quantum number, which corresponds to the Bohr energy level, is given by one of the integers in the sequence beginning 1, 2, 3.... To this number is appended a letter, such as S, P, D or F, which is related to the orbital angular momentum; the letters stand for integers in the series beginning 0, 1, 2, 3.... A third symbol, which for the hydrogen atom is always a fraction such as 1/2, 3/2 or 5/2, is related to the sum of the orbital and spin angular momenta. If an analysis is subtle enough to reveal the hyperfine structure (an interaction that depends on the magnetic properties of the nucleus), a fourth label, represented by the symbol F, is also needed. It is related to the sum of the orbital, electron-spin and nuclear-spin momenta. Generally, stationary states that differ in any label have different energies, although that is not always the case.

The fine structure splits the $n = 2$ state of the hydrogen atom into three sublevels and the $n = 3$ state into five sublevels. A simple tabulation of the possible combinations of these levels implies that there could be 15 components of the Balmer-alpha line. Actually several of these transitions are forbidden by conservation laws, and the Dirac theory predicts that the line should have seven components. At the time Dirac proposed his theory experimenters could clearly recognize only two components, although in 1925 G. Hansen detected a third component as a broadening on the side of one of the others.

The factor limiting the resolution of the lines was (and still is) Doppler broadening caused by the thermal motions of the atoms. Because of the Doppler effect the light emitted by an atom appears to be shifted to a higher frequency if the atom is approaching the observer and to a lower frequency if the atom is receding. Atoms in an ordinary light source move in all directions, and so the emitted light is broadened in wavelength by the ratio v/c, where v is the average speed of the atoms and c is the velocity of light. The broadening

is exceptionally troublesome for hydrogen, whose atoms are lighter than any others and therefore move faster. At room temperature the fractional line width is about 12 parts per million. The speed of the atoms, and hence the line width, is proportional to the square root of the absolute temperature. The line width can be reduced by a factor of three or four by cooling the hydrogen discharge lamp from room temperature (300 degrees Kelvin) to the temperature of liquid hydrogen (20 degrees K.).

The Lamb Shift

In the 1930's several investigators tried to resolve the fine structure of hydrogen more completely in order to test the predictions of the Dirac theory. By then it was possible to examine the spectrum of deuterium, the hydrogen isotope with a mass twice that of ordinary hydrogen. The width of a spectral line varies inversely as the square root of the atomic mass, and so the width is reduced in a deuterium spectrum by a factor of about 1.4. Even so, no additional components of the Balmer-alpha line were resolved, and the third component was still not completely separated from its stronger neighbor.

In the course of these investigations, however, some observers reported small discrepancies between the measured spectra and the predictions of the Dirac theory. Of particular interest was the pair of fine-structure levels $2S_{1/2}$ and $2P_{1/2}$. These levels differ in orbital angular momentum and therefore in the average geometrical distribution of the electron, but they have the same total angular momentum and the Dirac theory predicts they should lie at the same energy. Thus although they are distinct states, they should not give rise to any splitting in the spectrum. Simon Pasternack interpreted experiments carried out by Robley C. Williams and by William V. Houston and his colleagues as showing that the $2S_{1/2}$ and $2P_{1/2}$ levels do not exactly coincide. Others, however, disagreed.

The question was settled in 1947 in a brilliant experiment conducted by Willis E. Lamb, Jr., and Robert C. Retherford. They found that the $2S_{1/2}$ level was shifted upward with respect to the $2P_{1/2}$ level by about 1,060 megahertz, or millions of cycles per second. Compared with the frequency of the Balmer-alpha line, which is about 450,000 gigahertz (billions of cycles per second), that is a shift in the position of the $2S_{1/2}$ component by a factor of little more than two parts per million.

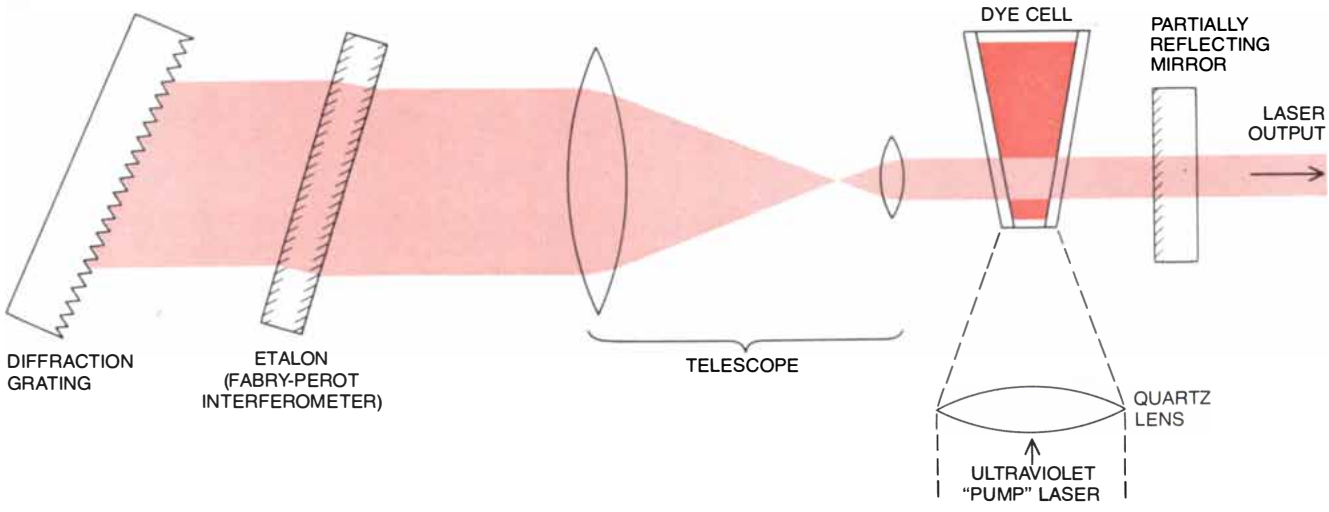
Lamb and Retherford did not attempt to resolve the two components in the optical spectrum. Instead they employed radio-frequency energy to stimulate a direct transition between the $2P_{1/2}$ and the $2S_{1/2}$ levels. The frequency of the radio waves is 500,000 times lower than the frequency of the Balmer-alpha line; since the Doppler broadening of the line is reduced by the same factor, it could be neglected entirely. This small splitting of the energy levels is now called the Lamb shift.

The need to supply an explanation for the Lamb shift inspired another fundamental revision of physical theory, introduced by Richard P. Feynman, Julian Schwinger and Sin-itiro Tomonaga. Their point of departure was the novel idea that the electrons observed in experiments may differ in certain properties, such as mass and charge, from the hypothetical "bare" electron that is imagined to exist in an empty universe. This difference, which had previously been ignored, can alter the properties of an electron in a bound system. The theory that results from these ideas is called quantum electrodynamics.

In quantum electrodynamics the Lamb shift arises from an interaction between the electron and "zero-point fluctuations" of the electromagnetic field. Such fluctuations, which are always present, cause the electron to jitter about, so that it perceives the nucleus as being somewhat smeared out in space. The result is a seeming alteration of the force between the two particles when they are close together. Because the electron spends more time near the nucleus in the S state than it does in the P state, the energy of the S state is slightly increased.

This model of the hydrogen atom has a further complication. If the electron is regarded as a pointlike particle, it can respond to fluctuations of infinitely high frequency, which have infinite energy. The infinities are avoided, and so is the need to answer questions about the size and structure of the electron, by comparing an electron bound in an atom with a hypothetical bare electron. The small measured energy is found by subtraction. Such calculations are complex, but the predictions of the theory are exceedingly accurate. Indeed, quantum electrodynamics is the most precise theory ever devised, and no discrepancies with experiment have been found.

The fine-structure splitting of various energy levels of the hydrogen atom has been revealed in considerable detail through measurements at radio frequencies. At the same time improved spectroscopic methods have led to precise determinations of the Rydberg constant from measurements of the wavelength



DYE LASER emits light that can be confined to a narrow range of wavelengths but can also be tuned over a broad range. The laser medium is a fluorescent dye that emits light with a wide and essentially continuous spectrum when the dye is "pumped" by another laser. From this broad emission band a much narrower interval of wavelengths is selected for amplification by an optical cavity that is tuned to resonate at a particular frequency. Coarse tuning is provided by a diffraction grating, which disperses the light so that only selected wavelengths are reflected back into the dye cell. An etalon further narrows the bandwidth by passing only those wavelengths that interfere constructively during many internal reflections. A telescope in the cavity spreads the beam so that many grooves of the grating are illuminated. The monochromatic but tunable light has been employed in spectroscopic measurements that eliminate Doppler broadening.

THE NEW ERA.
SENSOR TOUCH CALCULATORS.

Sharp, which made the first solid-state electronic desk-top calculator, as well as the first hand-held, has become the first to eliminate push buttons with its exclusive Sensor Touch™ line of calculators. Now, numbers are entered at the touch of a finger. An electronic tone assures you that each entry has actually taken place.

To eliminate buttons and reduce thickness in calculators, Sharp invented a new way to bond onto Polymid film. Our 5808 Scientific Calculator is only 5mm, yet it does complex algebra, trigonometry, statistical analysis, and other advanced mathematics.

For general calculating, credit Sharp with developing the 8145, which is literally the size of a credit card: 3.8mm thin; Sensor Touch Entry Panel; power-saving 8-digit liquid crystal display (Sharp was the first to develop liquid crystal display for calculators); a Storage Computer™ that holds any number, even an English-to-metric conversion, for later use; Safe Guard™ circuitry that protects the numbers in memory even after the power goes off; Sharp's A.P.O.—Automatic Power Off—which does just that after 7 minutes of nonuse; and many other features.

Our 8130 model, called the "Super Thin Man" is the calculator that does everything you'd want and makes it all easier with an independently addressable 4-key memory for convenient storage and retrieval of totals.

Then there's our CT-600, a 24-function calculator with two memories, a world clock with an alarm, a 200-year calendar, and a stopwatch that counts up or down.

What, you might ask, will we think of next?

The world's thinnest calculator. The 8152...our newest model. 1.6mm. 8-digit. With Sensor Touch Entry, Storage Computer and Safe Guard circuitry.

The CT-610, a calculator with sophisticated microcomputer technology, marries a desk-top, 8-digit calculator with a digital synthesized tuning radio and an incredibly precise quartz clock.

The Sensor Touch CT-510. An 8-digit dual display calculator that offers you the time here and around the world, plus the date, an alarm, a stopwatch and more.

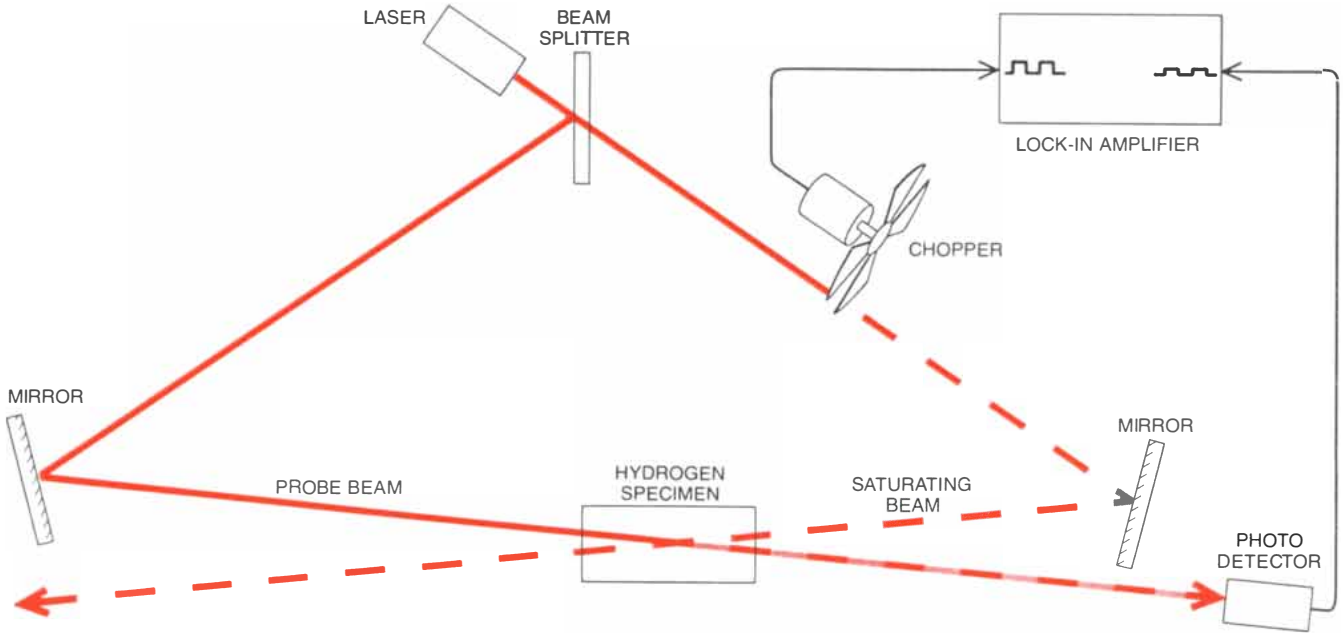
We could go on. The point is, whether you're looking for the most complex, vari-featured calculators or just the most basic, you've only one place to look:

Sharp. Where you can touch the new era.

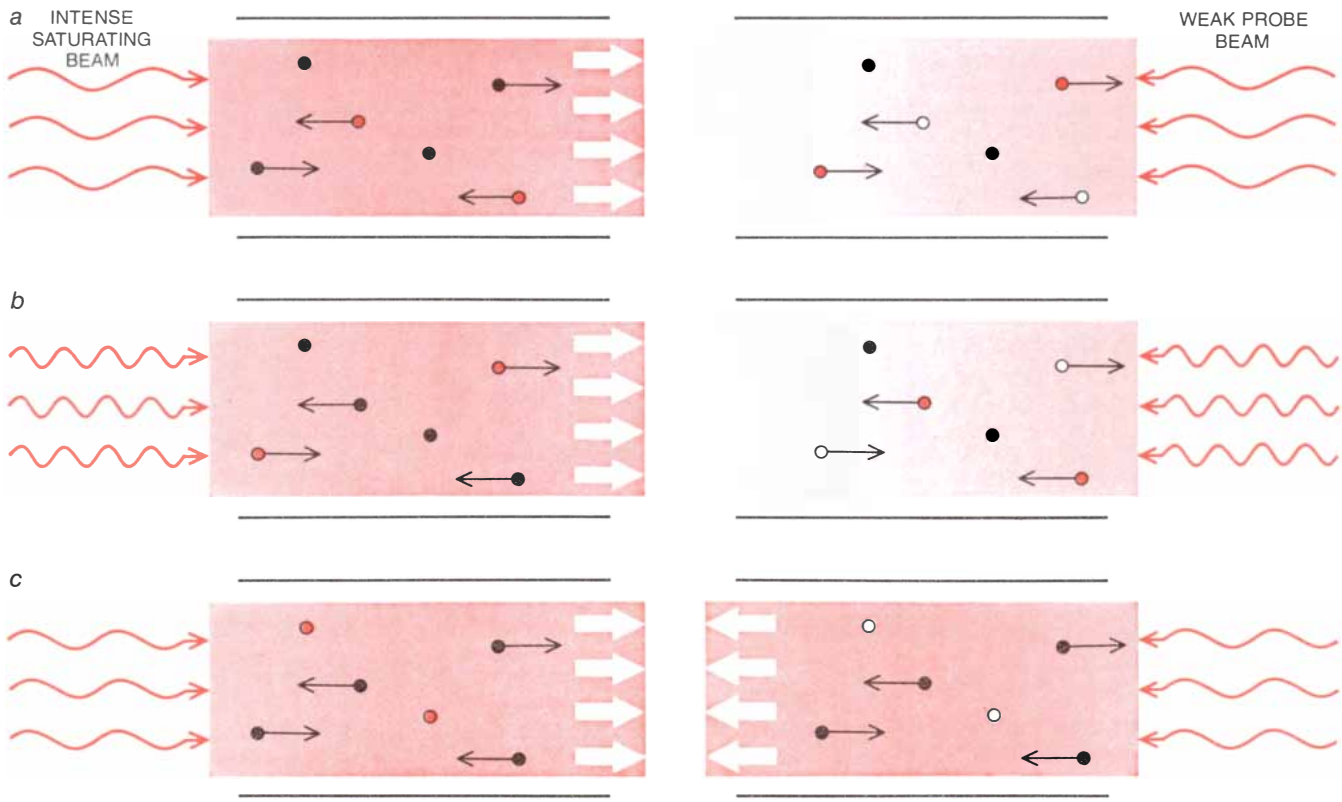
SENSOR TOUCH CALCULATORS

Sharp Electronics Corporation, Paramus, N.J. 07652

SHARP



SATURATION SPECTROSCOPY suppresses Doppler broadening by labeling a group of atoms that happen to have no component of motion along the optical axis. The light from a dye laser is split into an intense saturating beam and a weaker probe beam, which pass through a specimen of atomic hydrogen in opposite directions. The saturating beam is strong enough to reduce significantly the population of atoms in the state capable of absorbing the laser wavelength: the beam bleaches a path through the gas. The probe beam therefore encounters a smaller absorption and registers a higher intensity at the detector. The two beams can interact in this way, however, only when they are both absorbed by the same atoms in the gas, and that can happen only when they are both tuned to the wavelength of atoms that have no Doppler shift. In practice the saturating beam is interrupted by a mechanical "chopper," and enhancement in the transmission of the probe beam is detected by tuning the laser through a range of wavelengths and searching for a signal at the chopping frequency.



COUNTERPROPAGATING BEAMS from a tunable dye laser undergo opposite Doppler shifts. When the frequency of the saturating beam (measured in the laboratory frame of reference) is below that of some selected component of the spectrum (*a*), only atoms moving to the left can absorb the light, because their motion shifts the frequency into resonance with the atomic transition. The probe beam, with the same frequency, is absorbed only by atoms moving to the right, and so it is not affected by the saturating beam. When the laser frequency is above that of the spectral component (*b*), the saturating beam is absorbed by atoms moving to the right and the probe beam by atoms moving to the left, which are again different groups of atoms. Only when the laser frequency matches the atomic-transition frequency (*c*) do the two beams interact with the same atoms: those atoms that are effectively standing still. The probe beam then finds that the atoms that might have absorbed it are unable to do so because they have already absorbed radiation from the saturating beam. Transmission of the probe beam therefore increases. Only the component of motion that is directed along the optical axis is depicted.

of one component of the Balmer-alpha line. Bryan P. Kibble, William R. C. Rowley, Raymond E. Shawyer and one of us (Series) employed a light source cooled in liquid helium to determine the Rydberg constant to a precision of better than one part in 10 million. Even the Lamb shift was observed in 1948 by one of us (Series), and similar shifts were measured in ionized helium by one of us (Series), by Gerhard Herzberg and by Frederick L. Roesler. The accuracy of the optical experiments, however, does not approach that of the radio-frequency methods.

Width of Spectral Lines

Although these experiments have been highly successful, the hydrogen atom is such a fundamental testing ground for the laws of physics that it is important to probe deeper as new techniques become available. For now the most promising approach is to improve the resolution of measured spectra; it is resolution that limits the sharpness of observed spectral lines and thereby limits the precision with which their wavelengths can be determined.

The earliest spectroscopic studies of atoms were limited by the resolution of the instrument itself, which was simply a prism or a diffraction grating, a glass plate ruled with many closely spaced lines. The grating has the effect of dispersing light by reflecting different wavelengths through different angles. If the entire visible spectrum is dispersed through, say, 15 degrees, then some components of the Balmer-alpha line might be separated by less than .05 second of arc.

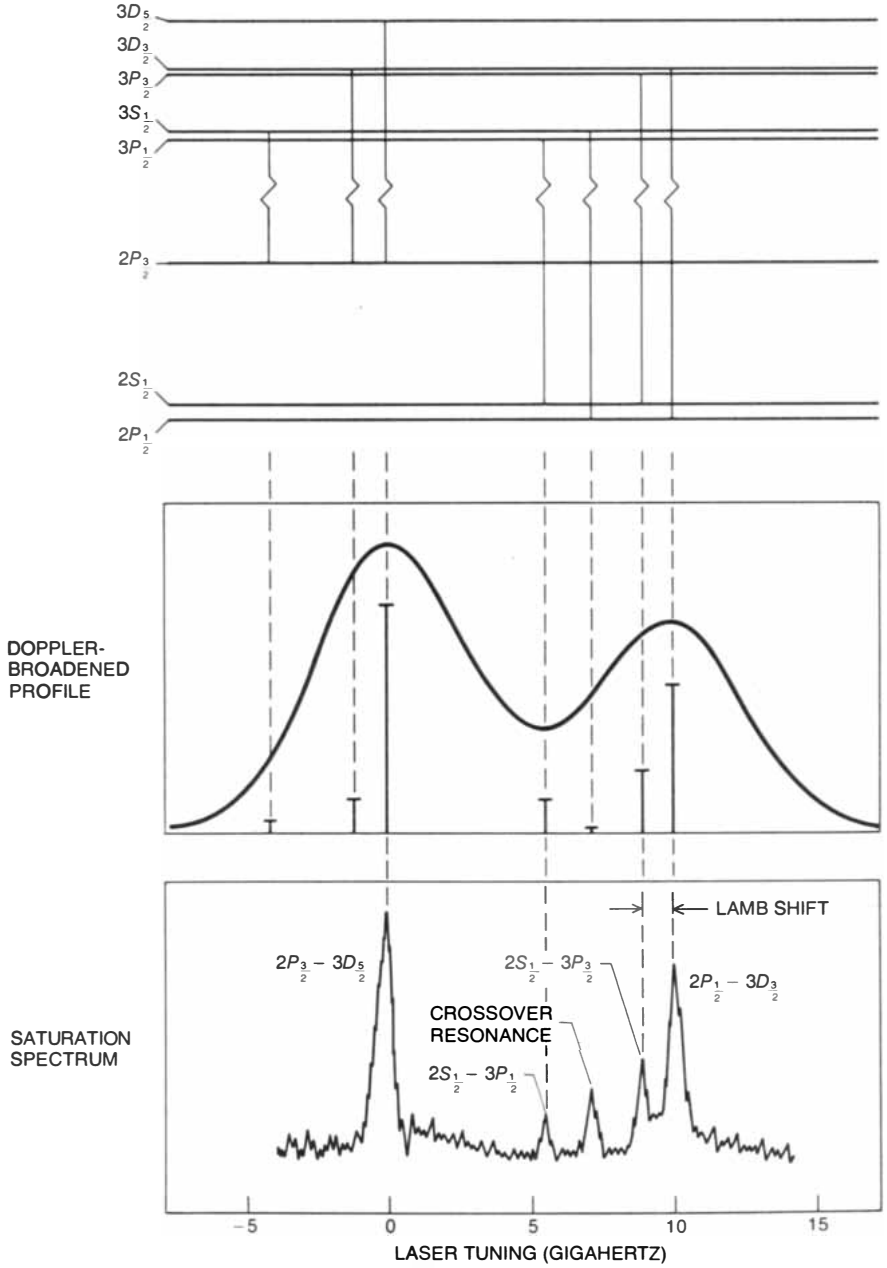
With the application of interferometers to the examination of spectral lines the intrinsic instrumental limitation was removed. The interferometer invented by Michelson and another type devised by the French opticians Charles Fabry and Alfred Perot are sensitive instruments for comparing two wavelengths. They are capable of distinguishing two sharp peaks that differ in frequency by one gigahertz or less. Interferometers fail to fully separate the several components of the Balmer lines not because the instruments are inadequate but because the components themselves are not sharp. Each component is distributed over a range of wavelengths that is often greater than the spacing between the components.

A part of the broadening is the "natural" width of the spectral line. Even under the best possible circumstances of observation no component of a spectrum can ever be absolutely monochromatic. This limitation stems from the fact that the "stationary states" of an atom are not truly stationary. Atoms do, in fact, radiate energy after a finite lapse of time. This is simply to say that excited states do eventually decay.

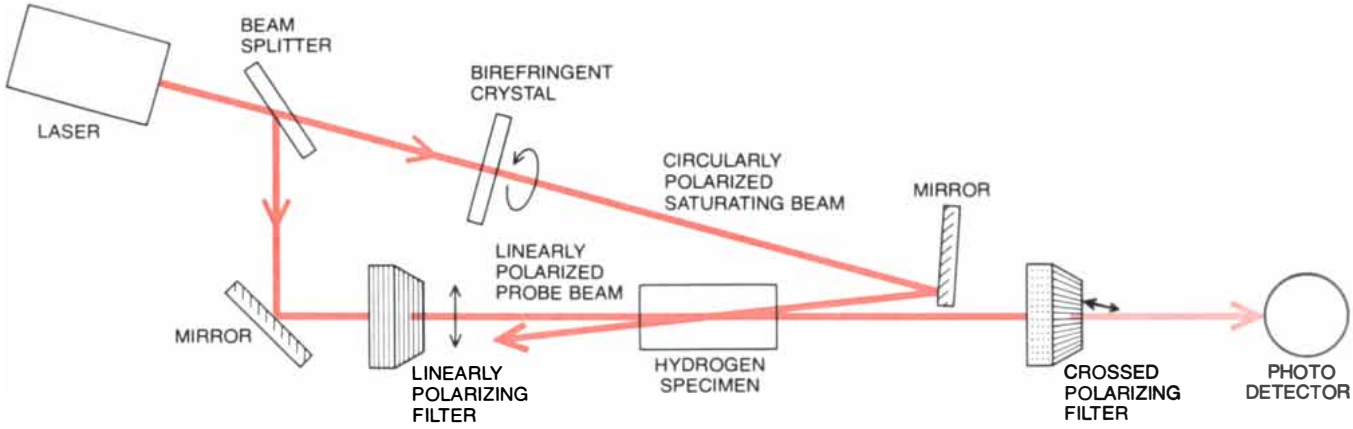
The natural line width sets an ultimate limit on spectroscopic resolution, but until recently the limit has seldom been approached. That is because the lines are broadened to a much greater extent by another effect: the Doppler shifting of frequencies as a result of the motion of the atoms.

One method of greatly reducing Doppler broadening is to measure the spectrum not of a gas but of a collimated atomic beam. Such a beam can be created by allowing the atoms of a gas to escape from a vessel through a series of pinhole openings. All the atoms that can pass through the holes will be moving in the same direction, and so light crossing the beam at right angles to the beam will show only small Doppler shifts. The number of atoms in the beam, however, declines as the collimation improves. Moreover, it is not always possible to prepare a beam of atoms in the necessary state. To study the Balmer spectrum of hydrogen by absorption, for example, the atoms must be in the $n = 2$ state; such atoms cannot be formed into beams dense enough for conventional spectroscopic measurements.

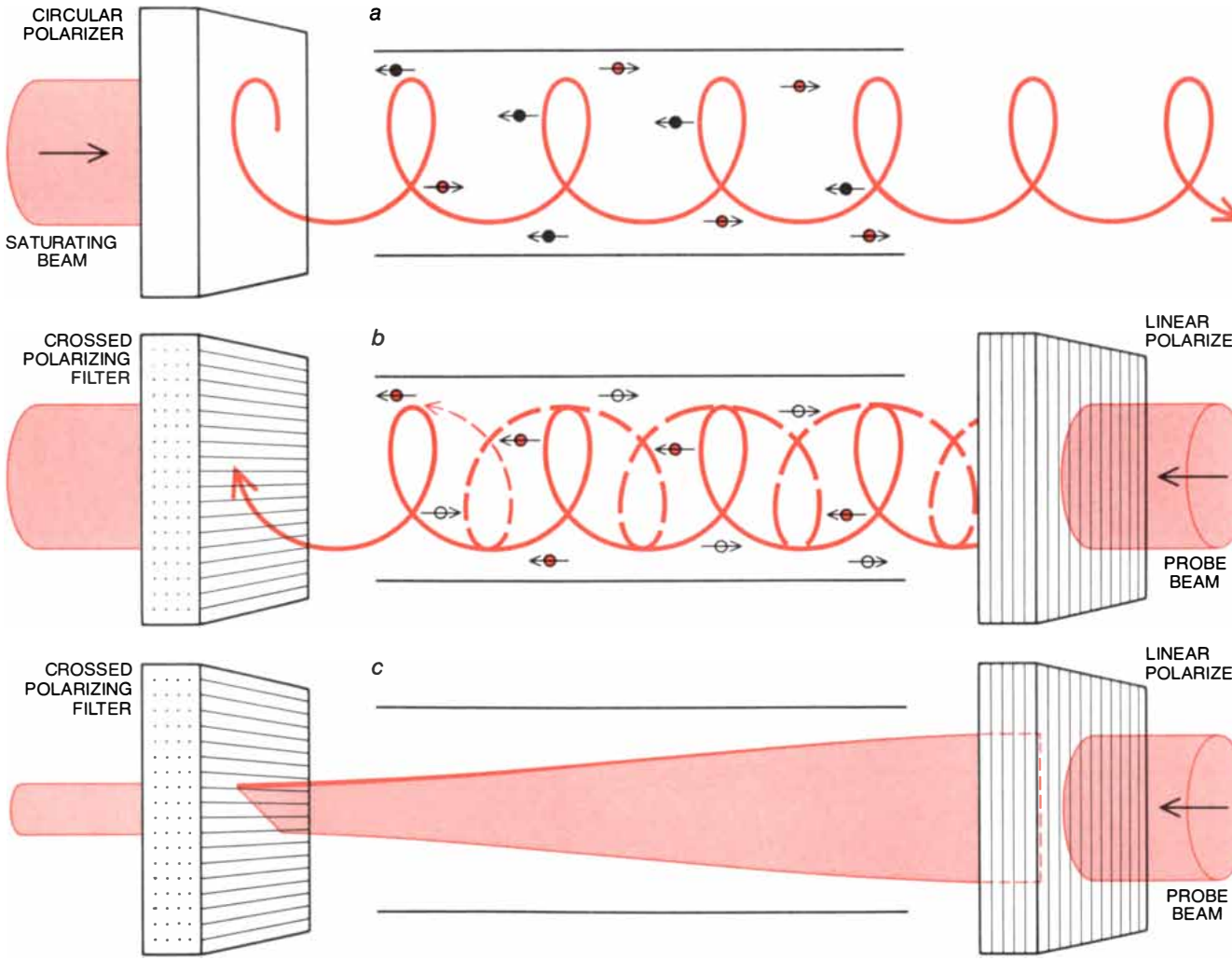
The techniques of laser spectroscopy introduced since about 1970 have significantly improved resolution because



BALMER-ALPHA LINE is resolved into a series of sharp peaks by saturation spectroscopy. Theoretical calculations indicate that the line has seven components resulting from fine-structure splittings and the Lamb shift but ignoring the hyperfine structure of much smaller scale. In a Doppler-broadened profile of the line only two components can be distinguished. By examining only those atoms that have no Doppler shift saturation spectroscopy clearly resolves four components. (A fifth peak midway between two other peaks is an artifact called a crossover resonance.) Frequency scale measures tuning of the laser from an arbitrary starting point.



POLARIZATION SPECTROSCOPY exploits the fact that small changes in the polarization of light can be detected more easily than small changes in intensity. The output of a laser is again split into two beams, but the saturating beam is given a circular polarization and the probe beam a linear polarization. Because the probe beam encounters crossed polarizing filters very little of it can reach the detector unless the hydrogen gas can in some way rotate the beam's plane of polarization. Such a rotation can be introduced by the circularly polarized saturating beam, but only if the two beams interact with the same population of atoms. In this way atoms that happen to be stationary at a given moment are singled out, as in saturation spectroscopy, by the ability to interact with light waves moving in opposite directions.



CHANGE IN THE POLARIZATION of a probe beam is induced by the prior passage of a saturating beam. Because the saturating beam is circularly polarized it is selectively absorbed by atoms that have a particular orientation (*a*), and those atoms are thereby removed from the population of atoms that can absorb radiation at the laser wavelength. Most of the remaining atoms have the opposite orientation. The linearly polarized probe beam can be regarded as a combination of two circularly polarized waves that have the opposite sense of rotation but equal intensity (*b*). When the probe beam passes through the specimen, one of the circularly polarized components is more strongly absorbed because the atoms have a prevailing orientation. The selective absorption of one circularly polarized component changes the polarization of the remaining light (*c*), and so a portion of the probe beam can penetrate the analyzing filter. The change in polarization can take place, however, only if the two beams are absorbed by the same population of atoms, namely those that have no Doppler shift.

they include methods for eliminating Doppler broadening. Three such methods will be discussed here. Two of them work by selecting for observation only those atoms in a gas whose component of motion along the optical axis happens to be zero. The third method allows all the atoms in a gas to contribute to a measurement, but it eliminates velocity broadening by arranging two Doppler shifts in opposite directions.

The Tunable Dye Laser

Lasers provide light of unequaled brightness, directionality and spectral purity. The early lasers, however, could emit light at only a few discrete wavelengths, determined by the spectral lines of the active atoms or ions. Such an instrument is of use in spectroscopy only if the transition to be observed happens to coincide with one of the available laser lines.

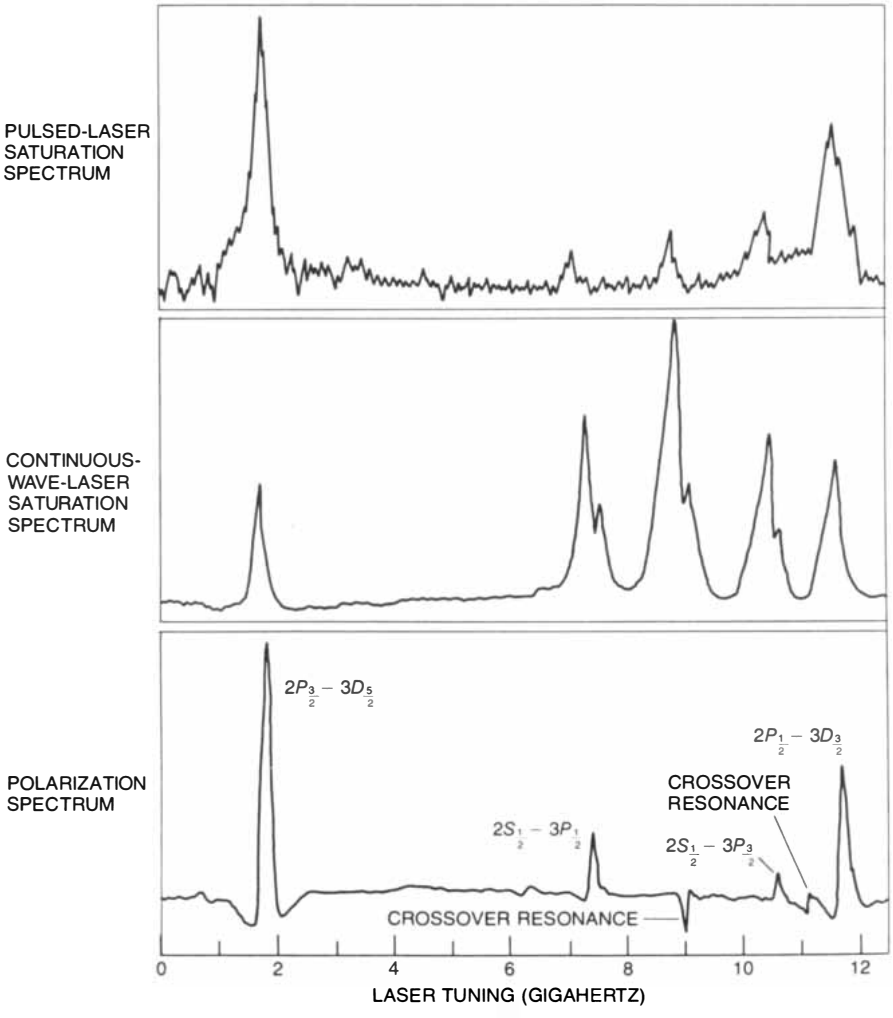
In 1966 Peter Sorokin of the Thomas J. Watson Research Center of the International Business Machines Corporation and independently Fritz Schäfer of the University of Marburg discovered that liquid solutions of certain fluorescent organic molecules can serve as the medium for a laser with a broad and continuous band of wavelengths. The organic molecules are dyes, which by nature have intense absorption bands; that is why they are strongly colored. In the laser the dye molecules are "pumped," or driven to an excited state, by an intense external light source, such as another laser.

Although the dye laser is capable of amplifying light over a wide range of wavelengths, it was soon learned that a particular wavelength could be selected by designing an optical resonator that will allow only the chosen wavelength to pass repeatedly through the amplifying dye cell. In 1970 one of us (Hänsch) devised a simple pulsed dye laser of high spectral purity. The dye, enclosed in a glass cell, was pumped by pulses of ultraviolet radiation from a nitrogen laser. The resonator consisted of a partially reflecting glass plate at one end of the laser and a diffraction grating at the other. The grating disperses the light emitted by the dye, with the result that only a narrow range of wavelengths is reflected back along the optical axis. A telescope in the laser cavity expands the beam before it strikes the grating, so that many grooves are illuminated with well-collimated light. The line width of the laser is further reduced by inserting a tilted Fabry-Perot interferometer into the cavity. The interferometer is made up of two partially reflecting surfaces whose spacing is precisely controlled; the type of interferometer placed in the laser resonator is called an etalon, and it consists of a single glass plate whose two parallel surfaces are coated with partially reflecting layers. Light waves with different wavelengths take slightly different paths through the etalon. As a result of multiple internal reflections some wavelengths interfere destructively and are extinguished, whereas other wavelengths interfere constructively and are reinforced.

The line width of this laser is a few thousandths of an angstrom, or about one part per million, which is comparable to the resolution of the largest grating spectrographs. The width can be further reduced by a factor of 100 or so by a second Fabry-Perot interferometer outside the laser cavity, which acts as a passive filter. Although at any given setting the light is highly monochromatic, the laser can be tuned over the full range of the dye absorption profile. Gross tuning can be accomplished by changing the angle of the diffraction grating or of the etalon. Finer adjustments can be made by enclosing the grating and the etalon in an airtight chamber and altering the pressure, which changes the index of refraction of the enclosed gas.

Saturation Spectroscopy

There is an ultimate limitation on the wavelength resolution of any pulsed laser. The precision with which the wavelength can be known is determined by the number of cycles available for measurement, and hence by the length of the pulse. Narrower lines require a continuous-wave laser. In 1970 Otis G. Peterson, Sam A. Tuccio and Benjamin B. Snavely of the Eastman Kodak Research Laboratories demonstrated the first continuous-wave tunable dye laser. Its principles of operation were the same as those for the pulsed dye laser, but the dye solution was made to flow rapidly in order to avoid overheating. Highly refined continuous-wave dye lasers are

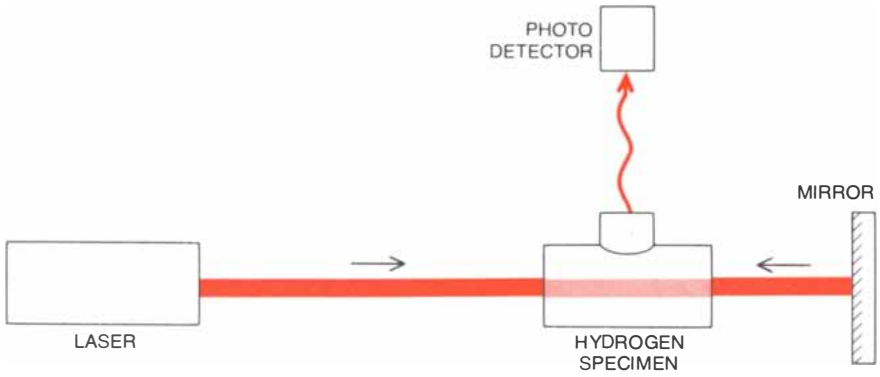


HIGHER-RESOLUTION SPECTRUM of the Balmer-alpha line was recorded by polarization spectroscopy. For comparison saturation spectrum shown on page 103 is repeated at the top and an improved saturation spectrum, made with a continuous-wave laser instead of a pulsed one, is given in the middle. In the polarization spectrum no additional components can be detected, but the four peaks that are resolved are much sharper. The two small peaks between the tall outer ones (excluding crossover resonances) have a measured width of about 35 megahertz, almost 10 times narrower than corresponding lines in the pulsed saturation spectrum. The polarization spectrum has an unusual appearance (it makes excursions below the background level as well as above it) because the graph records not the intensity of transmitted light but the rate of change in intensity as the laser is scanned through its frequency range.

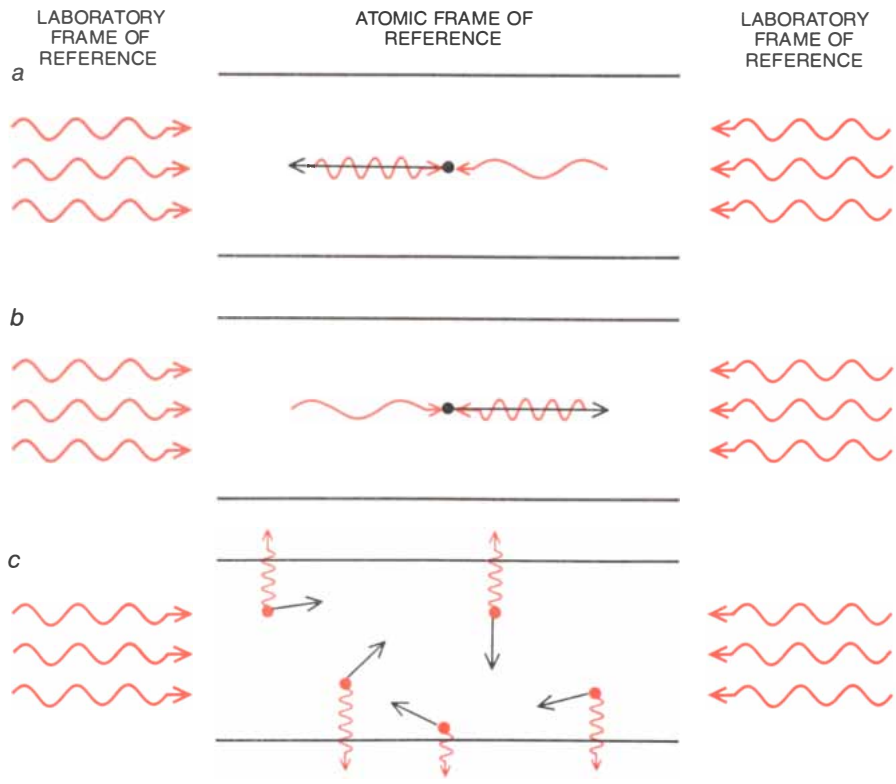
now available commercially. In several laboratories line widths as small as a few parts in 10^{12} have been achieved through electronic stabilization of the laser frequency.

Given a source of light that is monochromatic but tunable, an absorption spectrum can be measured by passing the light through a sample of the gas and scanning continuously through the frequencies surrounding a line in the spectrum. The intensity of the transmitted light can be monitored by a photoelectric detector, which should reveal a dip at the wavelength of each component of the line. This scheme, however, would merely measure the Doppler-broadened profile of the line. To take full advantage of the potential resolution of the laser some means must be found for eliminating Doppler shifts.

One such method has been given the name saturation spectroscopy. It depends for its operation on two properties of laser light: small line width and high intensity. When an atom absorbs a quantum of light, it is excited to a higher energy level and is thereby temporarily removed from the population of absorbing atoms. In an intense beam of light the atoms of a gas may absorb quanta faster than they can return to their original state. As a result the population of absorbing atoms is substantially depleted. The light bleaches a path through the gas, and a second beam of the same wavelength would find the absorbance of the gas reduced.



TWO-PHOTON SPECTROSCOPY does not select atoms without a Doppler shift but instead cancels the Doppler shifts of all the atoms in a gas. The dye laser is tuned to a frequency exactly half that of a selected atomic transition. The beam passes through the specimen and then is reflected back on itself, creating a field of standing waves. Atoms in the gas can make the selected transition by simultaneously absorbing two photons, or quanta of light, coming from opposite directions and therefore having opposite Doppler shifts. The excited states created in this way decay to some lower energy level, emitting photons that can be detected and counted.



OPPOSITE DOPPLER SHIFTS CANCEL in two-photon spectroscopy. An atom moving to the left (*a*) sees the wave coming from the left shifted to a higher frequency, but the frequency of the wave coming from the right is shifted down by an equal amount. For an atom moving to the right (*b*) the frequency shifts are the opposite of these. A stationary atom perceives no Doppler shifts. Regardless of the atom's velocity the sum of the two frequencies is constant and equal to twice the laser frequency (measured in the laboratory frame of reference). All the atoms can therefore absorb two photons (*c*) and afterward reradiate at some higher frequency.

Lamb was the first to recognize the potential of this effect for high-resolution spectroscopy. He pointed out that the two waves traveling in opposite directions inside a laser could work together to saturate the absorption of a particular class of atoms in the active medium of the laser. Assuming that the line width of the light is narrow enough, it must always interact only with those atoms whose motion shifts the wavelength into resonance with a component of the atomic spectrum. If the frequency of the light is lower than that of the line component, for example, then both waves in the cavity will interact with atoms moving toward them, and these will be different atoms for the two oppositely directed beams. There is one tuning of the laser, however, for which both beams interact with the same atoms: the tuning that corresponds to the exact wavelength of the spectral component, where the atoms that can absorb the light are those standing still or moving transversely to the optical axis. When the laser is tuned to this wavelength, the population of atoms in the absorbing state is depleted and the output of the laser declines. This "Lamb dip" was first observed by Ross A. McFarlane, William R. Bennett, Jr., and Lamb. It was employed for high-resolution spectroscopy by Abraham Szöke and Ali Javan of the Massachusetts Institute of Technology, but it has proved to be of limited utility. Only the laser transitions themselves can be studied, or molecular lines that happen to coincide with gas-laser wavelengths, and any gas to be examined must be placed inside the delicate resonator of a continuous-wave laser.

In 1970 Christian Bordé of the University of Paris and independently one of us (Hänsch) devised a technique for measuring saturation spectra outside the laser resonator. The light of a tunable laser is split into an intense saturating beam and a weaker probe beam. The two beams are then made to traverse the same path through a gas cell, but in opposite directions. If the laser is detuned slightly from the frequency of an atomic transition, then the two beams interact with different atoms and the saturating beam has no effect on the absorption of the probe beam. When the laser is tuned

so that it is absorbed by stationary atoms, however, the saturating beam depletes the population of atoms in the selected state and the probe beam experiences a smaller absorption. In effect the saturating beam labels all the atoms that happen to have a zero component of motion along the optical axis and ignores all others.

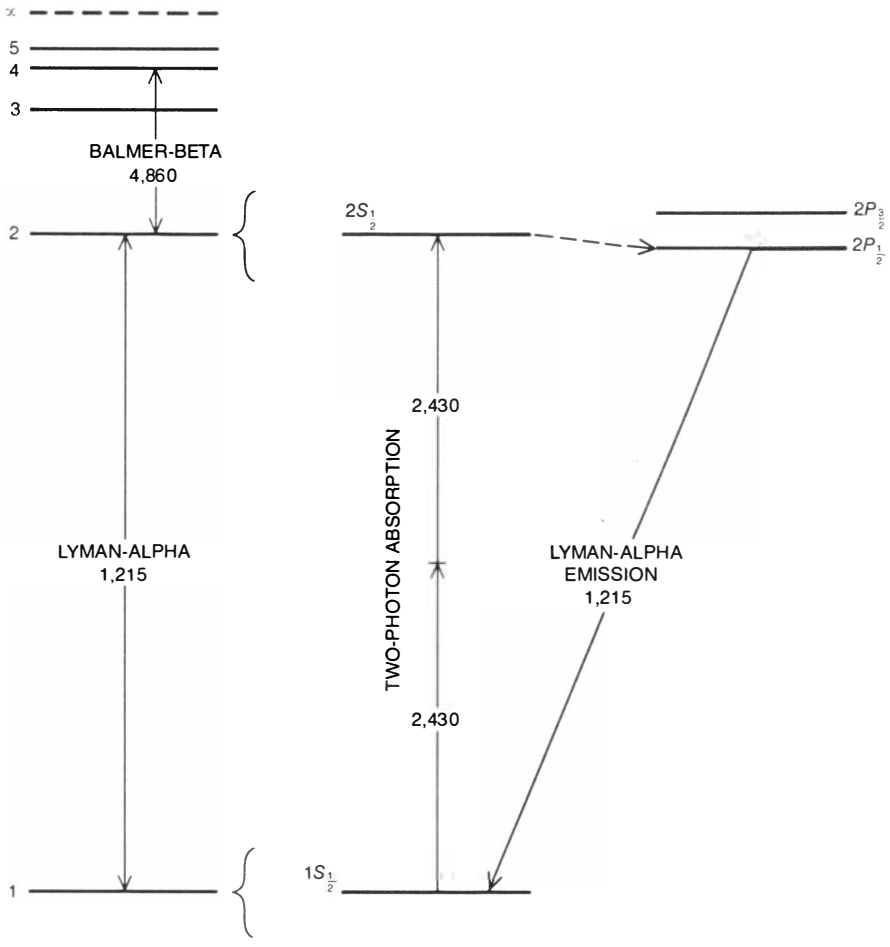
In practice the change in absorption is small and is easily obscured by noise. In order to help distinguish the signal from the background, the saturating beam is chopped, or interrupted at a known frequency, whereas the probe beam is allowed to operate continuously. The signal can then be detected by examining the amplitude of the probe beam for any modulation at the chopping frequency.

Two of us (Hänsch and Schawlow) with Issa S. Shahin, who was then a graduate student at Stanford University, examined the Balmer-alpha line by this method. The hydrogen atoms were excited to the absorbing $n = 2$ level in a low-pressure gas-discharge tube. The two light beams from a pulsed dye laser were passed through a section of the discharge plasma about 15 centimeters long, where they showed an absorption of about 50 percent.

Of the seven fine-structure components expected theoretically, four were visible, the two strongest ones being resolved far more sharply than in any previous spectrum. What is more, the Lamb shift that splits the $2S_{1/2}$ and $2P_{1/2}$ levels was directly observed for the first time in the optical absorption spectrum.

The saturation spectrum also included a fifth peak, but it was a spurious one, called a crossover resonance, which appears midway between any two transitions that share the same upper or lower level. When the laser is tuned to the frequency midway between these two line components, atoms moving in one direction can absorb the saturating beam, and atoms moving the other way can absorb the probe beam. As a result the absorption is saturated not by stationary atoms but by two classes of moving ones.

The sharpness of the peaks in the saturation spectrum suggested that the Rydberg constant could be determined with improved accuracy by measuring the absolute wavelength of one fine-structure component. Such a measurement was undertaken by a group of workers at Stanford led by one of us (Hänsch) and including in particular Munir Nayfeh. The component chosen was the $2P_{3/2}$-to-$3D_{5/2}$ transition, which is the strongest one, the one with the smallest unresolved hyperfine splitting and the one least perturbed by the electric field of the discharge tube. It was first necessary to examine the effects on the line position of variations in gas pressure, in current and voltage in the discharge tube and in laser intensity, and to correct



GROUND, OR LOWEST, STATE OF HYDROGEN can be reached only through transitions whose wavelengths are in the vacuum-ultraviolet region of the electromagnetic spectrum; such transitions can be observed more conveniently by two-photon spectroscopy than by any other technique. The Lyman-alpha line, a transition from $n = 1$ to $n = 2$, has a wavelength of 1,215 angstroms, and radiation at twice this wavelength, or 2,430 angstroms, is required for two-photon spectroscopy. Even the latter wavelength is inaccessible to dye lasers and must be generated by doubling the frequency (or halving the wavelength) of a laser tuned to 4,860 angstroms. The frequency doubling is accomplished by a crystal that under intense illumination emits the second harmonic of the incident frequency. The original laser wavelength (4,860 angstroms) happens to correspond closely to the wavelength of the Balmer-beta line. Absorption of two photons in atomic hydrogen stimulates only the $1S_{1/2}$-to-$2S_{1/2}$ transition. The excited atom is converted by weak encounters with other atoms to a $2P_{1/2}$ state, from which it returns to the ground state by emitting a photon having a wavelength of 1,215 angstroms.

for any systematic errors. In order to measure the absolute position of the line rather than the interval between two lines an accurate reference length was needed. The chosen standard was a helium-neon laser whose frequency was electronically locked to a particular hyperfine component of an absorption line of molecular iodine vapor; this line was also defined by saturation spectroscopy. The wavelength of the stabilized helium-neon laser was known accurately in terms of the international standard of length, which is a spectral line of krypton emitted under defined conditions. The data were evaluated in 1974. The result, $R = 10{,}973{,}731.43 \pm .10$ reciprocal meters, was more accurate than the best previous value by a factor of almost 10.

A related method of Doppler-free spectroscopy was introduced in 1976 by Carl Wieman, a graduate student at Stanford working with one of us (Hänsch). By exploiting the fact that small changes in the polarization of light can be detected more easily than changes in intensity, the technique achieves greatly improved sensitivity. Fewer atoms can be observed at lower laser intensity, thereby avoiding systematic line shifts and line broadening.

Polarization Spectroscopy

As in saturation spectroscopy, the laser light is divided into two beams, one more intense than the other, which traverse the sample in opposite directions. In this case, however, the weaker, probe beam sees the gas sample placed between crossed polarizing filters, so that little light reaches the detector as long as the sample does not change the polarization of the probe. The saturating beam acts on the sample to bring about such changes. It can do so because it is first passed through a quarter-wave plate, a

piece of birefringent material cut and polished so that it gives light a circular polarization. Light is said to be circularly polarized when the electric field rotates (either clockwise or counterclockwise) rather than oscillates in a plane as it does in linearly polarized light.

The probability that an atom will absorb circularly polarized light depends on the orientation of the atom's angular momentum with respect to the rotating field. Initially the orientation of the atoms is random, but the saturating beam depletes the gas of atoms that are oriented so as to absorb light of one circular polarization, leaving an excess of atoms with the opposite sense of rotation. When the probe beam, which is linearly polarized, passes through the same region of gas, the oriented atoms can alter its propagation.

A linearly polarized wave can be regarded as a superposition of two circularly polarized waves of equal intensity, one wave rotating clockwise and the other counterclockwise. When this beam passes through the gas, it encounters atoms that tend to absorb only one of its circularly polarized components, thereby attenuating it and leaving the other component the stronger. As a result when the probe beam emerges from the gas, it is elliptically polarized, and the axis of the ellipse is rotated from the plane of the original linear polarization. The beam has therefore acquired a component that can pass through the crossed polarizing filter. All of this can happen, however, only if the saturating beam and the probe beam interact with the same atoms, namely those that have no Doppler shift.

No light comes through the crossed polarizing filter if the two beams interact with different sets of atoms, and so there is essentially no background signal. Hence real peaks in the spectrum are not easily obscured by noise or by intensity fluctuations of the laser. When the signal is very small, it is often advantageous to uncross the polarizing filters slightly. The detector then registers a finite background signal, which can either increase or decrease, depending on the direction in which the plane of polarization is rotated.

John Goldsmith, a graduate student at Stanford, Erhard Weber, who was at Stanford on leave from the University of Heidelberg, and one of us (Hänsch) have recently obtained an improved spectrum of the Balmer-alpha line by laser polarization spectroscopy. The saturating and probe beams were generated by a low-power, continuous-wave dye laser with a bandwidth of less than one megahertz.

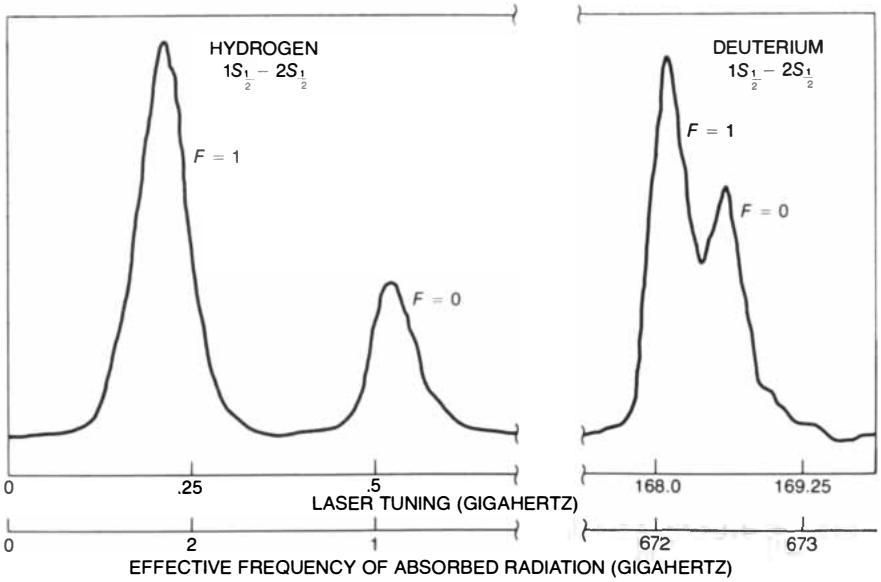
The spectrum was recorded in the form of a derivative, a mathematical function that measures the rate of change in the probe-beam intensity as the laser is tuned through a band of frequencies. In the polarization spectrum the two largest peaks are sharper than they were in the earlier measurement by saturation spectroscopy, and three smaller components are narrower still. The line width for these components is about 35 megahertz, almost 10 times sharper than in the previous measurement. These three components all represent transitions that begin in the long-lived $2S_{1/2}$ state, and each of them is known to be split into two hyperfine components separated by some 178 megahertz. It would seem that the hyperfine splitting should be clearly resolved, but only single peaks are observed. The explanation of this discrepancy is that in each case one of the hyperfine states has a net angular momentum of zero and cannot be polarized. Thus each of the observed peaks is not an unresolved mixture of two transitions but is a single hyperfine component.

The $2S_{1/2}$-to-$2P_{1/2}$ component was chosen for a new absolute wavelength measurement, again calibrated against the iodine reference line as a standard of length. The new value of the Rydberg constant derived from this measurement is $R = 10{,}973{,}731.476 \pm .032$ reciprocal meters. The measurement is almost three times as accurate as the previous one, and it places the Rydberg constant among the most precisely known of all fundamental constants. Only the speed of light and the ratio of the magnetic moments of the electron and the proton have been determined with comparable precision.

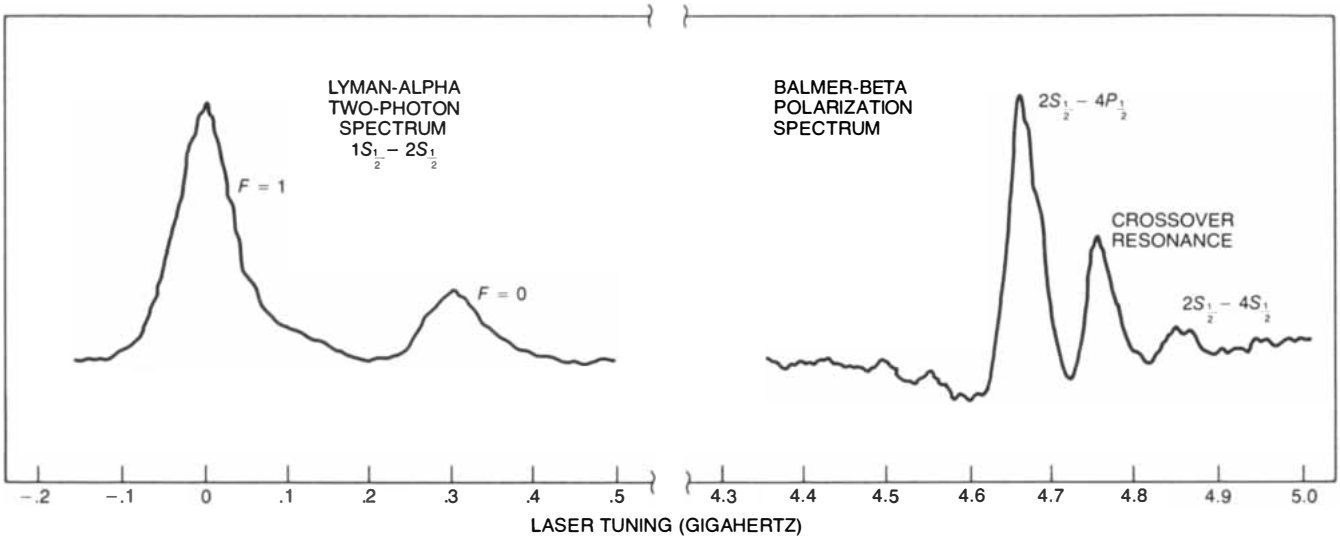
Two-Photon Spectroscopy

These methods provide convenient access to the lines of the Balmer series, those generated by transitions between the $n = 2$ energy level and higher levels. Other states cannot be reached as readily. One of particular importance is the $S_{1/2}$ state with $n = 1$, which is the ground state of the hydrogen atom. Even the nearest group of states, those with $n = 2$, is so far above the ground state that the wavelength of the corresponding line, which is designated Lyman-alpha, falls in the vacuum-ultraviolet region of the electromagnetic spectrum. Precision spectroscopy is particularly difficult at those short wavelengths, which are far beyond the range of the dye laser.

With intense lasers, however, the difficulties of vacuum-ultraviolet spectroscopy can be avoided by exciting each atom with two photons, which together provide the necessary excitation energy. The probability of an atom's absorbing two photons simultaneously is generally rather small, but it grows with the square of the light intensity. Two-photon spectroscopy is of particular interest for hydrogen because Doppler broadening can be avoided by an ingenious method first suggested in 1970 by L. S. Vasilenko, V. P. Chebotayev and A. V. Shishaev of the Institute of Thermophysics at Novosibirsk in the U.S.S.R. The method was first demonstrated in 1974 by three groups of investigators working independently. Instead of the laser beam's being split into two components the beam is reflected back on itself by a mirror, so that waves moving to the right and to the left travel along the same axis. The gas cell is placed in the standing-wave field that results from this superposition. The laser is then tuned to a frequency exactly half that



HYPERFINE STRUCTURE of the $1S_{1/2}$-to-$2S_{1/2}$ transition was resolved by two-photon spectroscopy both in normal hydrogen and in deuterium. Transitions between the two hyperfine levels in hydrogen are responsible for the radiation at a wavelength of 21 centimeters observed by radio astronomers. The natural width of the components is probably very small, perhaps as little as one hertz; here they are resolved to within about 100 megahertz. The shift between hydrogen and deuterium is caused mainly by the difference in the mass of the nuclei.



SIMULTANEOUS MEASUREMENT of the $1S_{1/2}$-to-$2S_{1/2}$ transition and the Balmer-beta line was employed to determine the magnitude of the Lamb shift for the $1S_{1/2}$ state of hydrogen. The Balmer-beta line was measured by polarization spectroscopy at the fundamental wavelength of the laser, 4,860 angstroms. The frequency-doubled output of the same laser was simultaneously employed to record a two-photon spectrum of the $1S_{1/2}$-to-$2S_{1/2}$ transition at an effective wavelength of 1,215 angstroms. If the Bohr energy levels were exact, the two transitions would be observed at the same laser tuning, but the splittings of the energy levels introduce a small discrepancy. From the measured difference in the laser tunings for the two lines, the magnitude of the $1S_{1/2}$ Lamb shift was calculated with a precision 50 times greater than that of other measurements. Value for hydrogen was found to be 8,161 ± 29 megahertz.

of a selected atomic transition. At the proper frequency each atom in the gas can simultaneously absorb two photons coming from opposite directions.

Suppose an atom moving along the optical axis in the standing-wave field encounters two photons approaching from opposite directions. One of the photons is Doppler-shifted toward the blue, meaning higher frequencies, but the other photon is shifted toward the red by an equal amount. For a stationary atom both photons have the same energy. In all cases the total energy or frequency of the two absorbed photons is constant, regardless of the velocity of the atom. As long as the laser frequency is detuned slightly from half the transition frequency none of the atoms can absorb the two counterpropagating photons. (There is a small background signal, however, from moving atoms that absorb two photons propagating in the same direction.) When the laser wavelength matches the atomic transition, a sharp increase is observed in the number of excited atoms. The Doppler-free signal is strongly enhanced because all the atoms in the sample can contribute to it rather than just those that happen to be stationary. In this respect the method differs fundamentally from both saturation spectroscopy and polarization spectroscopy, in which a population of atoms that are free of Doppler shifts is selected. In Doppler-free two-photon spectroscopy one Doppler shift is made to compensate for another.

A two-photon transition of particular interest is the transition from the $1S_{1/2}$ state to the $2S_{1/2}$ state. Because the latter state is a very long-lived one the line is expected to be extraordinarily narrow. Theoretical estimates suggest a natural line width of about one hertz, or less than one part in 10^{15}, which would make the transition one of the narrowest resonances known in any physical system. The $1S_{1/2}$ state is in itself interesting because its Lamb shift is the largest one predicted by quantum electrodynamics. An atom at the energy level $n = 1$ cannot have orbital angular momentum, and so there is no nearby P state from which the magnitude of the shift can be determined. The shift is simply a displacement of the $1S_{1/2}$ energy level from the position it would have in the absence of quantum-electrodynamical effects. The magnitude of the shift can be determined only by an absolute measurement of some spectral component, such as the $1S_{1/2}$-to-$2S_{1/2}$ transition.

A two-photon excitation of this transition was first observed in 1974 by Siu Au Lee, Wieman and others at Stanford. The Lyman-alpha line has a wavelength of 1,215 angstroms; a photon with half this energy corresponds to a wavelength of 2,430 angstroms, which is still in the ultraviolet and beyond the reach of tunable dye lasers. The Stanford workers therefore employed a dye laser operating at 4,860 angstroms, or twice the required wavelength. The intense light from this source was shined on a crystal that served as a frequency doubler. Because of the high intensity of the applied electromagnetic field the crystal not only reradiated light at the original frequency but also was driven to emit overtones, including the second harmonic, at 2,430 angstroms. About 2 percent of the energy appeared at this wavelength.

The frequency-doubled ultraviolet radiation was reflected by a mirror to form a standing wave inside a low-pressure gas-discharge tube. Hydrogen atoms that absorbed two photons returned to the ground state by emitting a single, far-ultraviolet photon at a wavelength of 1,215 angstroms. These photons escaped through a window at the side of the chamber and were detected by a photomultiplier.

The $1S_{1/2}$-to-$2S_{1/2}$ transitions of both hydrogen and deuterium were measured with this system. Each of these lines has a large hyperfine splitting, which is well resolved in the two-photon spectra. Transitions between the two hyperfine states in hydrogen are responsible for the 21-centimeter radiation that is an important signal in radio astronomy. There is also a large splitting between the hydrogen lines and the deuterium lines, caused chiefly by the difference in nuclear mass. This isotope shift was determined to an accuracy 1,000 times better than earlier measurements. A further improvement by a factor of 10 could provide a new value for the important ratio of electron mass to proton mass.

The greatest interest in the $1S_{1/2}$ state is in the determination of the ground-state Lamb shift. In the 1950's Herzberg, who was then at the Yerkes Observatory of the University of Chicago, was able to detect the shift, but he measured its value to a precision of only about 14 percent. Two-photon spectroscopy provided an opportunity for a far more accurate determination.

The most straightforward approach to measuring the Lamb shift would be to determine the absolute wavelength of the $1S_{1/2}$-to-$2S_{1/2}$ transition, but the value of the Lamb shift calculated from

that wavelength also depends on an assumed value of the Rydberg constant. The Stanford group found a better method, exploiting a peculiar but not accidental coincidence: the fundamental wavelength of the dye laser, namely 4,860 angstroms, nearly coincides with the Balmer-beta line. If Bohr's formula were correct, the correspondence would be exact: the interval from $n = 1$ to $n = 2$ (Lyman-alpha) would be just four times the interval from $n = 2$ to $n = 4$ (Balmer-beta). Actually the levels are split and shifted somewhat by relativistic and quantum-electrodynamical corrections. The corrections for the $n = 2$ and $n = 4$ levels are known with great precision, however, and so a comparison of the two transitions can determine the ground-state Lamb shift.

Three such measurements have been carried out by the Stanford group, the most recent and the most accurate one by Wieman. The Balmer-beta reference line was observed by polarization spectroscopy, while the same laser simultaneously measured the Lyman-alpha line by two-photon spectroscopy. The Balmer-beta spectrum not only resolved the fine-structure components of the line but also revealed the splitting of these components in the weak axial electric field of the gas-discharge tube. In principle the ground-state Lamb shift could be determined by measuring the difference between the Balmer line and the $1S$-to-$2S$ line and then comparing the result with the predictions of the Dirac theory. Known corrections would be applied to the positions of the $n = 2$ and $n = 4$ levels; any remaining difference would be attributed to the Lamb shift of the $1S$ state. In practice it was not quite as easy as that, because allowance had to be made for possible systematic errors and imprecisions in the apparatus. The result eventually derived was $8,161 \pm 29$ megahertz. The experiment is still far short of the accuracy of the theoretical calculations, which give a value of $8,149.43 \pm .08$ megahertz, but it represents a significant step beyond Herzberg's earlier measurement.

Higher Resolution

Many formidable obstacles would have to be overcome in order to approach the one-hertz natural line width of the $1S_{1/2}$-to-$2S_{1/2}$ transition. It would not be enough to further reduce the bandwidth of the laser, because there are several other sources of line broadening. One of them is a second-order Doppler shift, a consequence of the special theory of relativity. An atom that is moving appears to oscillate slower than one at rest, so that the moving atom effectively has a lower resonant frequency no matter what the direction of motion is. The counterpropagating beams of two-photon spectroscopy cannot cancel this shift, which should amount to about 50 kilohertz for hydrogen atoms at room temperature.

Another source of line broadening is the finite transit time of a moving atom in the laser beam. Because a moving atom is exposed to the laser light only briefly, even the light from a continuous-wave laser is perceived by the atom as a pulse, whose short duration limits the potential resolution of the spectrum. In principle both effects could be reduced by slowing the atoms, or in other words by cooling the gas. In this regard it is encouraging that Daniel Kleppner of M.I.T. has recently shown that atomic hydrogen can be cooled to the temperature of liquid helium (4.2 degrees K.) without condensing, even though diatomic hydrogen molecules condense at 20 degrees K. Two of us (Schawlow and Hänsch) have recently pointed out that laser light itself might cool a gas to a low temperature. This may seem paradoxical, since an intense light usually supplies heat, but it must be recalled that laser light, being coherent, represents a state of low entropy, or low disorder. A state of low entropy can be made to yield a low temperature.

Several laboratories are now attempting to apply the technique of two-photon spectroscopy to the $1S_{1/2}$-to-$2S_{1/2}$ transition of positronium, a hydrogen-like "atom" made up of an electron and its antiparticle, the positron. The properties of this exotic atom are predicted with great precision by quantum electrodynamics, and there is no need to apply corrections for the structure of the nucleus. The experiment is difficult, however, because positronium has a lifetime of only about 140 nanoseconds.

The hydrogen atom has been a focus of attention for almost a century, but the information embodied in this simple system of particles is by no means exhausted. Recent unified field theories, such as those of Steven Weinberg of Harvard University and Abdus Salam of the Imperial College of Science and Technology, predict subtle effects in hydrogen and other atoms that might be observed as small changes in the polarization of emitted or absorbed light. The changes are so small that there is some question of whether they can even be detected, but several groups of investigators are now attempting to measure them accurately enough to test the theories. As the technology of coherent light sources evolves it seems safe to predict that the exploration of the hydrogen spectrum will continue for decades. So far all the recent findings tend to confirm the predictions of theory. It is worth remembering, however, that several major upheavals in 20th-century physics were instigated by the discovery of minute discrepancies between theory and observation in this very spectrum. The hydrogen atom may hold surprises yet to come. On the other hand, perhaps the greatest surprise would be none at all.

Taking you where you want to go...

Wherever you'd like to go, whenever you'd like to get there, chances are the world's growing fleet of DC-9 and DC-10 jets have just what you're looking for. Our big DC-10s make more flights—more than 700 daily—than any other wide cabin jet. They fly to more places, too—165 cities—on every continent but Antarctica. And our DC-9 twinjets, on which more than a billion passengers have already flown, carry another third of a million people every day.

Giving you power...

Solar One, America's first sun-powered generating plant, was designed by McDonnell Douglas for the Department of Energy. We're helping to build it, too—a field of computer-driven mirrors that follow the sun across the sky, focus its energy on a tower-mounted boiler, and produce steam to drive the turbines that produce electricity. Excess heat goes to an oil-rock "storage battery," so that the plant can keep on working hours after the sunset. Energy is a problem. Solar One may be part of the solution.

Keeping you free.

McDonnell Douglas builds The Champion: The F-15 Eagle, America's air superiority fighter, the most effective combat plane in the world. The Eagle rules the skies, guarding the Free World. It also continues the tradition of excellence established by our F-4 Phantom, introduced in 1958 and still in production.

And now the F-4 and the F-15 are being joined by other great McDonnell Douglas combat aircraft. The F-18A Hornet for the Navy and Marines, and the AV-8B advanced vertical takeoff jet.

Advanced: the right word for everything McDonnell Douglas does.

This is just a sample of the things McDonnell Douglas is creating to bring technology to life—your life—helping to make it easier, better, safer. If you'd like to know more about us, write McDonnell Douglas, P.O. Box 14526, St. Louis, Mo. 63178.

MCDONNELL DOUGLAS

EQUAL OPPORTUNITY IN PROFESSIONAL CAREERS. SEND RESUME: BOX 14526, ST. LOUIS, MO. 63178

The Perception of Surface Blacks and Whites

What shade of gray a surface appears is related to the perceived distribution of light and shadow, which in turn depends on the perceived spatial relation between the surface and its neighbors

by Alan L. Gilchrist

The lens of the human eye projects onto the retina a two-dimensional image of the three-dimensional physical world. Partly because the retinal image is two-dimensional most investigators of color perception have assumed that depth perception has nothing to do with the process by which the human visual system determines the color of objects. The experiments I shall describe here invalidate that assumption. I have found that a change in the perceived spatial orientation of a surface can change its perceived color from black to white or from white to black.

Traditional explanations of color perception assign no special role to the central nervous system. Because that system governs depth perception my work assigns it a major role. Traditional investigations have also largely ignored the perception of the intensity and color of the light illuminating surfaces whose colors are under scrutiny. In my work I have tried to rectify this imbalance, and for the sake of simplicity I have dealt mainly with what are called neutral colors: white, black and gray.

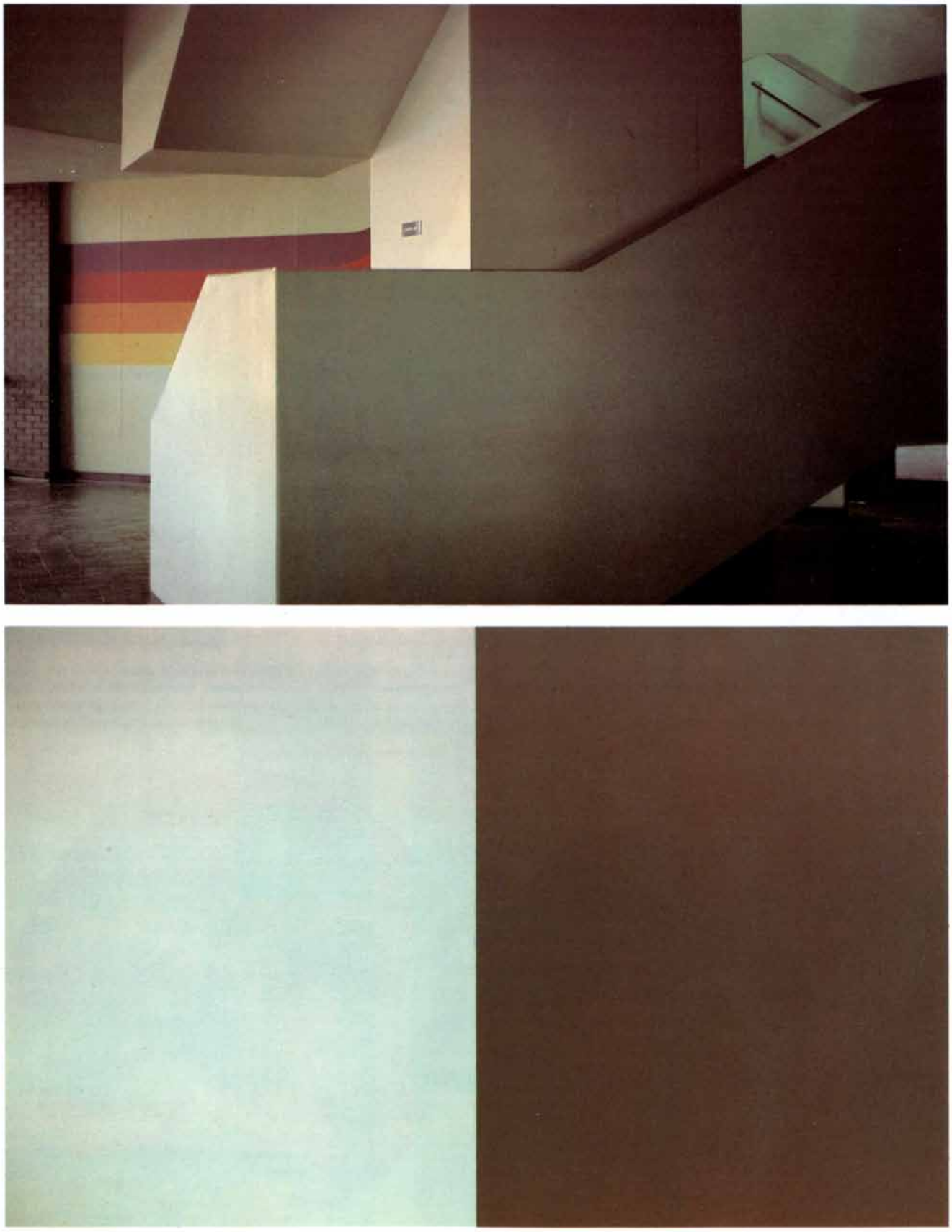
When white light, which consists of a balanced mixture of all the colors in the spectrum, strikes a colored surface, the surface reflects one wavelength (one color) more than it does the others. This dominant wavelength corresponds to the physical color of the surface. In general the light that illuminates objects in nature is not white but is an unbalanced mixture of various colors. A colored surface reflects some wavelengths in such a mixture more than others, but now the dominant wavelengths do not correspond to the physical color of the surface. The mixture of wavelengths in the light the surface reflects to the eye depends not only on the physical color of the surface but also on the mixture of wavelengths in the light that illuminates the surface. Therefore the visual system should not be considered an instrument that measures the wavelength and the intensity of reflected light; such measurements would reveal little about surface color. The visual system has the remarkable ability to correctly perceive the physical color of a surface in spite of wide variations in the mixture of wavelengths in the illumination. This is the phenomenon of color constancy.

Neutral surfaces leave unchanged the mixture of wavelengths that illuminates them, but they do alter the intensity of the light. The shades of gray from white to black are all neutral, and so they reflect various degrees of illumination. Surfaces that reflect between about 80 and 90 percent of the illumination are called white, whereas those that reflect between about 3 and 5 percent are called black. In short, the lightness, or perceived grayness, of a neutral surface corresponds to its reflectance: the percentage of illumination it reflects. Again, the visual system has the remarkable ability to determine the lightness of a neutral surface in spite of wide variations in the intensity of the illumination. Such is the phenomenon of lightness constancy, which is analogous to the phenomenon of color constancy for colored surfaces.

A physicist would determine the reflectance of a surface by comparing the intensity of the reflected light with the intensity of the illumination. The human visual system also determines reflectance, but it does so by comparing the intensity of the light reflected from a surface with the intensity of the light reflected from neighboring surfaces. For the visual system to make this comparison there must be constant relative displacement between the retinal image and the retina. This happens naturally because the eye constantly and involuntarily flicks back and forth in tremors with a frequency of between 30 and 150 cycles per second. It has been found that the visual field goes blank in one to three seconds if the relative displacement is eliminated by artificially stabilizing the image (using a special apparatus that causes the image to remain still on the retina even as the eye moves back and forth). The eye tremors indicate that the receptor cells of the retina function only in the presence of a change of stimulation. Consider what happens to an individual receptor cell during the tremors (assuming that there are no large voluntary eye movements). A cell in the interior of a patch of homogeneous stimulation within the retinal image receives no change of stimulation, whereas a cell at the boundary of a patch does receive such a change.

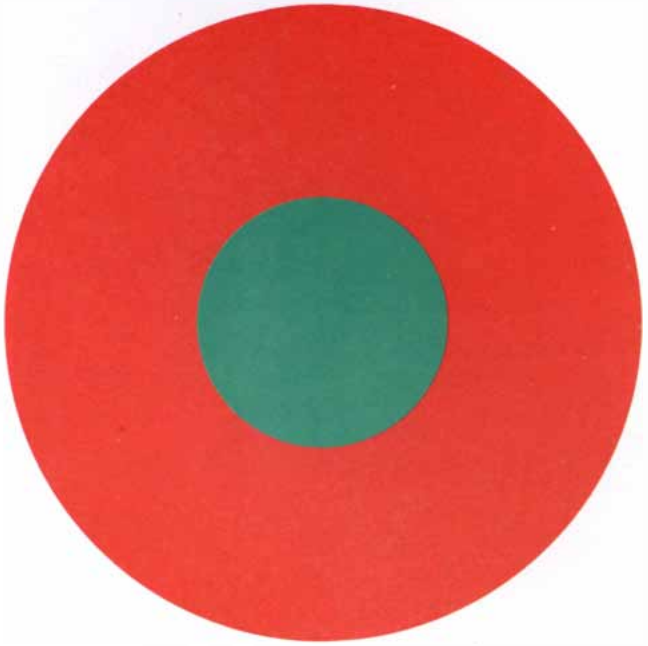
The relevance of all of this for color vision was elegantly demonstrated by John Krauskopf of Bell Laboratories. He built a display that consisted of a disk of one color, say green, surrounded by an annulus, or ring, of another color, say red. He arranged matters so that the green-red boundary between the disk and the annulus was stabilized, that is, moved in such a way as to follow the tremors. In this way all displacement between the green-red boundary and the receptor cells was eliminated. As a result the retinal boundary disappeared and the observer simply saw a single large red disk.

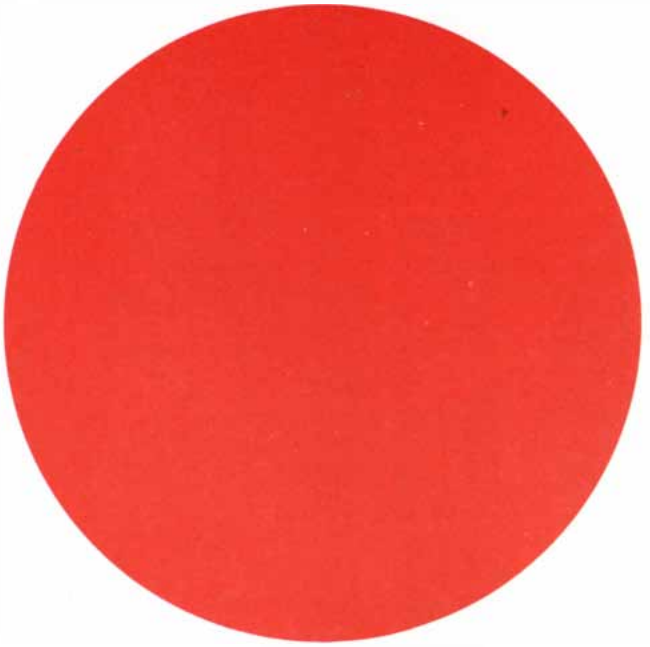
The most straightforward interpretation of this experiment is that the eye sends the brain only information about changes in light across boundaries, with areas where no change is reported being filled in by the brain as homogeneous. In Krauskopf's display, and in the normal viewing of a large red disk, the eye extracts information from the outer edge. Since no change is reported within the boundary, the brain treats the interior of the disk as homogeneous. Consider what this implies. Krauskopf's observers perceived red even in the center of the display, even though green light was striking the corresponding region of the retina. Hence if the color of the light (green) striking the disk region of the



PERCEIVED SPATIAL ORIENTATION affects color perception. In the top photograph it is apparent that the right side of the stairwell is a white surface in shadow and that the left side is an illuminated white surface. Because the visual system recognizes that the stairs turn a corner it correctly perceives that the stairwell is differentially illuminated. The bottom photograph is a closeup view of the stairwell corner. Here the two sides appear to form a plane rather than a corner. The sides are perceived to be illuminated equally, and so the visual system attributes the difference in grayness to the sides themselves. Hence the shadowed side is not correctly perceived as being white.

ILLUMINATION EDGES, or illumination differences, are of two kinds: attached and cast (projected). Attached edges are the result of changes in the spatial orientation of surfaces; cast edges are the result of shadows. In this photograph both kinds of edges are present in abundance. For example, there are attached edges where the walls turn corners to form the indentations that hold the windows. The boundaries of the trapezoidal patches of light projected on the floor by sunlight streaming through windows are examples of cast edges.





RETINAL RECEPTOR CELLS gather color information by detecting changes in light. To detect such changes there must be constant relative displacement between the retinal image and the retina, and so the eye scans a surface by constantly flicking back and forth in tremors with a frequency of between 30 and 150 cycles per second. In the display at the left the boundary between the red ring and the green disk was moved in such a way as to follow the tremors. That eliminated all displacement between the boundary and the cells. As a result retinal boundary disappeared and observer saw a red disk (*right*). The experiment was done by John Krauskopf of Bell Laboratories.

retina had nothing to do with the color (red) that was perceived there, then by the same token the color of the light (red) striking the annulus region might have nothing to do with the color (red) that was perceived there.

Krauskopf's display suggests that even under natural conditions the perceived color of a surface depends not on the light emanating from the surface but on the change in the light across the boundary of the surface. The perceived color of a surface cannot depend solely on the change in the light at the edge of the surface because such a change is strictly relative. The change in the light at the edge of a surface depends as much on the color of the background as it does on the color of the surface.

If perceived color depended only on the change in the light at the edge of a surface, the surface would look radically different in color when seen against different backgrounds. Conversely, two surfaces of radically different color could look the same. For example, the change in the intensity of reflected light from a white surface to a middle-gray background is the same as the change in the intensity from a middle-gray surface to a black background. Of course, white and middle-gray surfaces would not look the same when they were seen against these respective backgrounds. Such changes in background color can result in what are called contrast effects, but these effects are not nearly large enough to make the white surface on the middle-gray background look the same as the middle-gray surface on the black background. It is a remarkable empirical fact that perceived surface color remains largely constant in spite of changes in the background, which in turn give rise to changes in the edge information.

The visual system extracts only relative information from an edge, and such information is a small part of what the visual system needs in order to perceive color, a part that can be properly interpreted only in the context of the relative information from many other edges. There is evidence that the visual system integrates neural signals emanating from distinct edges. Recording changes in light and then integrating them is mathematically equivalent to keeping a point-by-point record of the light. Simply recording the changes, however, is more efficient and requires less transmission capacity. Consider an analogous situation in the stock market. To draw a graph showing how the price of a certain stock varies from day to day, one could either find out the price of the stock each day or find out the price only on days when it had changed. The latter method requires less transmission capacity, and so it might be preferred if getting the stock quotation required a transcontinental telephone call. Yet the information collected by this method could be easily integrated to give the stock prices for each day. I believe something similar to this process of information extraction and integration is going on in the visual system. Such a process has also been suggested by Edwin H. Land of the Polaroid Corporation.

It turns out, however, that the model of reflectance-edge extraction and integration is quite limited in its application. The model works well when the objects whose colors are under scrutiny are uniformly illuminated. It even works when the illumination is changing uniformly over the entire visual field, because changes in the overall illumination have no effect on the relative changes in the light at the boundary between two shades of gray. This is just a special case of the general phenomenon that a change in illumination does not change the ratio of the amount of light reflected off adjacent surfaces of differing reflectance. For example, consider two surfaces whose reflectances are respectively 25 and 50 percent. The ratio of reflected light will remain 1:2 regardless of the intensity of the illumination. It has been demonstrated that the neural signal generated at an edge in the retinal image remains the same if the intensity ratio at the edge remains the same.

The model of reflectance-edge extraction and integration runs into trouble when the objects under scrutiny are not uniformly illuminated. The trouble begins with spatial changes in the illumination. Such illumination edges are of two kinds: attached and cast (projected). Attached edges arise from changes in the planarity, or spatial orientation, of surfaces. If two walls of the same color meet at a corner and one wall receives more illumination than the other, then there is an attached illumination edge where they meet. Such edges would also appear in a white plaster-of-Paris sculpture where irregularities in the sculpture's shape are clearly visible because the intensity of the incident illumination varies as the angle between the light source and various areas of the sculpture changes. On the other hand, the boundary of a shadow is a cast illumination edge. So is the boundary of a spotlight projected onto a stage and the boundary of a bright trapezoidal patch of light projected onto a floor by sunlight streaming through a rectangular window.

The presence of illumination edges in the retinal image makes the simple model of reflectance-edge extraction and integration unworkable. If changes in illumination were extracted and integrated right along with reflectance changes, gross errors in perceived lightness would result. For example, a white surface in shadow might appear as a gray surface. Nevertheless, the model can be further developed to eliminate this difficulty. What is necessary is that before the edges are integrated into an intensity image they must be classified either as reflectance changes or as illumination changes. Once this classification is made, the visual system could separately integrate the two kinds of edge. The result would be two distinct images: one representing surface reflectance and the other representing surface illumination.

My own work has centered on the conditions under which the visual system can distinguish reflectance edges from illumination edges and on how the distinction might be made. I have been guided by the belief that the classification of edges cannot be carried out at the level of the retina. The fact is that spatial changes in illumination and changes in reflectance generate identical edges on the retina. Some workers have suggested that the retina could classify the edges based on their sharpness, with reflectance edges tending to be sharp and illumination edges tending to be gradual. This simple explanation does not work, however, because illumination edges (both cast and attached) are often sharp whereas reflectance edges are sometimes gradual. And yet the visual system can correctly identify the two kinds of edge. Rather than classifying edges solely on the basis of their sharpness on the retina the visual system must classify them on the basis of their relation to all the other edges in the entire retinal image. This means that not only the eye but also the central nervous system play a role in the classification of edges.

That the central nervous system is involved in color perception is suggested by a simple phenomenon Irvin Rock of Rutgers University pointed out to me. He noted that the shadowed side of a corner (for example, where two walls meet) usually seems to be approximately the same shade of gray as the illuminated side. When the corner is artificially made to appear flat by viewing it through a hole, however, the shadowed side seems to be an extremely dark gray, often even black. When the corner is seen this way, no shadow is perceived; the darkness is attributed to the surface itself.

Here is a situation where the edge at the corner can be perceived as either an illumination edge when it is seen correctly in depth or as a reflectance edge when it is made to look flat. This suggests that the visual system relies on depth information to determine whether it is the illumination or the reflectance that is changing at an edge. Moreover, the situation seems to challenge the prevailing view that lightness is determined by intensity ratios between adjacent regions of the retinal image independently of where these regions are perceived to

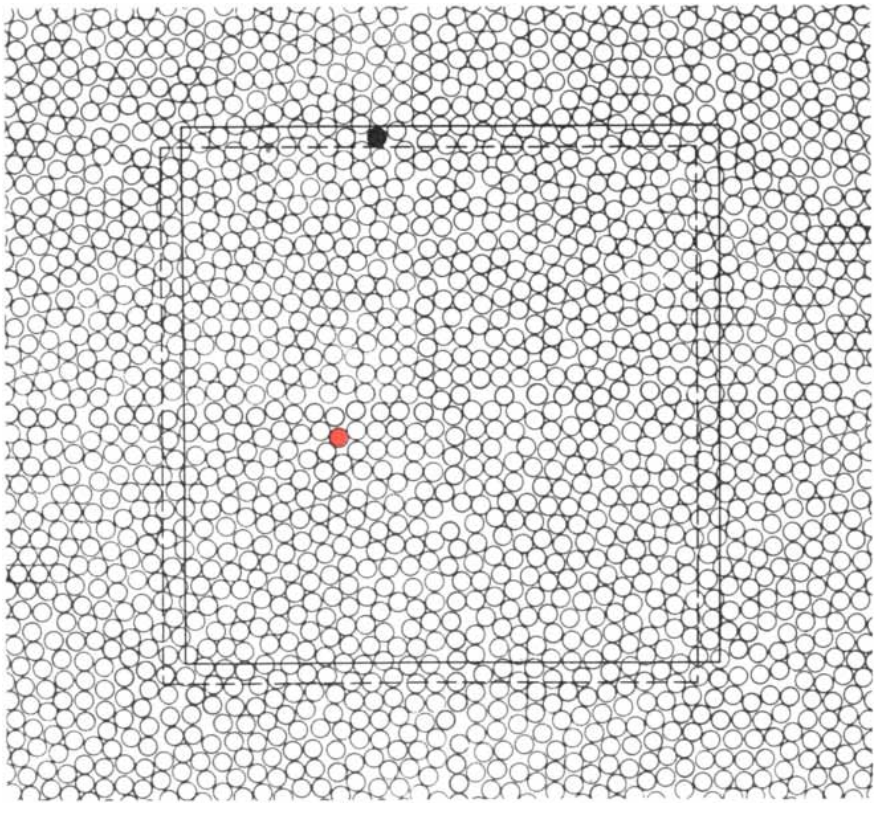
be in three-dimensional space. To demonstrate conclusively that depth perception (and hence the central nervous system) plays a role in color perception, I tried to reproduce the corner phenomenon under strict laboratory conditions.

I suspended in midair from a hidden support two white surfaces that met to form an outside right-angled corner. Behind the surfaces I put a uniform background of medium intensity. One of the white surfaces received 30 times more light than the other surface. Observers viewed the display and indicated the apparent lightness of each surface by selecting a matching sample from a chart of various shades of gray. A second group of observers viewed the surfaces when they were made to seem flat by being looked at through a hole with one eye.

When the two surfaces looked flat, the difference in intensity was perceived as a difference in reflectance, as I had expected. In other words, the illuminated side (a value of 30) looked white and the shadowed side (a value of 1) looked black. To my surprise, however, the observers saw exactly the same thing when the surfaces appeared to be at right angles to each other. This is essentially what earlier investigators had found in similar circumstances. In a few studies the perceived spatial pattern had been found to have a small effect on perceived lightness, but in most other studies, including my own, there was no such effect.

I was so impressed by the strength of the effect of spatial orientation under natural conditions that I felt something significant about the natural context of the corner must have been lost in my attempt to reproduce the situation in the laboratory. I tried in several ways to enrich the context of the laboratory display. For example, I placed objects in the vicinity of the suspended corner so as to make the lighting conditions manifest to the observers. All these attempts failed to change the results.

Finally I found the effect of spatial orientation when I put a black surface next to and coplanar with each of the original white surfaces. As in the original display, the 30 : 1 ratio in intensity between the white surfaces was seen as a difference in reflectance when they appeared to be coplanar. When the two surfaces were seen at right angles to each other, the same intensity ratio was perceived as an illumination difference. In other words, when the surfaces were seen as being coplanar, the illuminated surface looked white and the shadowed surface looked black, but when the surfaces were seen as an outside corner, the shadowed surface was seen correctly as a white surface in shadow.



STIMULATION OF A RECEPTOR CELL (*black dot*) changes as the edge of the retinal image moves across the cell when the image shifts from one position (*solid rectangle*) to another (*broken rectangle*) because of the eye tremors. Cells such as the colored one that the boundary of the retinal image does not move across do not receive a change of stimulation. Since the green-red boundary in Krauskopf's display did not cross cells, the display appeared to be red.

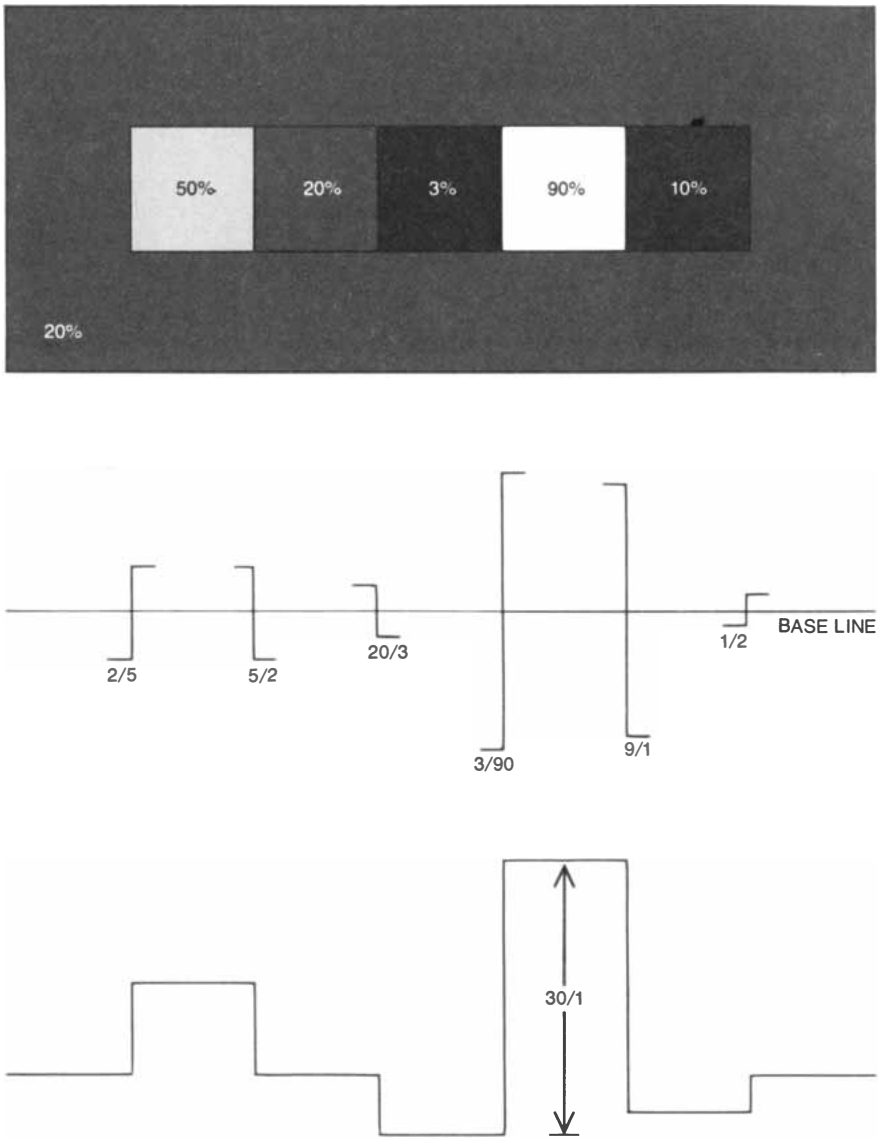
A reasonable interpretation of these results is that the visual system operates to account for the entire pattern of retinal intensities as generated by either changes in illumination or changes in reflectance. The original two-surface display could be seen as being uniformly illuminated because the two intensity levels (30 and 1) could be treated as resulting from different shades of gray, namely white (30) and black (1). The more complex four-surface display involves a 900 : 1 range of intensities. (The illuminated white surface is 900 times more intense than the shadowed black surface.) Since the range in intensity of shades of gray that can result solely from reflectance is normally limited to 30 : 1, the four-surface display also requires the perception of some variations in illumination. Which edges will be treated as representing illumination changes and which as representing reflectance changes depends on the overall organization of the scene, particularly the perceived three-dimensional layout.

Next I modified the four-surface display to rule out the possibility that the results were caused by some kind of contrast effect operating between adjacent areas of the two-dimensional retinal image. The horizontal plane of the modified display consisted of a large white square attached to a small black trapezoidal tab that extended outward toward the observer [*see illustration on page 118*]. The vertical plane consisted of a large black square attached to a small white tab that extended upward. In other words, the tabs extended out into midair from the corner formed by the two large squares, and so each tab appears on three sides against a background that is not in the same plane. As in the original four-surface display, the horizontal plane received 30 times more light than the vertical plane. The tabs are trapezoidal in order to create an illusion in depth perception.

Seen with both eyes the tabs correctly appear as trapezoids lying in their actual planes. In this case the vertical tab appears to be almost white and the horizontal tab almost black. On the other hand, seen through an aperture with only one eye each trapezoidal tab appears to be a small square lying in the same plane as the larger square that surrounds it on three sides. In this case the perceived colors are the reverse: the vertical tab appears to be black and the horizontal tab appears to be white.

In both cases, however, the retinal image is the same: each tab is seen against a square background that surrounds it on three sides. This means that the relation between the target and its background in the retinal image is irrelevant to the target's perceived shade of gray. The shade of gray turns out to depend on the relation between the intensity of the target and the intensity of whatever surface

seems to lie in the same plane, even if that coplanar surface does not provide the background of the target in the retinal image. Therefore when the display is seen with both eyes, the shadowed white vertical tab appears to be coplanar with the shadowed black vertical background, and so the tab appears to be white. In addition the illuminated black horizontal tab appears to be coplanar with the illuminated white horizontal background, and so the tab looks black. When the display is seen with one eye, the shadowed white vertical tab appears to be coplanar with the illuminated white horizontal background, and so the tab looks black. Moreover, the illuminated black horizontal tab appears to be coplanar with the shadowed black vertical background, and so the tab looks white.



REFLECTANCE OF A SURFACE is the percentage of illumination that the surface reflects. The top part of the illustration indicates the reflectances of several surfaces in a visual scene. The eye responds only to changes in the intensity of the reflected light. The middle part of the illustration indicates the kind of information that is generated at edges in the retinal image. Each vertical line indicates the change in intensity at an edge. The visual system integrates information from all the edges in order to reconstruct the original intensity pattern (*bottom*).

The model of edge extraction, classification and integration was put to a more direct test in a series of experiments designed to probe what observers would see if a reflectance edge were mistakenly perceived as an illumination edge or vice versa. The corner experiments I have described showed that a change in the perceived type of edge results in a change in the perceived lightness of the region bounded by the perceptually altered edge. If perceived lightness depends on the visual system's integration of a series of spatially remote edges, then even stranger perceptual mistakes will manifest themselves. For example, if a single reflectance edge in a complex display is made to appear as an illumination edge, then in certain circumstances this ought to have as great an effect on the perceived lightness of regions that are remote from the altered edge as it has on the regions that are bounded by the altered edge.

At the State University of New York at Stony Brook, Stanley Delman and I designed one such edge-substitution experiment. We studied a familiar display, known as simultaneous lightness contrast, that is often treated in psychology textbooks. Two middle-gray squares are placed respectively on adjacent white and black backgrounds. It turns out that for some reason the gray square on the white background looks slightly darker than the gray square on the black background. This effect, however, was not the subject of our experiment. We wanted to reproduce the same pattern of reflected light that the eye receives from this display when the boundary between the white and black backgrounds was perceived as the boundary between high and low levels of illumination.

To this end we affixed to a wall a large rectangular piece of middle-gray paper that would serve as the immediate background for both target squares. The difference in immediate background intensities was created by casting a beam of light with a sharp edge across half of the gray rectangle. Part of the beam also illuminated the background wall on three sides of the rectangle. The intensity of the beam was set so that the intensity ratio of the illuminated half of the gray rectangle to the shadowed half was 30 : 1, which is equal to the intensity ratio of white paper to black paper, as was the case in the original display. Next we needed to add the target squares. They could not be the same shade of gray (that is, have the same reflectance), because they would then reflect different amounts of light to the eye owing to their unequal illumination.

In the original lightness-contrast display that we were trying to reproduce the two squares reflected exactly the same amount of light. It is also true that in the original display the squares had the same reflectance. That did not concern us, however, because our goal was to reproduce not all the facets of the original display but only the information the visual system received: the pattern of light intensities that reached the eye. To achieve this goal the target square on the shadowed side had to have a reflectance 30 times as great as the other target square, which would be receiving 30 times as much illumination. Therefore we placed a white square on the shadowed side and a black square on the illuminated side.

In this way we thought we had reproduced in a logical manner the light intensities of the original pattern. Now it was time to test our results photometrically and empirically. Photometric measurements showed that the targets reflected equal amounts of light, that the

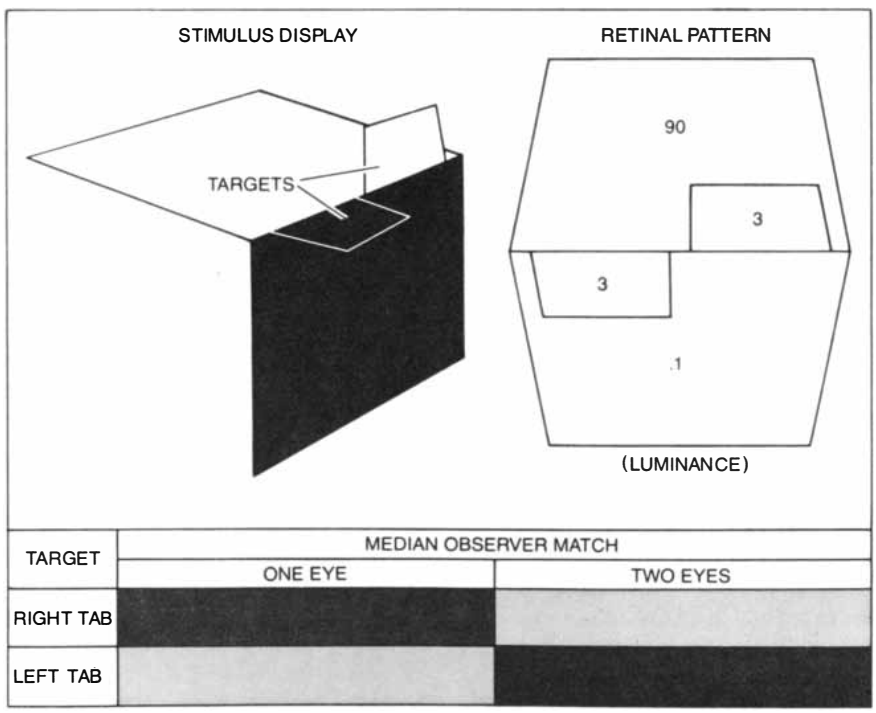
immediate background on one side reflected 30 times more light than the immediate background on the other side and that on a logarithmic scale the target intensities were halfway between the intensities of the two immediate backgrounds. All of this was true of the original display as well. Moreover, the reproduction preserved the geometry of the original.

Empirical considerations also indicated that our reproduction was faithful. Under control conditions observers viewed our reproduction through a rectangular aperture in a cardboard screen. The screen limited the observers' view so that the only parts of the display that were visible were the two square targets and their immediate backgrounds: the illuminated and shadowed halves of the gray rectangle. Under these conditions the display looked just like the original. The immediate backgrounds looked respectively white and black even though they were actually a middle gray that had been differentially illuminated. The illumination edge between the immediate backgrounds was perceived as a reflectance edge, and the square targets looked middle gray, one slightly darker than the other.

All of this served merely to assure us that we had correctly reproduced the pattern of light reflected by the original display. The real test came when we had observers view the display without looking through the confining cardboard screen. Under these conditions the observers could clearly see that a beam of light was illuminating half of the rectangular piece of paper since the beam of light illuminated the part of the background wall that was behind this half of the rectangle. Now the illuminated target was correctly perceived as black whereas the shadowed target was correctly perceived as white.

The perceived lightness of each target changed radically from the screened display to the unscreened one in spite of the fact that the amount of light each target reflected remained the same, that the amount of light each immediate background reflected remained the same and hence that the neural signal generated at the retinal edge of each immediate background must have remained the same. This clearly means that color perception is not simply a function of the amount of light a surface reflects, of the intensity ratio of a surface to its immediate background or of the neural signal generated at the retinal edge of a surface.

It seems that a much more relativistic process is going on in the visual system. The boundary between each target square and its immediate background gives only the relation between the light reflected by each of these areas. If one target is to be compared with the other, the relation between the immediate backgrounds must be taken into account. If the immediate backgrounds are perceived to differ radically in reflectance, then this perception, together with the edges between the targets and the backgrounds, suggests that the targets are nearly the same in lightness. If, on the other hand, the immediate backgrounds are perceived to differ radically in illumination, then the targets appear to lie against the same (reflectance) background and hence appear to differ extremely in lightness.



STIMULUS DISPLAY, consisting of an illuminated horizontal plane and a shadowed vertical plane, was viewed by observers who looked down at it from an angle of 45 degrees. Because the targets were trapezoids each seemed to lie in the same plane as one of the larger background squares when the display was viewed with one eye through an aperture. When the targets were viewed with both eyes, they were seen in their actual spatial positions. In both cases the retinal pattern remained the same, although the perceived lightness changed. The chart in this illustration indicates what the observers saw. The perceived shade of gray turned out to depend on the relation between the intensity of the target and the intensity of whatever surface appeared to lie in the same plane, even if that surface was not the background of the target in the retinal image.

It is now time to consider how the model of edge classification incorporates the above results. In the screened case where the display looked like the original lightness-contrast display the changes at the edges could simply be extracted and integrated to form an intensity profile of the display. In the unscreened case the same intensity profile would be generated if the edges were not classified either as illumination edges or as reflectance edges. If, however, the edges are classified before the integration, and if the integration is done only within each class of edges, then two separate profiles are generated. In that case the reflectance profile of the screened display should be different from the reflectance profile of the unscreened one. The difference is that the middle edge representing the relation between the immediate backgrounds is present only in the reflectance profile of the screened display. This means that in the unscreened display the targets look sharply different in reflectance (namely as white and as black) rather than appearing to be two shades of middle gray. Of course, the edge between the immediate backgrounds that is missing from the reflectance profile would manifest itself in the illumination profile. Thus the illumination profile would show a region of high illumination next to a region of low illumination. In the screened display the illumination profile would simply be uniform.

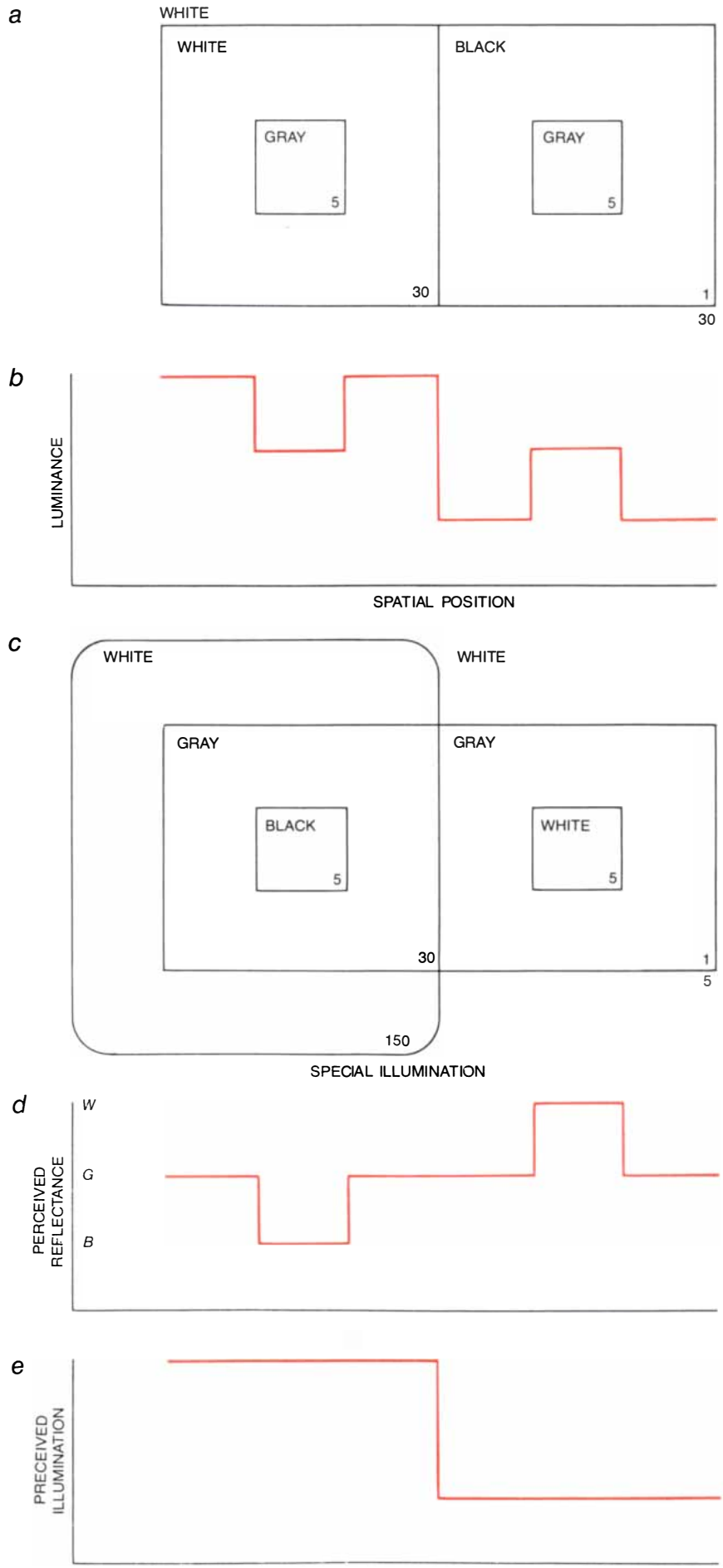
Alfred Yarbus of the Academy of Sciences of the U.S.S.R. did a similar experiment using the technique by which the image on the retina is stabilized even as the eye moves back and forth. He placed a white and a black region next to each other on a red background [*see illustration on page 122*]. Disk-shaped targets of the same shade of red as the background were placed in the center of both the white and the black region. Under normal viewing conditions the disk on the white region would look slightly darker than the disk on the black region. Yarbus altered the conditions so as to eliminate the perceptual boundaries between the white and the black region and between these regions and the red background. This he accomplished with an apparatus that caused the physical boundaries to move along with the eye,

as Krauskopf had done with his green disk and red annulus. Since the eye registers only changes in light intensity when there is constant relative displacement between the retinal image and the retina, it could not detect the presence of the white and black regions. As a result the disks appeared to stand on a homogeneous red background. Yarbus did not report his results quantitatively, but he indicated that one disk (the one objectively on the white region) appeared very dark whereas the other disk appeared very light.

Yarbus' results were essentially the same as mine, although there were significant differences in our respective methods that are worth exploring. In Yarbus' experiment and in my unscreened display the targets were made to look as though they had a common background by removing from the reflectance profile the edge that partitioned the background. As far as perceived lightness goes the effects are the same. What is fascinating is that Yarbus removed the edge at the point of extraction, whereas I removed the edge at the point of classification. To put it another way, in Yarbus' experiment the eye never detects the edge, whereas in my unscreened display the eye detects it and then classifies it correctly as an illumination edge. In Yarbus' experiment there is no reason to believe the observers saw the targets as being differentially illuminated.

Yarbus' experiment gives direct evidence for the validity of the concept of edge extraction and integration. The edge of each disk did not change; rather, boundaries that partitioned the background disappeared. As with Krauskopf's green disk and red annulus, the region bounded by the disappearing edge seemed to change color. Here the white and black regions of the background disappeared, leaving only a single uniformly colored background. Yarbus' work, however, reveals an additional phenomenon: the disks appeared to change color even though their boundaries remained constantly visible. The obvious conclusion is that information at one edge (here the edge between a disk and its immediate background) is integrated in some way with the information at a remote edge (here the edge between the immediate background and the surrounding background) before the color of any region is finally perceived.

It would be helpful at this point to discuss why the visual system determines surface colors by comparing intensities of reflected light. As I have mentioned, making such a comparison involves extracting information from edges. This suggests that it is more useful to think of the retinal image as a pattern of edges than it is to think of it as a mosaic of color patches, as traditional accounts of color perception would



have it. The edges are of two kinds, and so the retinal image is actually a dual image.

The retina is a light-sensitive surface designed to register gradients in the optic array. There are basically two kinds of change in the physical world that result in gradients of stimulation on the retina: change in surface reflectance and change in illumination. Each kind of change is capable by itself of producing a complex pattern of stimulation on the retina. From a large expanse of uniformly illuminated wallpaper the retina receives a pattern arising solely from changes in surface reflectance. On the other hand, from the rumpled white sheets of an unmade bed or from a snow-covered landscape the retina receives a pattern of stimulation arising solely out of changes in illumination. Normally these two kinds of change are at work simultaneously. It is as if two separate patterns have been laid on top of each other on the retina. The task of the visual system is to disentangle the two patterns.

The duality of the retinal image implies that each point in the visual system has at least two values: a reflectance value and an illumination value. The perception of transparency involves a similar two-value phenomenon, because a surface of one color is seen behind a surface of another color. The same phenomenon is involved in reflections. For example, when an observer looks through a window, the observer sees two scenes at once: the actual scene outside the window and another scene reflected on the inside of the window. As a result it is possible to see two colors in the same place.

Consider an experience I had recently. A book with a red cover had been left on top of the dashboard of my car in such a way that I could see a red reflection of the book as I looked out through the windshield. I was surprised to find that distant objects retained their normal colors as they were viewed through the red reflection. Even green objects seen through the red reflection looked

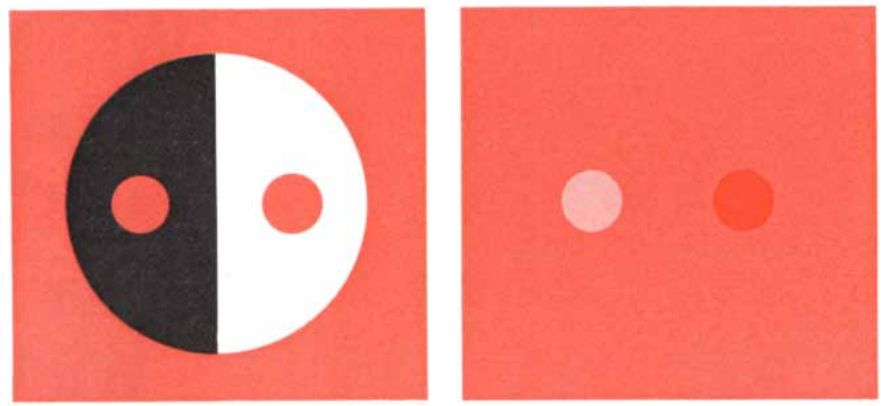
LIGHTNESS-CONTRAST DISPLAY (*a*) gives rise to an intensity pattern (*b*) that consists solely of reflectance edges. The two gray squares are perceived as being almost the same shade of middle gray. The identical intensity pattern can result from an illumination difference and an altered display (*c*). When observers viewed the display, they could clearly see that a rectangular beam of light was illuminating half of it (*e*), and so the illuminated target was correctly perceived as being black and the shadowed target was correctly perceived as being white (*d*). When the display was seen through an aperture that allowed the observers to view only the gray rectangle, the visual system did not recognize that the display was differentially illuminated, so that it looked like the original reflectance display.

green. This interested me because I of course knew that when red and green light are mixed in isolation, they form yellow. Then when I held up my hand to block the rest of the scene and viewed just a patch of the red and green through a small opening between my fingers, I did see yellow. The opening in my hand had imposed the same boundary on the red and the green light, and so they mixed to form yellow. In the original situation, however, the boundary of the red reflection did not coincide with the boundary of the green object. As a result the red and the green light appeared as elements of separate images rather than as yellow. Apparently when one speaks of mixing colors, one should actually speak of mixing edges (changes in the light) rather than of mixing light itself.

Such considerations indicate that the retina does not act as a photocell in measuring the intensity and color at each point in a scene. Man-made measuring devices are designed to respond to only one physical quantity at a time. For example, if a voltmeter is sensitive to resistance or, even worse, to temperature, the meter is considered to be defective. To cite such devices as models of the human sensory system can be extremely misleading. Unlike man-made measuring devices, the human visual system seems capable of processing multiple variables simultaneously. Accounts of color perception relying on inappropriate measuring-device models maintain that the visual system gains information about surface reflectance by sacrificing information about illumination. Such accounts unnecessarily limit the visual system to simple situations, when in fact it can handle complex situations by extracting information about changes in reflectance and illumination.

I have discussed here how the visual system deals with retinal images composed of both reflectance edges and illumination edges. It is also possible to study the perception of images that show only one kind of edge. I wondered how things would look if all the reflectance edges were missing from a situation. Alan Jacobsen and I built two miniature rooms that consisted entirely of illumination edges. We furnished each room identically with the same number of objects of varying size and shape. We painted one room, including all its contents, a matte (nonglossy) white, and we painted the other matte black. Observers viewed each room through an aperture in one of its walls, which prevented them from seeing the light bulb that illuminated the room.

It is important to keep in mind that each of these rooms projects a complex nonuniform pattern of stimulation on the retina of the observer. In the room there are sharp edges at corners and fuzzier ones across walls and curved surfaces. Some of the edges are the result of cast shadows, but all of them are



OBSERVERS viewed the display at the left under conditions where the white and black immediate backgrounds were not perceived because the boundary between them and also the boundaries between each of them and the surrounding red background were moved back and forth in such a way that they followed the eye tremors. The target disks then appeared to lie on a homogeneously red background (*right*). The boundary of the disk at the right still carried the information that the disk was darker than its immediate background, and so it appeared a very dark red. The boundary of the disk at the left still carried the information that the disk was lighter than its immediate background, so that it appeared a very light red. This experiment, done by Alfred Yarbus of the Academy of Sciences of the U.S.S.R., indicates that a change in the information at an edge affects regions that are spatially remote from the altered edge.

illumination edges, since we eliminated the reflectance edges by covering everything with paint of a single reflectance.

In presenting these displays to naïve observers we had a number of questions in mind. We wondered whether the edges in the room would be perceived as changes in reflectance, as prevailing theories of lightness perception would predict, or as changes in illumination, as the changes actually were. In other words, would each room be correctly perceived as a single shade of gray? And if it was so perceived, what shade of gray would it be? Would the black room look black and the white room white?

It turns out that 22 of our 24 observers saw each room as being uniform in lightness. They had correctly attributed the variations in intensity to variations in illumination. All of them saw the white room as white. As for the black room, all of them saw it as consisting of only a single shade of gray, although the perceived shade varied from observer to observer. The shades ranged from black to middle gray and averaged to a dark gray.

The model of edge extraction, classification and integration suggested that the illumination edges in the two rooms would be perceived as such and hence that the reflectance of each room would seem uniform. Yet there is nothing in the model so far to predict that observers could identify at least roughly the correct shade of gray. What information makes this identification possible? A plausible hypothesis is that the identification is made on the basis of the intensity of the light. In our experiment the average intensity of the light reflected from the white room is higher than the average intensity of the light reflected from the black room. By modifying our experiment, however, we invalidated the hypothesis. We lowered the intensity of the light source in the white room and raised the intensity of that in the black room until every point in the black room reflected more light than the corresponding point in the white room. The results were the same: the white room looked white and the black room looked dark gray.

We now have a promising lead to how the visual system determines the shade of gray in these rooms, although we do not yet have a complete explanation. (John Robinson helped me to develop this lead.) We believe the operative factor is the effect of indirect illumination on the retinal pattern. Every point in our rooms received illumination from two sources: directly from the light bulb and indirectly from light reflected from other surfaces in the room. In the black room, which had a reflectance of about 3 percent, there was scarcely any indirect illumination. Direct light accounted for almost all the light shining on the various surfaces. Direct light generates sharp edges, and so the intensity profile of the black room revealed a wild pattern of up-and-down swings. Although the relative pattern of direct illumination turned out to be the same in the two rooms, there was much more indirect illumination in the white room, whose reflectance was about 90 percent. Such abundant indirect light had the effect of smearing out the sharp edges resulting from the direct illumination, so that the intensity profile of the white room revealed a smoother pattern of gradual changes.

When we reduced the illumination in the white room, the shape of the intensity profile did not change. The profile merely showed a uniform decrease in intensity. If the white room were painted

The Working Trees: Class of 2020.

By the time the seedlings in our tree nursery become productive "working trees," America will require more than twice the wood fiber it does today.

Since doubling our land resource is virtually impossible, Potlatch has turned to other means to satisfy future demand.

One of the ways we plan to increase yields is through genetically-improved trees.

Many seedlings in our nursery originate from "parents" significantly superior in certain desirable characteristics. Such traits can be combined and enhanced through testing and crossbreeding.

Our objective is to develop trees with better growth rates, fiber quality and insect and disease resistance.

A better class of working trees will help us get the most out of the land under our productive control.

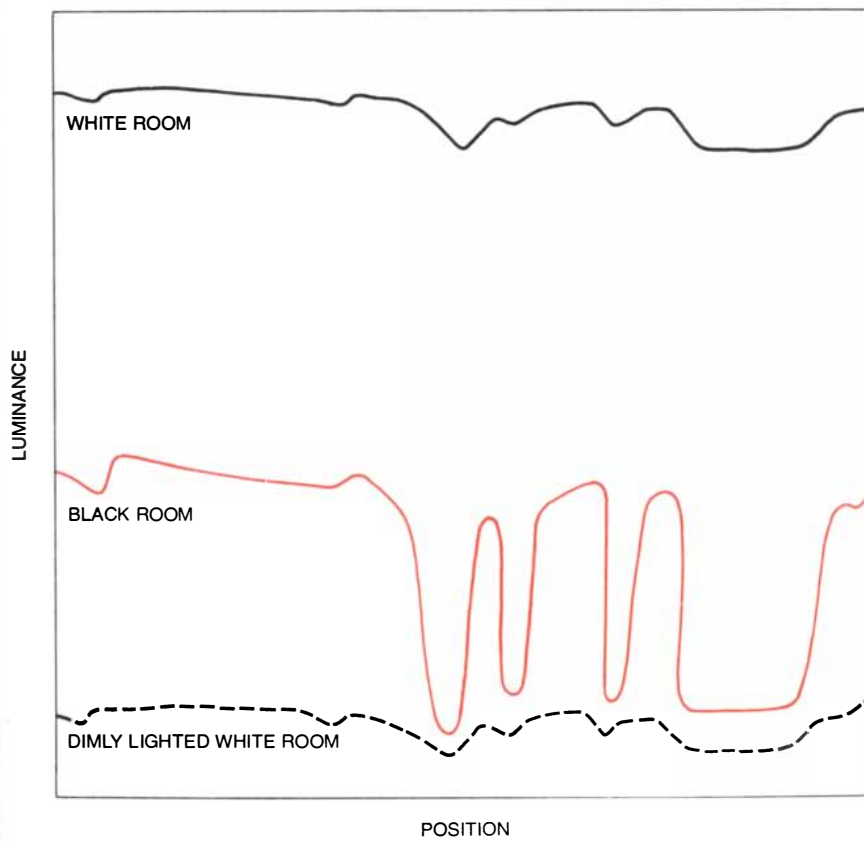
Potlatch Corporation, Public Information Office, P.O. Box 3591, San Francisco, CA 94119.

Potlatch

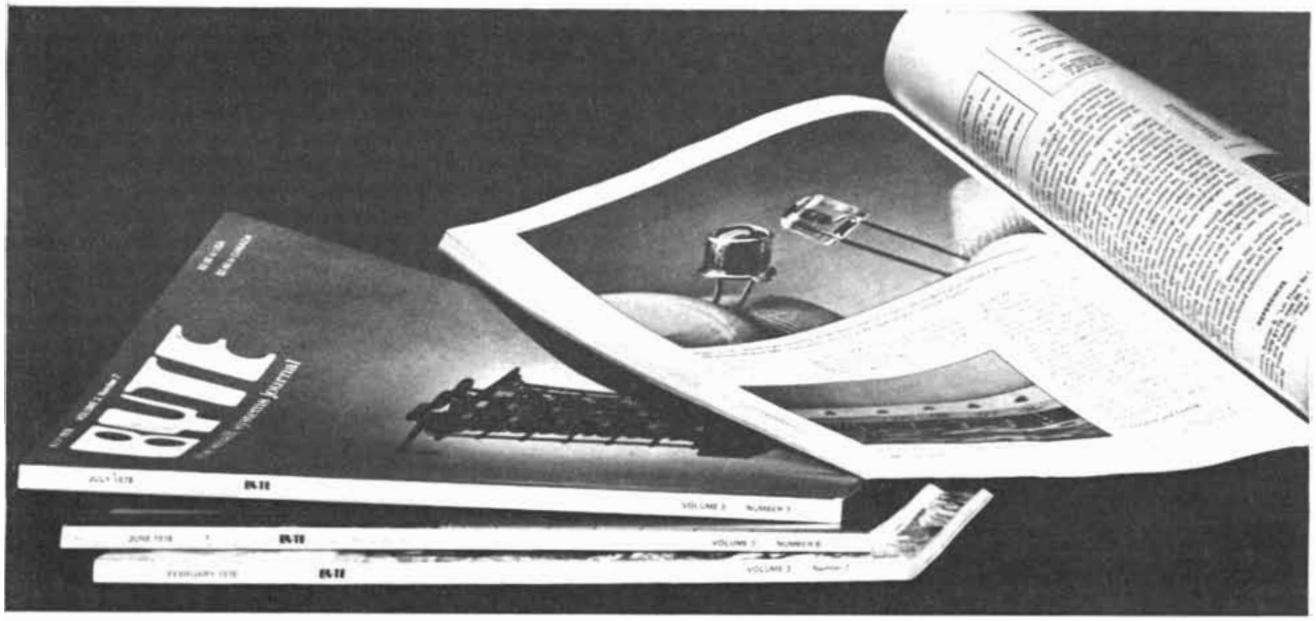
a darker shade of gray, the shape of its profile would change. Our work demonstrates that black rooms have sharper gradients than white rooms. In some way the visual system deciphers the shade of gray of a room from the shape of the intensity pattern of the room. Our work in this area is only at the beginning. As we learn more about the information stored in the intensity pattern, we hope to understand the deciphering process.

My co-workers and I have begun to extend our analysis of shades of gray to the chromatic colors. In one experiment we compared the observations of subjects who viewed a white room illuminated with blue light with the observations of those who viewed a blue room illuminated with white light. All the subjects were able to tell whether the blueness came from the surfaces or from the illumination, in spite of the fact that the light reflected off corresponding patches in the two rooms might be identical. This result clearly indicates that the perception of illumination also plays an important role in the perception of chromatic surface color.

The question of whether or not surface-color perception can be explained without reference to the perception of illumination has been debated since the end of the 19th century, when Hermann von Helmholtz first proposed that surface color could be determined only after illumination had been estimated. Helmholtz, however, was unable to specify how the illumination is estimated, so that most accounts of color perception have failed to refer to perceived illumination. The experiments I have described here suggest that Helmholtz was correct in his emphasis on perceived illumination but incorrect in his ideas about its relation to perceived surface color. He assumed incorrectly that the eye measures light. Current work shows that the eye compares light by extracting edge information, and so the visual system acquires information about illumination in exactly the same way that it acquires information about the colors of the surfaces in a scene. It turns out that the perception of illumination and the perception of surface color are parallel processes in the visual system involving the decomposition of the retinal image into separate patterns of illumination and surface color.



INDIRECT ILLUMINATIONS may play a role in how the visual system determines the grayness of a surface. The top curve depicts the intensity profile of a furnished room that was painted entirely white; the bottom curve, the same room dimly lighted; the middle curve, an identical room painted entirely black. In the white room abundant indirect illumination smeared out the sharp intensity edges resulting from the direct illumination, and so the intensity profiles (*top and bottom*) are comparatively smooth. In the black room there was very little indirect illumination, so that the profile (*middle*) revealed a pattern of up-and-down swings characteristic of edges generated by direct illumination. Perceived grayness does not depend on the intensity of reflected light, since the black room reflected more light than the dimly lighted white room.



Get the inside story...

on software, hardware, simulations, computer games, robotics, computers and calculators, languages and compilers, custom systems design and scores of applications.

Read BYTE®.

BYTE, the leading magazine in the personal computer field, will keep you aware of fast-paced changes in the ever growing world of microprocessors. You'll find **BYTE** tutorials invaluable, reports on home computer applications instructive, and the reviews of computer products stimulating.

Tutorial information in **BYTE** is of interest to both the beginner and experienced computer user. **BYTE**'s editorial content explores the fun of using and applying computers for personal satisfaction in activities from electronic music to electronic mail, from games to pragmatic personal business. In **BYTE** is found authoritative yet easily read coverage of principles of computer hardware and software design, approaches to novel applications, and essentials of proven interest to personal computer enthusiasts.

Each month serious computer users read **BYTE.** They're rewarded with timely articles by professionals, computer scientists and competent amateurs. Isn't it time you got the inside story? Subscribe now to **BYTE,** the Small Systems Journal.

Fill in and mail the coupon today. Read your first copy of **BYTE**. If it is everything you expected, honor our invoice. If it isn't, just write "Cancel" on the invoice and mail it back. You won't be billed and the first issue is yours at no charge.

Allow 6 to 8 weeks for processing. If you have any questions, dial toll free 800-258-5485. This is the number of the **BYTE** Subscription Department.

BYTE Subscription Dept. P.O. Box 590 Martinsville, NJ 08836
Please enter my subscription for

☐ One year $15 (12 issues) ☐ Two years $27 ☐ Three years $39

☐ Check enclosed entitles me to 13 issues for price of 12 (North America only)
☐ Bill Visa ☐ Bill Master Charge ☐ Bill me (North America only)

Card Number _____ Expiration Date _____

Signature _____ Name (please print) _____

Address _____

City _____ State/Province/Country _____ Code _____

Foreign Rates (To expedite service, please remit in U.S. Funds)
☐ Canada or Mexico ☐ One year $17.50 ☐ Two years $32 ☐ Three years $46.50
☐ Europe, one year (air delivered) $25
☐ All other countries, one year (surface delivered) $25. Air delivery available on request.

7139

Ion Channels in the Nerve-Cell Membrane

The nerve impulse is generated by the flow of sodium and potassium ions through molecular channels embedded in the nerve membrane. The operation of these ion channels is now being intensively studied

by Richard D. Keynes

The fundamental unit of information transmitted from one part of an animal's nervous system to another is a single action potential, or nerve impulse. Unlike the dots and dashes of the Morse code, the shape of each impulse is the same, with an amplitude of about 100 millivolts and a fixed duration of about a millisecond; the information content depends simply on the frequency of impulses carried by a given nerve fiber and on the connections the fiber makes at each end.

The transmission line for the action potential is the axon, or nerve fiber, the elongated cylindrical process that arises from each peripheral nerve cell. The transmission mechanism depends to some extent on the structural similarity between the axon and an insulated electrical cable, but if the core and the sheath of a cable had the electrical resistances typical of a nerve fiber, the signal would be rapidly attenuated with distance. The action potential, however, is self-sustaining and maintains a constant amplitude throughout its passage along the fiber. This much has been known for some time, but the molecular details of how the process is achieved are still being worked out.

Where invertebrate animals need to be capable of making a quick mechanical response in a situation of danger (examples of which would be the flip of a lobster's tail, the withdrawal of a worm into its burrow and the jet-propelled escape of a squid) they have evolved giant axons to ensure rapid conduction of the impulses that activate the contraction of their muscles. As far as electrophysiologists are concerned, the giant axon par excellence is that of the squid. It provides them with a readily dissected single nerve fiber more than half a millimeter in diameter, into which complex electrode assemblies can be inserted for electrical measurements and within which the ionic composition can be changed at the will of the experimenter.

The evolution of the nervous system has taken a different course in vertebrate animals, where nerve-impulse conduction in many of the nerve fibers is speeded up considerably by the presence of an insulating sheath of the fatty material myelin, which enables the impulse to jump quickly from one node of Ranvier, where the myelin is interrupted for a short distance, to the next. This myelinization is a purely structural modification of the fibers, and the excitable membrane of the axon behaves in just the same way where it is exposed at the nodes of a myelinated fiber as it does all along a nonmyelinated fiber.

Myelinization confers an appreciable evolutionary advantage on vertebrates because it allows a 25-fold reduction in fiber diameter for the attainment of a given conduction velocity. As a result, however, myelinated fibers are harder to isolate by dissection and furnish a less favorable material than giant nonmyelinated fibers for studying the properties of the excitable membrane.

The basic events underlying the generation of the action potential were elucidated 25 years ago by Alan L. Hodgkin and Andrew F. Huxley, working on the squid giant axon at the Marine Biological Association of the United Kingdom in Plymouth (a contribution for which they shared a Nobel prize in 1963). They showed that the electrical excitability of the nerve membrane depends on its possession of a voltage-sensitive ionic permeability system that enables it to utilize energy stored in the ionic concentration gradients set up by an energy-dependent ion pump.

As in most living cells, the interior of the axon—the axoplasm—is rich in potassium ions and poor in sodium ions, whereas in the body fluids outside the axon the relative concentrations of these ions are reversed. Because of this unequal distribution of ions, together with the fact that the membrane in its resting state is much more permeable to potassium than to sodium, the nerve has a resting potential of from 60 to 70 millivolts, the inside being electrically negative with respect to the outside.

The event that triggers the electrical activity of the nerve is a depolarization of the membrane, that is, a reduction of the potential difference across it, which may be brought about either by an excitatory input from another nerve cell or by an impulse advancing from farther down the fiber. This depolarization opens channels through the membrane that are highly selective for sodium, and sodium ions begin to flow into the fiber, driven by the large inwardly directed electrochemical gradient. The entry of positively charged ions further depolarizes the membrane and therefore opens even wider the gates controlling the sodium channels, so that once the rise in sodium permeability has been initiated the permeability builds up explosively until all the channels are open. As a result the membrane potential first falls to zero and then is reversed. The inward flow of sodium slows down, however, as its equilibrium potential is approached, finally ceasing altogether because of the onset of a process termed inactivation, which inexorably closes the channels. After a slight delay a second set of channels selective for potassium open up, and potassium ions flow outward down their electrochemical gradient until the original resting potential is restored.

Because of the cable structure of the axon, the wave of depolarization spreads along it at a velocity that (in a 600-micrometer squid axon or a myelinated 12-micrometer frog axon) is about 20 meters per second. In a squid axon the rise in the internal sodium concentration during a single impulse is very small and only about a millionth part of the internal potassium is lost, so that there is no measurable change in the ionic pumping rate afterward. In the long run, however, the exchange of ions that provides the immediate source of energy for the propagation of the action potential has to be achieved by the sodi-

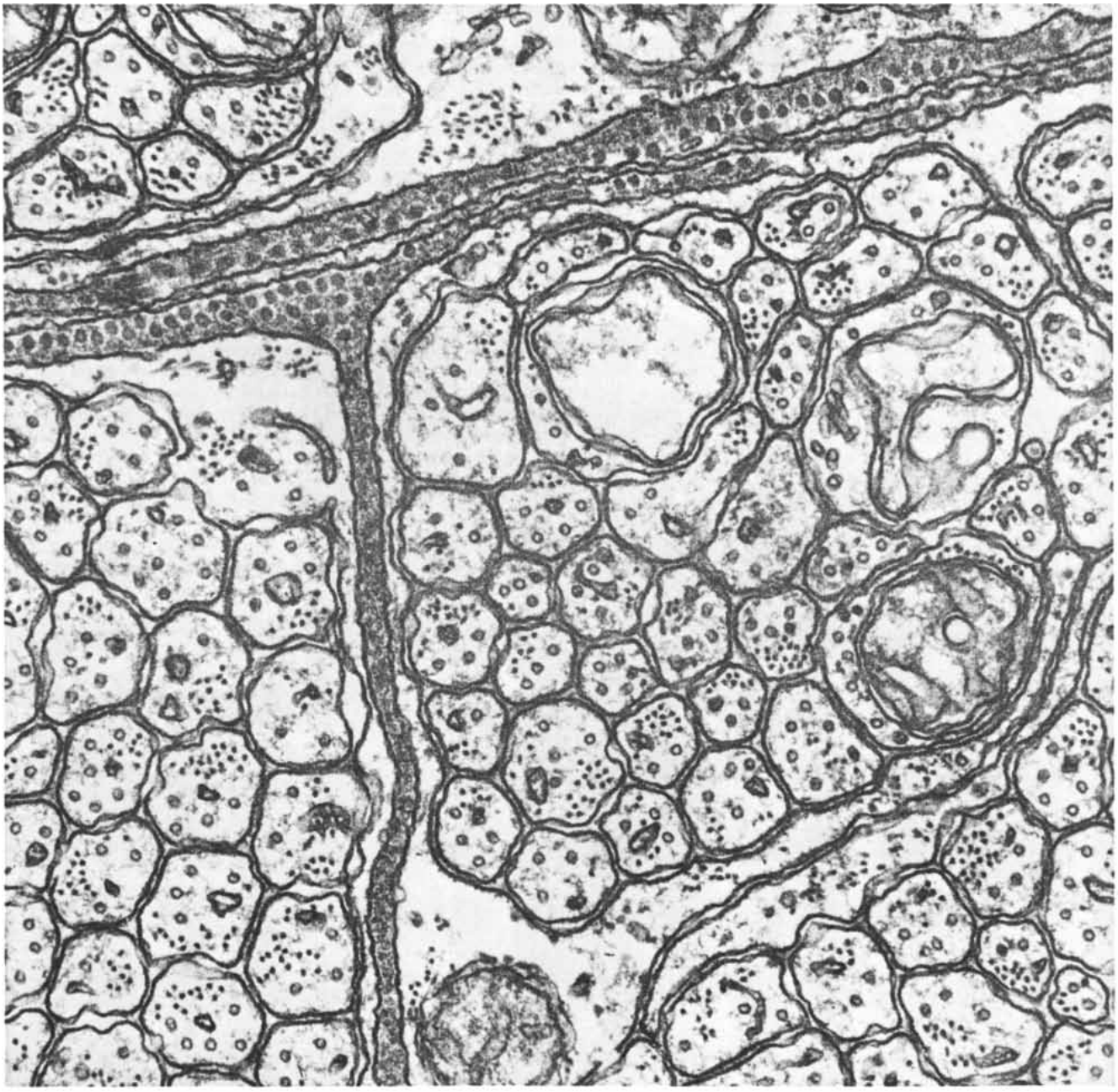
um pump, which is fueled by cellular metabolism.

In order to explain this sequence of events Hodgkin and Huxley postulated the existence of parallel sets of ion channels that were selective for sodium or potassium and were controlled by the electric field across the membrane. They then proceeded to devise a method for making quantitative studies of the properties of the sodium and potassium channels. The aim of their "voltage clamp" system was to make it possible to alter the membrane potential instantaneously to a new level and hold it there, at the same time measuring the flow of ionic current that resulted.

The voltage-clamp technique involved inserting two electrodes into the axon, one electrode to monitor the internal potential and the other to allow current to pass across the membrane to a large external electrode. A feedback amplifier was connected in the circuit in such a way that the monitored potential was compared with the control voltage set by the experimenter, and just enough current was caused to flow into the axon to reduce any discrepancy between the internal and the external electrode to zero.

The membrane potential could thus be clamped at a predetermined level or sequence of levels, and the time and voltage dependence of the ionic current could be measured under various experimental conditions. Since the original introduction of the voltage-clamp technique it has been adapted for application to myelinated nerve fibers and many other types of excitable tissue and is indeed the principal tool employed in the majority of investigations into the properties of excitable membranes.

The voltage-clamp records obtained by Hodgkin and Huxley showed that the sodium and potassium systems behaved



BUNDLES OF AXONS, or individual nerve fibers, within the olfactory nerve of the pike are shown in transverse section in this transmission electron micrograph. Each axon is a membranous tube filled with a viscous fluid, the axoplasm. The axon membrane has separate molecular channels for sodium and potassium ions; the flows of these two ions across the membrane give rise to the nerve impulse. In the pike axon the sodium and potassium channels are sparsely distributed, with a density of only a few channels per square micrometer of membrane. This micrograph, which magnifies the fibers 80,000 diameters, was made by Alexander von Muralt of University of Bern.

in a distinctly different way when the membrane was depolarized. The sodium permeability rose quickly to a peak and then declined to zero as inactivation set in, whereas the potassium permeability rose with an appreciable time lag but was not inactivated regardless of how long the depolarization was maintained. The time courses of the permeability changes were analyzed in detail, and the results were embodied in a famous set of mathematical equations describing the properties of the two types of ion channel in quantitative terms. This enabled Hodgkin and Huxley to attain their primary objective of testing the correctness of their analysis by predicting from their equations the precise shape of the conducted action potential and also such features as the conduction velocity and the amount of sodium and potassium exchanged. The calculations were triumphantly successful.

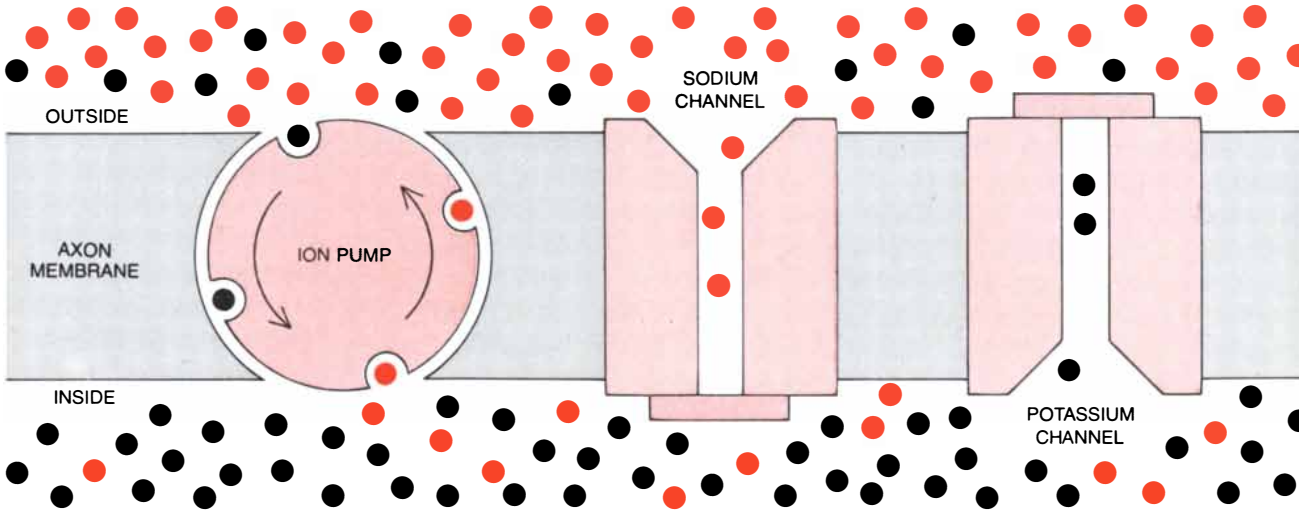
This classic piece of research was a major step forward, but it left unanswered two fundamental questions. First, how do the ion channels manage to discriminate so markedly between singly charged positive ions as alike as those of sodium, potassium and the other alkali metals? Lithium ions penetrate the sodium channel just about as well as sodium ions do, but potassium ions pass through only a twelfth as easily, and for rubidium and cesium ions the fraction is too small to be measured. Second, how does the opening and closing of the channels come to depend so strongly and with such appropriate timing on the electric field across the membrane?

The mechanism of channel specificity has been examined by Bertil Hille of the University of Washington in an elegant series of experiments on voltage-clamped myelinated frog nerve fibers. Hille examined the permeability of the node of Ranvier to a wide variety of small organic positive ions, and he found that only molecules measuring less than about three angstrom units by five angstrom units in cross section could get through the sodium channels. In part, therefore, selection depends on a good fit between the dimensions of the penetrating ion and those of the channel. Even among positive ions of the appropriate size, however, there were striking differences in permeability. For example, hydroxylamine ($OH–NH_3^+$) and hydrazine ($NH_2–NH_3^+$) passed readily through the channel, whereas methylamine ($CH_3–NH_3^+$) did not. To explain the discrepancy Hille proposed that the sodium channel is lined at its narrowest point with oxygen atoms, one of which bears a negative charge. Positive ions containing hydroxyl (OH) or amino (NH_2) groups are able to slide through the channel by making hydrogen bonds to the oxygens, but methyl groups (CH_3), which do not form hydrogen bonds, are too wide to pass through.

In measurements of the ionic current when the sodium outside was replaced by other ions either capable or incapable of passing through the channel Hille obtained evidence for the binding of sodium ions to fixed negatively charged sites somewhere in the channel, and for what is known as "single file" behavior. This phenomenon, in which the ions pass through the channel one behind the other, was first observed by Hodgkin and me some years ago in radioactive-label experiments on the potassium channels of cuttlefish giant axons. Hille suggested that there are four energy barriers in the sodium channel and that only one ion is allowed inside at a time.

The highest of these energy barriers constitutes the selectivity filter. The relative ease with which hydrated sodium ions surmount it depends on their ability to undergo a stepwise removal of stabilizing water molecules by interacting with an ionized carboxylic acid group (COO^-) in the wall of the channel. The geometry of the situation is such that the somewhat larger potassium ions cannot shed their shell of water molecules nearly as easily, and so they encounter a much higher energy barrier. A convincing picture of the molecular architecture of the portion of the sodium channel responsible for its selectivity is hence beginning to emerge.

Much of the recent progress toward understanding the voltage-sensitive gating mechanism of the sodium channels has been made with the help of a remarkable nerve poison called tetrodotoxin. This molecule binds to the outward-facing mouth of the sodium channel, thereby blocking the passage of sodium ions with an affinity that makes it one of the most toxic substances known. It is found in quite large amounts in the ovaries, liver, skin and intestines (and to a lesser extent in the muscle) of puffer fishes of the family Tetraodontidae, of which the species *Spheroides porphyreus* is known to the Japanese as fugu. Fugu is a popular delicacy in Japan, but before eating it one is advised to make sure that the chef is properly certified for his knowledge of the correct procedure for discarding the really toxic parts of the fish! Tetrodotoxin is also present in the tissues of two animals quite unrelated to the puffer fishes, the California newt *Taricha torosa* and the Australian blue-ringed octopus *Hapalochlaena maculosa*, and a molecular analogue differing only



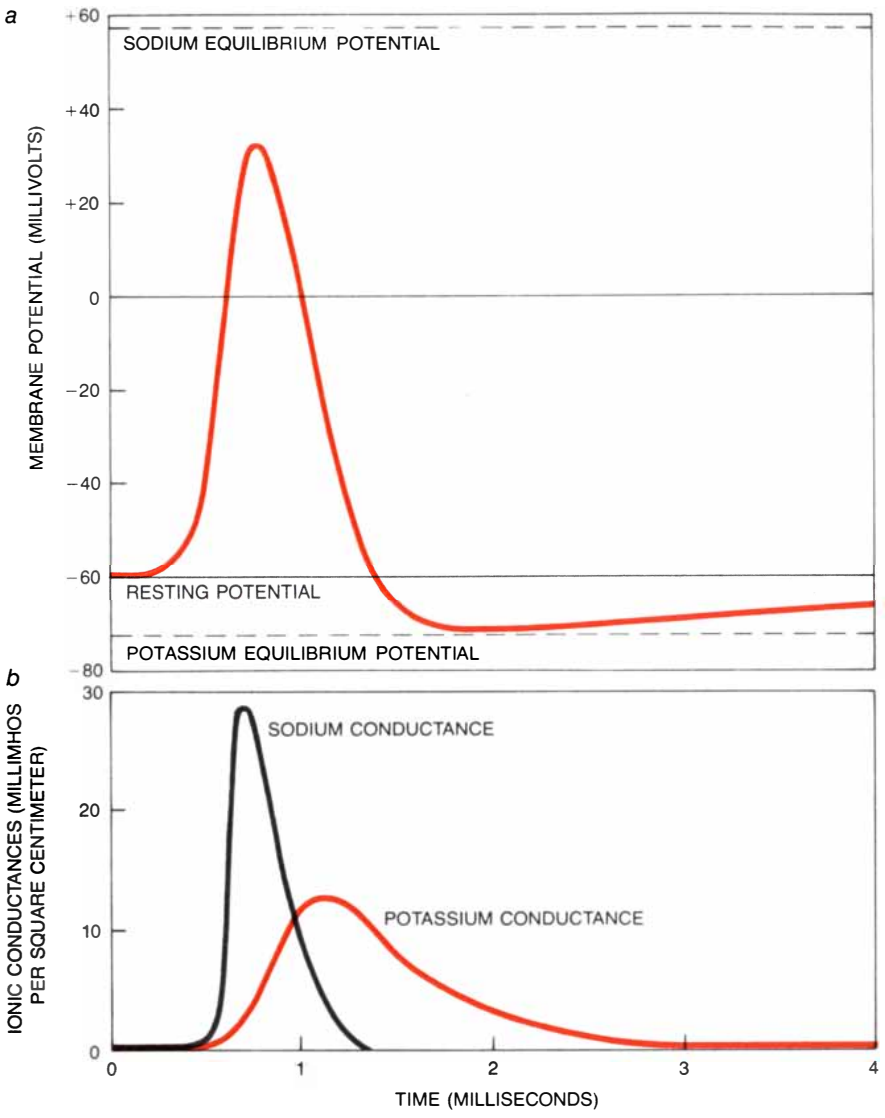
TWO KINDS OF ION CHANNELS traverse the axon membrane. An energy-dependent "pump" driven by the hydrolysis of ATP transports sodium ions (*color dots*) and potassium ions (*black dots*) "uphill" and establishes concentration gradients of these ions across the membrane. The second type of channel enables the ions to flow "downhill" in response to changes in the voltage across the membrane. In this diagram the separate sodium and potassium channels are shown in the resting state, with the charged gates held closed by the membrane potential. When the resting potential is reduced, the channels open (the sodium channels quickly and the potassium channels slowly) and give rise to a pulse of current that propagates down the axon. This article concerns the channels that allow downhill ion movements.

in one side group has recently been isolated from the Central American frog *Atelopus chiriquensis.*

Another paralytic poison of equally high toxicity, which acts in the same way as tetrodotoxin does in spite of its somewhat different structure, is saxitoxin. This molecule is elaborated by a species of the marine plankton *Gonyaulax,* which under favorable conditions can multiply so rapidly that it reaches a density of many millions per liter, creating a "red tide" that turns the surface of the ocean rusty brown by day and luminescent by night. Shellfish feeding in red-tide waters concentrate the toxin in their tissues and so become poisonous to man; fatal paralysis can result from the ingestion of as little as one milligram of saxitoxin. The independent creation of such unusual molecules as tetrodotoxin and saxitoxin on more than one occasion by the adventitious processes of natural selection provides a striking example of convergent evolution at work.

Tetrodotoxin is a complex organic molecule built up of several interlocking rings, at one end of which there is a positively charged guanidinium group. It seems probable that the guanidinium group mimics the similarly charged sodium ion and enables the molecule to enter the mouth of the sodium channel, where it sticks like a fat man caught halfway through a porthole, effectively plugging the channel. To explain why tetrodotoxin acts at such an extraordinarily low concentration, however, one must also suppose the rest of the molecule fits very exactly on the outer surface of the channel. To account for the equally tight binding of saxitoxin, with its two guanidinium groups and different rings, one must suppose the molecule occupies a neighboring but not identical receptor site. It is certain that the toxin receptor is located only at the outside of the channel, because even a strong solution of tetrodotoxin has no blocking effect when it is infused into the interior of an axon. Tetrodotoxin and saxitoxin are quite specific for the sodium channel; neither of them has the slightest effect on the conductance of the potassium channel.

When an axon is exposed to tetrodotoxin or saxitoxin, only one molecule binds to each of its channels, so that the toxins are admirably suited for counting the number of channels in the membrane. This approach was first recognized by John Moore, Toshio Narahashi and Trevor Shaw, working at Duke University, who devised an ingenious bioassay procedure for indirect measurement of the binding of tetrodotoxin to the small axons of the nerve in the leg of the lobster. They found that the sodium channels in this type of nerve were rather few and far between, there being only about 13 of them in each square micrometer of membrane.



NERVE IMPULSE can be represented by changes in the voltage across the axon membrane (*a*) or by changes in the conductance of the membrane to sodium and potassium ions (*b*), both on a time scale of milliseconds. During the impulse the inside of the membrane becomes briefly positive with respect to the outside. After impulse has passed the resting voltage is restored.

This observation was later confirmed and extended to other types of axon by J. Murdoch Ritchie and his colleagues at the Yale University School of Medicine. They employed tetrodotoxin and saxitoxin incorporating atoms of tritium, the radioactive isotope of hydrogen, so that the amount of toxin bound to the membrane could be measured directly. Their results showed that the smallest axons, those of the olfactory nerves in fishes, have the fewest sodium channels per square micrometer of membrane, only about three, whereas the somewhat larger fibers in the vagus nerve of the rabbit have about 25 channels per square micrometer. Working at the Marine Biological Association in Plymouth, Hans Meves and Rock Levinson found that the squid giant axon has a much higher density of sodium channels: as many as 500 per square micrometer. This higher figure fitted satisfactorily with several other lines of evidence, in particular with what would be expected from the magnitude of the sodium gating current. It is not much different from the number of channels calculated by Hodgkin to yield the maximum conduction velocity of the propagated impulse in a 500-micrometer axon.

A second respect in which tetrodotoxin has proved to be a tool of outstanding value to electrophysiologists is in making it possible to get a direct handle on the mechanism of ionic gating. As Hodgkin and Huxley pointed out in 1952, the strong dependence of the ionic permeability of the membrane on the potential difference across it means that the sodium and potassium channels must be controlled by the movements of ionic "gates" that behave as if they have a large charge. At the normal resting potential of about −65 millivolts the majority of the charged gating particles are in their resting, or "closed," configu-

ration and the channels are shut. In response to a shift of the membrane potential toward a more positive value, that is to say toward a depolarization, the gating particles begin to change to their "open" configuration, allowing ions to pass through the channels.

The operation of the gating system must necessarily involve the movement of charged particles in an electric field, and so one would expect the change in the configuration of the gates prior to the opening of the channel to give rise to a small displacement of charge across the membrane. In other words, one would expect to see a transient flow of displacement current immediately preceding the passage of ionic current. Hodgkin and Huxley predicted the existence of this gating current but were not able to measure it for the simple reason that at the time there was no way to prevent the large influx of ions through the sodium channels from obscuring it completely. In the short period for which a single sodium channel remains open during the passage of a nerve impulse about 100 sodium ions pass through the channel; during a long voltage-clamp pulse up to 500 ions might get through. The actual opening of the gate, on the other hand, involves the transfer of no more than four electronic charges within the membrane.

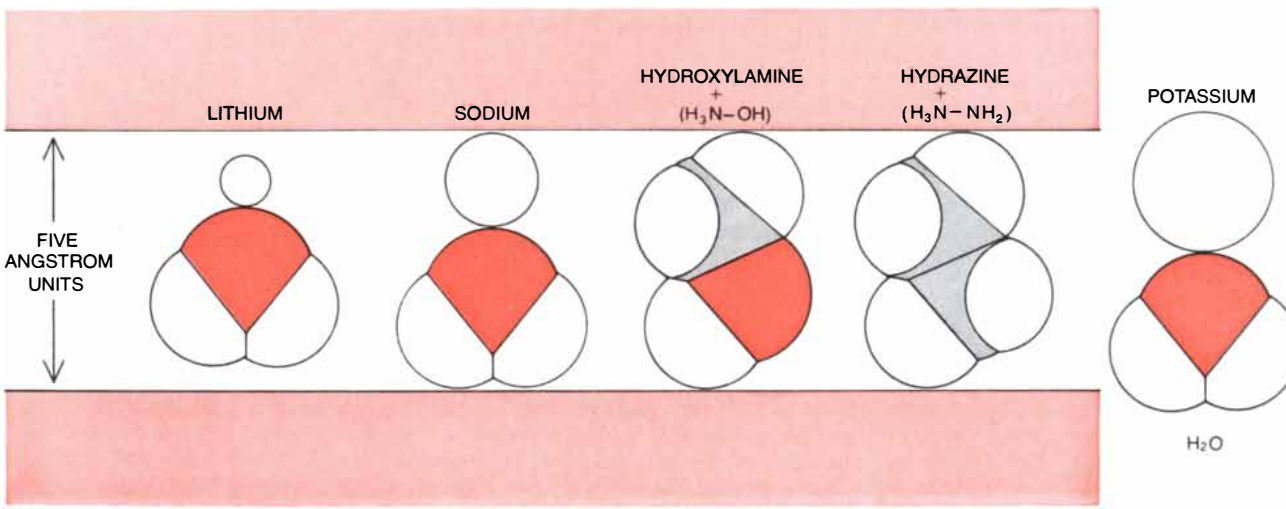
Since the gating current is much smaller than the ionic current, it cannot readily be detected without first tightly plugging the ion channels. With the availability of tetrodotoxin more favorable conditions for recording the gating current could be achieved, thanks to the fortunate circumstance that when the toxin is bound at the mouth of the sodium channel, thereby blocking it, the opening and closing of the gate itself is not hindered and the gating current continues to flow.

If, then, the sodium current in a squid axon is abolished by applying a sizable dose of tetrodotoxin (and also replacing the sodium in the bathing medium with "Tris" buffer), and if at the same time the potassium channels are blocked from the inside by perfusing the axon with cesium or tetraethylammonium ions, the gating current constitutes a substantial fraction of the displacement current recorded when the potential at which the membrane is clamped is suddenly altered. With this approach sodium gating currents were observed for the first time during the 1972–73 squid season by Clay Armstrong and Pancho Bezanilla at the Marine Biological Laboratory in Woods Hole, Mass., and shortly afterward by Eduardo Rojas and me at Plymouth. As in other instances, once the trailblazing observations had been made on the squid giant axon it proved to be not too difficult to extend them to other types of nerve, and successful studies on the gating current at the node of Ranvier in frog nerve have been carried out at, among other places, the University of Hamburg by Wolfgang Nonner, Berthold Neumcke and Robert Stämpfli.

Although with the aid of tetrodotoxin the gating current can be separated from the ionic current without too much trouble, the problem remains of distinguishing between the displacement current that arises from the charging and discharging of the large static capacity of the membrane and the displacement current arising from the movements of the gating particles. It can reasonably be assumed that for the true capacity transient the total amount of charge transferred across the membrane on the application of a voltage-clamp pulse varies in direct proportion to the size of the potential step. Hence the capacity transients for pulses equal in size but opposite in direction will always be precisely symmetrical.

The gating particles, on the other hand, will initially be distributed unequally between their two configurations, and if their total number is fixed, the charge transfer when they change states will exhibit saturation if the pulses are large enough. The amplitude of the gating current will consequently be asymmetrical with respect to the potential. Starting from a negative holding potential at which the gating particles are all in the "closed" configuration, the record of the displacement current for a hyperpolarizing (negative-going) voltage-clamp pulse will consist of a pure capacity transient, since the gates can close no further. The displacement current for a depolarizing (positive-going) pulse of the same size, however, will consist of an equal capacity transient of the opposite sign plus the gating current generated by the opening of the gates.

The method of recording the gating current initially adopted both by Armstrong and Bezanilla at Woods Hole and by Meves and us at Plymouth was therefore simply to add together with a signal averager the current records for equal and opposite voltage-clamp pulses in axons with blocked ion channels, automatically canceling the capacity transient. The technique has since been improved in certain respects by employing an on-line computer in place of the signal averager, but the principle remains the same. The gating current rises to a peak immediately after the potential step is applied, and thereafter it falls exponentially to zero as the gating particles arrive at their new configuration. As is shown in the top illustration on page 135, with increasing pulse size more charge is transferred but the relaxation time decreases.



IONIC SELECTIVITY of the sodium channel is achieved through its dimensions and its electrostatic properties. As is shown here, partially hydrated lithium ions or sodium ions and the small positively charged molecules hydroxylamine and hydrazine have the appropriate dimensions to pass through the sodium channel, which appears to have a cross section of three angstrom units by five angstrom units at its narrowest point. A hydrated potassium ion, however, is too big to pass through the channel. These models are of the space-filling type.

One conclusion from a consideration of the steady-state properties of the gating particles is that, as Hodgkin and Huxley had suggested earlier on the basis of kinetic evidence, the gating mechanism involves the cooperation of three (or possibly four) charged particles in each channel. The individual particles behave as if they carried only one electronic charge, but it could well be that they actually have a number of charged side groups that move only through a small fraction of the electric field. It is not possible to decide from the results how these side groups might be disposed within the particle. On comparing the rate at which the gating particles reach their new configuration with the rate at which the channels begin to allow ions to pass, however, it is evident that the gating-current kinetics cannot be reconciled directly with the Hodgkin-Huxley equations. In order for the data to fit it must be supposed that an "activated" configuration intervenes between the "closed" and "open" configurations, and that after the particles have attained it there follows a finite period of cooperative interaction of the particles before the channels open. During this period no further gating current flows.

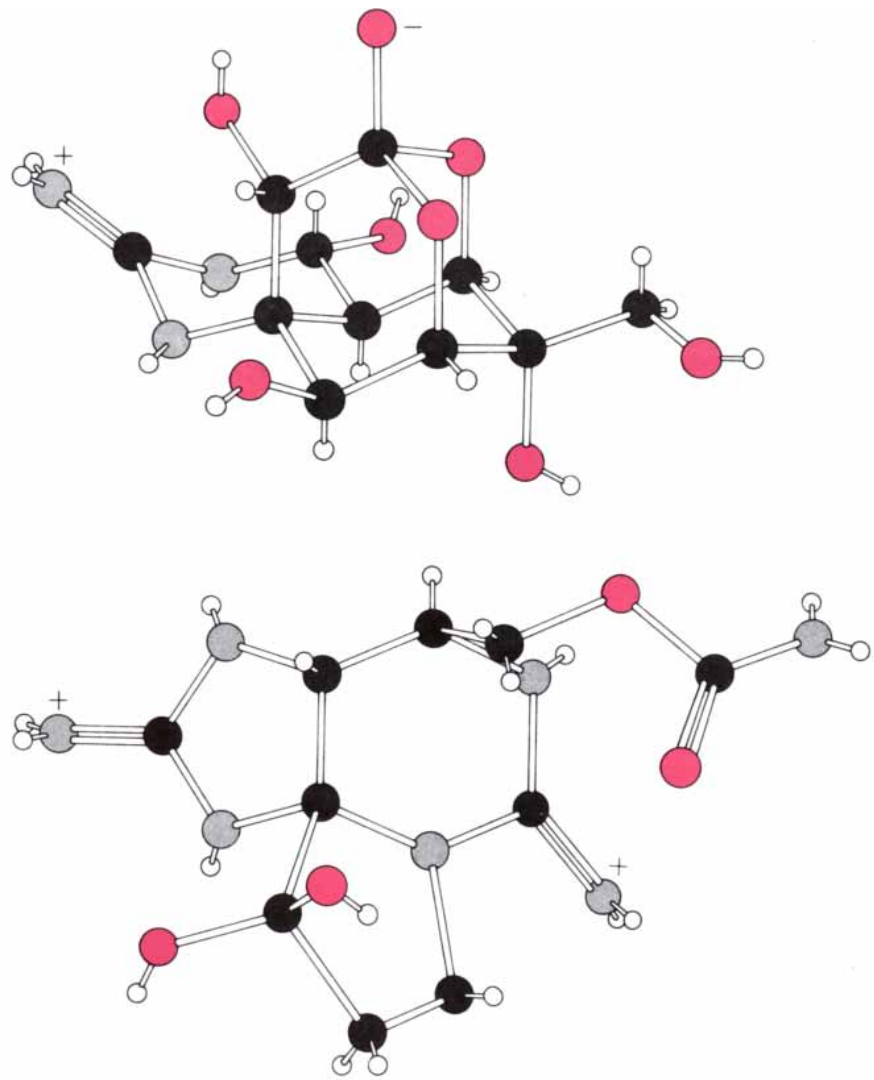
The other main conclusion concerns the mechanism of inactivation of the sodium channels, this being the process that closes the channels again a millisecond or so after they have first opened, even though the membrane remains depolarized. Everyone is agreed on dropping the idea that in addition to the charged particles opening the channels there is a separate set of blocking particles shutting them down with an appropriate time delay. The reason is that the blocking particles would also have to be charged, yet no asymmetrical displacement current that would correspond to their movement can be recorded. Armstrong and Bezanilla have proposed that inactivation involves a further rearrangement of the gating particles that slowly closes the conducting channels. This final configurational change is thought to be accompanied by little or no charge movement, and so it is electrically silent. It apparently leaves the gating particles in a partially immobilized condition, with the result that on the repolarization of the membrane their return to the resting state is too slow to be readily recorded.

The study of the sodium gating current had therefore led to the concept of the voltage-sensitive portion of the sodium channel as a system in which a multiplicity of subunits pass through a multiplicity of configurational states. A mathematical model incorporating the principal findings I have described can be constructed, but it is likely to have a limited appeal for experimentalists because like all such multistate models it embodies too many arbitrary parame-

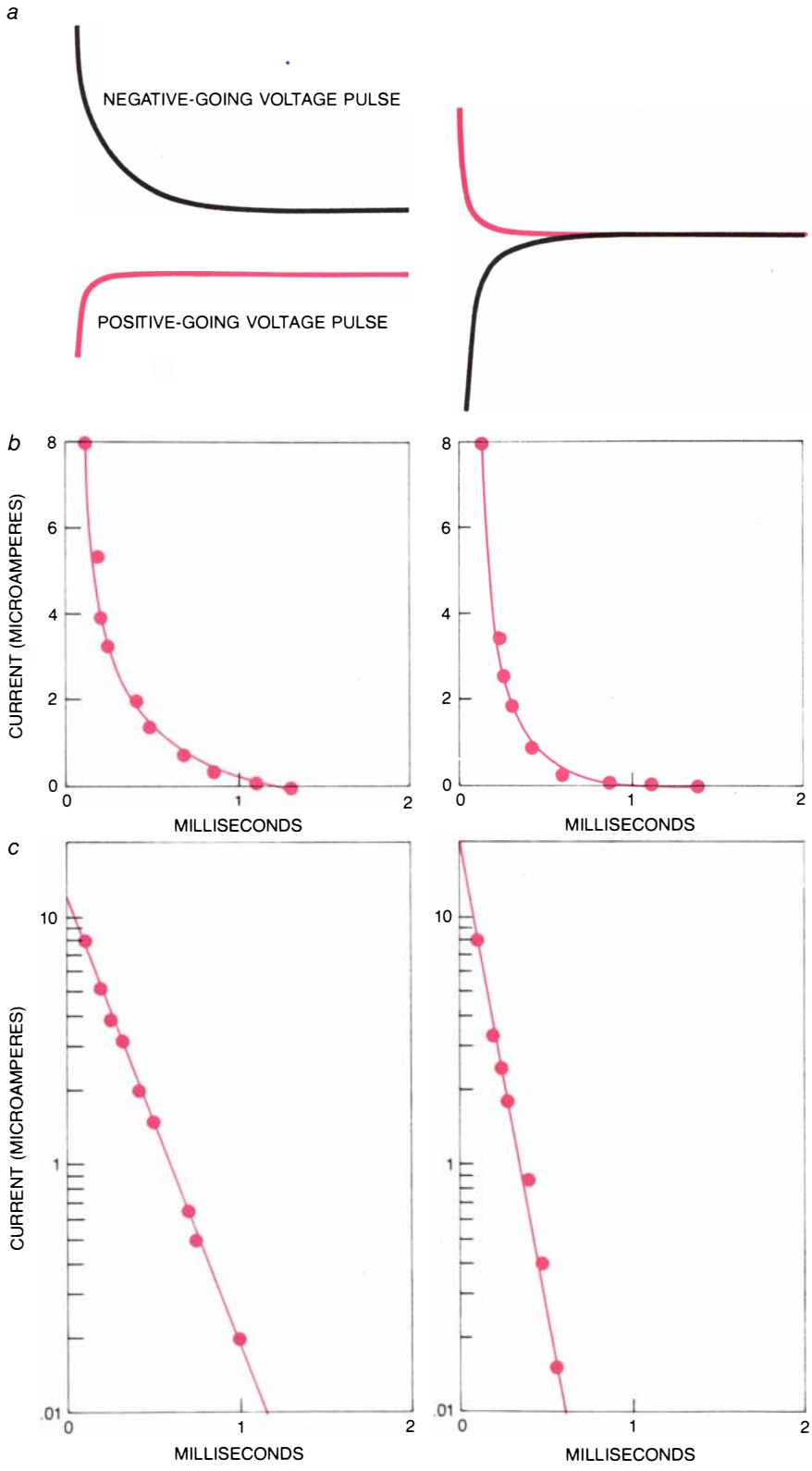
ters to provide a unique solution to the problem. Nevertheless, it cannot be denied that operationally the sodium channels may be (1) resting, (2) conducting or (3) inactivated or refractory, so that like it or not the system must exist in at least three distinct states. Additional intermediate states may also be necessary if the transitions between the main states are to be appropriately timed. It is to be hoped that from further and more accurate measurements of the gating current information about one or more of these transitions will emerge that will characterize them independently of the data derived from the overall kinetics of the system. This is the type of constraint one needs in order to conduct a realistic model-building exercise.

The construction of a plausible molecular model of the sodium channel is even farther from our grasp at the present time. It is clear that the selectivity filter is located near the outer end of the channel, and it seems probable that at this narrowest point the structure is a rigid one, unaffected by changes in membrane potential. The difficulty is to picture the central voltage-gated part of the channel, since there is no precedent to allow even a guess to be made as to how an alteration in the electric field across a large molecule can open up an aqueous crevice through it, let alone one whose opening and closing displays such complex kinetics. Any portrayal of the inner workings of the sodium channel must therefore be strictly fanciful.

My story has so far dealt almost entirely with the sodium channels, but this does not mean that the potassium channels are any less important. For two reasons, however, not nearly as much is known about them. First, there is no blocking agent available for the potassium channels that has anything like the specificity of tetrodotoxin for the



BALL-AND-STICK MODELS of tetrodotoxin (*top*) and saxitoxin (*bottom*), two powerful nerve poisons that bind tightly and specifically to sodium channels, reveal some common structural features. In both molecules a positively charged group mimics the positively charged sodium ion by entering the sodium channel, but the bulky multi-ring structure prevents the toxin from passing through. The two toxins have proved useful for studying sodium channels.

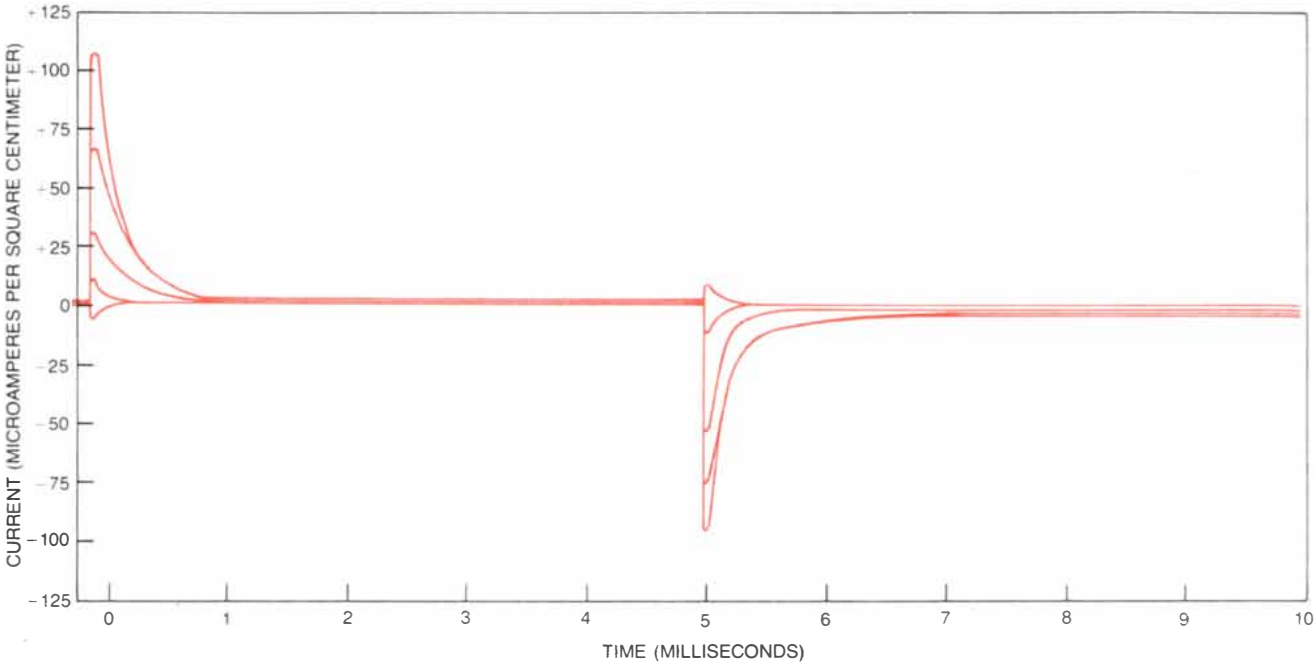


GATING CURRENT is associated with the movement of charged gating particles or subunits within the sodium channel. The subunits open or close the channel in response to shifts in membrane voltage. The gating current can be detected only after blocking the sodium current with tetrodotoxin and the potassium current with cesium, and balancing out the symmetrical capacity transient of the membrane. Under these conditions equal and opposite voltage pulses across the membrane give rise to slightly different current flows; the discrepancy constitutes the gating current. The traces in *a* are single-sweep records of the membrane current for voltage pulses of ±120 millivolts. The difference between them is plotted below on linear (*b*) and logarithmic (*c*) scales. A positive-going (but not negative-going) shift in membrane voltage causes a brief outward gating current that coincides with the opening of the sodium channels. The gating current falls exponentially to zero as the gating subunits arrive at their new configuration.

sodium channels. Although positively charged tetraethylammonium ions can enter the potassium channels from the inside and cork them up quite effectively, a relatively high concentration is needed to produce a complete block. Hence the trick of applying a radioactive label to the blocking molecules in order to count the potassium channels cannot be employed. Second, because the potassium channels open with a delay and about 10 times slower than the sodium channels, the potassium gating current is much harder to detect.

In one respect, however, the slowness of the potassium channels can be turned to advantage, because it results in their contributing the greater part of the electrical "noise" that arises from random fluctuations in the ionic current flowing across the membrane. Examination of the noise spectrum in different types of excitable membrane is a fairly new field of study, and it is one that promises to yield valuable information about some of the characteristics of ion channels that are inaccessible to other approaches. Already, in the hands of Charles F. Stevens of the Yale University School of Medicine, Franco Conti of the National Research Council of Italy and others, noise studies have provided the best estimates we have for the single-channel conductance and for the number of potassium channels in the squid giant axon. It turns out that there are only about 50 potassium channels per square micrometer of axon membrane, or roughly one potassium channel for every 10 sodium channels. The relative paucity of the potassium channels further reduces the probable magnitude of the potassium gating current.

So much for the results of electrophysiological measurements. Since it is now clear that the voltage-dependent changes in the ionic conductances of the nerve membrane are brought about by structural rearrangements within the ion channels, the question arises of whether these changes might be detectable by techniques other than electrical ones. With this in mind, a number of investigators, including myself, have looked for transient changes in several of the parameters that might be associated with the molecular configuration of the gating particles. These parameters have included the optical retardation of the membrane, the amount of light scattered by the axon, the emission and absorption of light after staining the membrane with a fluorescent dye, and the release and absorption of heat by the axon. Although in each case effects could be observed during the passage of nerve impulses or the application of voltage-clamp pulses to the membrane, it was regretfully concluded that they always arose from the response of the membrane as a whole and were not spe-



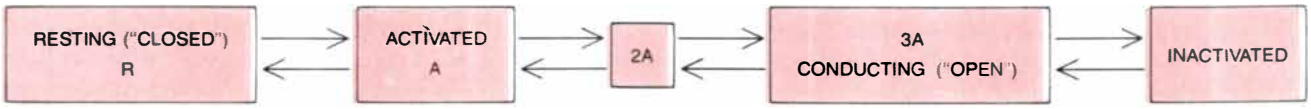
FAMILY OF GATING CURRENTS was recorded in a squid giant axon by the author with George Malachowski and Dirk van Helden at the Marine Biological Association of the United Kingdom in Plymouth. The membrane potential was held at −100 millivolts and was pulsed to −120, −80, −40, 0 and +40 millivolts. When voltage was shifted in a positive direction, the gating current flowed outward.

cifically related to events within the ion channels themselves.

In order to arrive at a complete understanding of the functioning of the ion channels their biochemistry has to be studied. It may reasonably be assumed that, like other large molecules involved in the transport of ions across cell membranes, the ion channels are large proteins. The channel protein is likely to have a central hydrophobic section, which is at home in the oily environment provided by the hydrocarbon tails of the membrane phospholipids, and hydrophilic portions that project into the aqueous medium on each side of the membrane. As with other membrane proteins, this dual nature makes it difficult to isolate and purify the protein for methods of structural study such as X-ray crystallography, because the very process of extracting it from the membrane is likely to alter its shape and inactivate it. Moreover, in order to apply purification techniques such as affinity chromatography a suitable molecule is needed to which the channel protein will bind selectively. Tetrodotoxin would be the logical choice, but until quite recently, when Michel Lazdunski of the University of Nice succeeded in preparing fluorescent derivatives of tetrodotoxin, the molecule has proved hard to manipulate chemically.

The key to success in work of this kind is often to find a starting material already highly enriched in the desired protein, as is exemplified by the exploitation of the electric organ of the electric ray *Torpedo* as a source of acetylcholine-receptor protein by Jean-Pierre Changeux of the Pasteur Institute in Paris. Here another type of electric organ has now come into its own: that of the electric eel *Electrophorus*. Although the electroplates of *Electrophorus* are not nearly as rich in sodium channels as those of *Torpedo* are in acetylcholine receptors, they do not serve too badly in this respect, and a large eel provides a kilogram or more of electric tissue. From studies of the loss of tetrodotoxin-binding capacity in microsomal preparations of sodium channels that had been bombarded with high-voltage electrons, my colleagues Clive Ellory and Levinson arrived at a value of 230,000 daltons for the molecular weight of the channel protein. Working with Michael A. Raftery at the California Institute of Technology, Levinson and others have been able to extract from the electric organ of *Electrophorus* an almost pure tetrodotoxin-binding protein whose molecular weight is of the same order. Although there are still many formidable technical obstacles to be overcome, electrophysiologists can hope that eventually they will be able to study sodium channels that have been incorporated into artificial membranes.

A dim outline of the organization of the ion channels in excitable membranes is therefore beginning to emerge. The further elucidation of the channels' precise structure and mode of functioning will resolve what is undoubtedly one of the most important questions still to be answered about the mechanism of biological excitability.



MULTISTATE MODEL of the sodium channel was proposed by the author on the basis of recent studies of sodium gating currents. Model shows the hypothetical sequence of configurational states through which the gating subunits pass during the opening and closing of the channel. In response to a positive-going shift in the voltage across the axon membrane the gating subunits undergo a configurational change to the activated state. They are then able to interact sequentially, and when all three subunits controlling a given channel have done so, the channel opens for the passage of sodium ions. After remaining open for a short time the channel is closed through a further configurational change that takes it to the inactivated state. The double arrows between the states indicate that each reaction can proceed in both directions. In practice, however, most of the channels pass through the full cycle from the resting to the inactivated state and back to the resting state with each impulse. A measurable gating current accompanies only the resting-to-activated step or its reversal.

Dolphins

These descendants of land mammals that took to sea have a large brain, learn quickly and exhibit a rich vocal repertory. Yet lack of evidence leaves open the question of how intelligent they are

by Bernd Würsig

Because dolphins have big brains, are quick to learn tricks devised by human trainers and exhibit a rich repertory of vocal signals, they are widely reputed to have a level of intelligence unmatched by any other animal and perhaps even the equal of man's. On the basis of observations of dolphins in several of their natural habitats I believe the effort to position them firmly in the spectrum of animal intelligence is premature. In the present state of marine technology it is simply impossible for human observers to spend more than brief and isolated moments with an animal that lives in the ocean and moves rapidly over great distances. When a great deal more is known about the behavior of dolphins, the question of their intelligence will answer itself. At present the most one can say is that dolphins are gregarious herding animals, comparable in their individual and social behavior to more easily observed herding and flocking mammals on land.

Dolphins evolved at least 50 million years ago from land mammals that may have resembled the even-toed ungulates of today such as cattle, pigs and buffaloes. After taking to the sea dolphins became progressively better adapted to life in the water: their ancestral fur was replaced by a thick coat of blubbery fat, they became sleek and streamlined, they lost all but internal remnants of their hind limbs and grew a powerful tail, their forelimbs were modified into steering paddles and apparently as a further aid in steering and stabilization many species evolved a dorsal fin.

More than 30 species of dolphins can be identified. They belong to the suborder Odontoceti, or toothed whales, of the order Cetacea. (The quite similar porpoise, which is often confused with the dolphin, is also a member of the whale family; it is distinguished from the dolphin mainly by having a less beaklike snout and also by its laterally compressed, spadelike teeth. It should be noted, however, that many American students of marine mammals refer to all small odontocete cetaceans as porpoises, regardless of their physical characteristics.) A few dolphin species live in fresh water, but most species have an ocean habitat. The freshwater species travel in small groups or are nearly solitary, whereas the ocean species (such as the Pacific spotted dolphin) may congregate in aggregations of several thousand. Such numbers are reminiscent of buffalo herds in North America and grazing animals on the Serengeti plains of Africa; one wonders about behavioral and ecological similarities between the dolphins and their distant terrestrial relatives.

In general highly social mammals have complex social signals and a rich behavioral repertory. Hence they can interact with other members of their group in sophisticated ways. Examples include signals of aggression that are useful in establishing and maintaining dominance hierarchies, signals for courtship, warning sounds or movements at the approach of a potential attacker and many other signals that contribute to the functioning of the group and its individual members.

The effort to accumulate data on the behavior and social systems of dolphins is made difficult by the fact that most of their communication goes on below the waves. It is extremely difficult to approach a group of dolphins in a boat and to stay with them long enough to begin to understand their social system. All one sees is a group of dorsal fins as the animals surface to breathe, and then they are lost to view as they move on underwater. Even in the rare circumstances when the water is calm and clear and the dolphins can be seen for more than a few minutes, the proximity of a boat may disturb them, so that it becomes difficult to separate what is natural in their behavior from what is unnatural, being merely a reaction to the boat.

As a result of these problems and others most of the early observations involved dolphins in captivity. Although captivity must be an awkward situation for animals that are accustomed to a life with few physical boundaries, the captive dolphins nonetheless yielded useful data simply because observers could watch them for extended periods of time. The best-known work of this kind was done by Margaret C. Tavolga of Marineland of Florida. She observed a group of 12 bottlenose dolphins (*Tursiops truncatus*) in a large tank for a total period of about five years.

Tavolga found that the group had a definite dominance hierarchy. The one adult male, which was the largest animal in the group, was more aggressive and less fearful than any of the females, subadult males or young dolphins. In general the larger animals were dominant over the smaller ones.

Similar data were obtained by Gregory Bateson at the Oceanic Institute of Hawaii. He found a dominance hierarchy in a group consisting of two spotted dolphins (*Stenella attenuata*) and five spinner dolphins (*S. longirostris*). The largest male threatened other dolphins (by lunging at them or showing his teeth) but was never threatened himself. The second-ranking dolphin, also a male, threatened the animals below him, and so on down the line. Bateson's findings showed, as the findings of other workers have shown, that the hierarchy is not as strict as it is among some other mammals. For example, the lowest-ranking male was still able to mate with a female without being challenged by the largest male.

If the dominance hierarchy is not necessarily related to the access of the males to the females (as it is, for example, in the harem of the elephant seal), one wonders what its function is in wild dolphin populations. Kenneth S. Norris and Thomas P. Dohl of the University of California at Santa Cruz have speculated that the function may be to organize the members of the group to deal with a variety of situations. For example, threats and chases by the larger dolphins could cause the smaller females and young animals to be herded into the center of the group, where they would be better protected from such potential predators as sharks and killer whales. Norris believes he has seen such struc-

THREE STAGES OF A LEAP are demonstrated (unintentionally) in the Sea of Cortés (the Gulf of California) off the coast of Mexico by three dolphins of the species *Delphinus delphis*. The repertory of leaps, spins and somersaults executed by dolphins is richly varied.

HERD OF DOLPHINS of the species *Delphinus delphis* was photographed leaping in the Atlantic. Dolphins are social mammals that sometimes congregate in quite large herds, as they are doing here, but more often they are found in subgroups of perhaps 20. Dolphins are also air-breathing animals, so that their leaps serve in part to enable them to breathe. The animals also leap during play and hunting.

turing in spinner dolphins and spotted dolphins.

For such a system to evolve it is helpful and perhaps necessary that the animals possessing it be genetically related. W. D. Hamilton of the Imperial College of Science and Technology and other biologists have argued that closely related animals should tend to protect one another more than they protect distant relatives, since close relatives share more genes. If a mother saves her offspring by herding it inside the group while putting herself in at least some danger, her apparent altruism will be adaptive for the group because a significant proportion of her genes will be preserved.

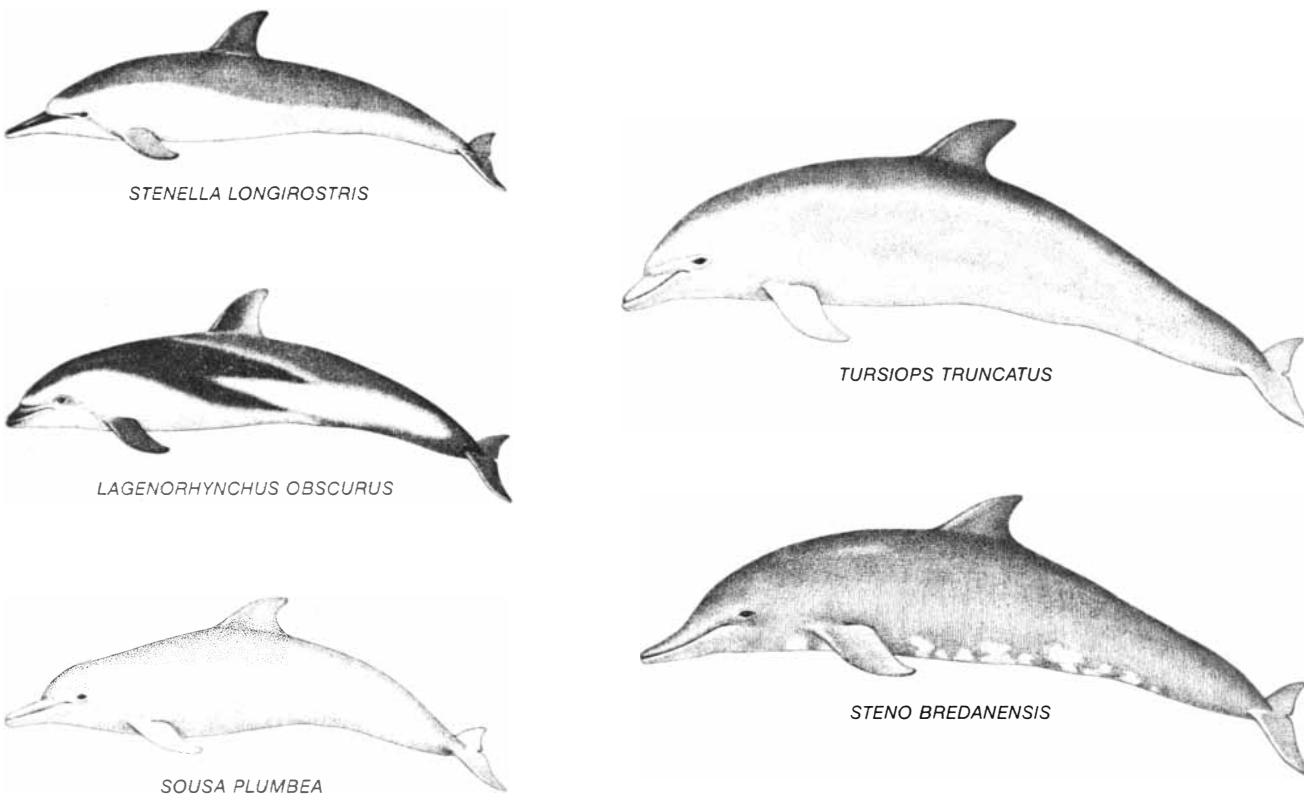
If dolphins herd group members in this way, it is likely that at least some of the animals in a group are related. Dolphins have also been seen to help an ailing member of the group reach the surface to breathe and to protect a group member from predators or other dangers. These patterns of behavior have often been cited as evidence of humanlike altruism and of great intelligence. It appears more likely that they represent an outgrowth of an evolved tendency to help related individuals.

Unfortunately the degrees of consanguinity in a wild dolphin population are not known. Only recently has any idea at all of social structure among dolphins been gained, and the knowledge is complicated by the existence of many different species and many more separate populations within species. From terrestrial mammals it has been learned that the social system represents in part an adaptation to the population's habitat. For example, wolves that live mainly on deer tend to travel in small packs, whereas wolves that hunt moose are found in larger and more highly organized societies. This difference is apparently related to the need for a coordinated effort by several animals to successfully bring down the larger prey.

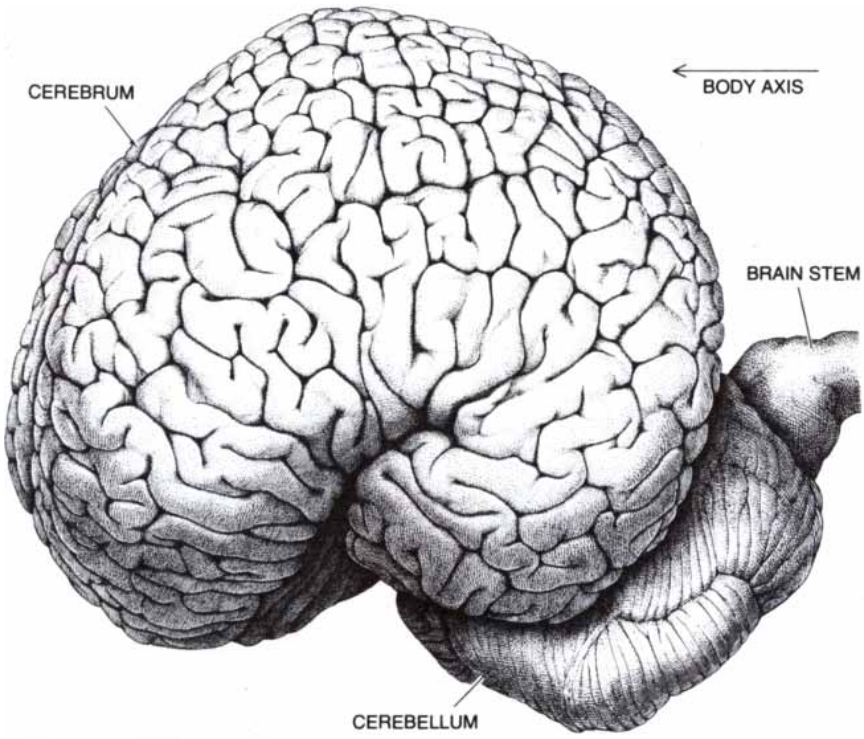
A few excellent studies that make comparisons possible among dolphin species and populations have been done recently. Norris and Dohl studied Hawaiian spinner dolphins from sea cliffs and from underwater. They found that these highly social mammals traveled in schools averaging 25 members. The structure of the school varied during the day in a predictable manner. In the morning the dolphins moved slowly and in tight groups, with individuals almost touching. They appeared to be resting. Later they became increasingly active, swimming faster, with individuals leaping clear of the water in the spins, somersaults and other displays for which dolphins in oceanariums are famous.

At such times the schools became more spread out, with animals often as much as 20 meters apart. Moreover, groups tended to join, so that 50 or more members might constitute the expanded school, with all the animals moving in the same direction. As night approached, the school moved several kilometers away from the shore, entering deeper water and beginning deep dives in order to feed on fish several hundred meters below the surface. The array consisted of many widely spaced groups within an area with a diameter of several kilometers. Because Norris and Dohl were able to recognize some individuals, they found that a given small group of individuals tended to stay together but often shifted as a unit from one school to another.

Norris and Dohl suggested that dolphins may form close groups while they are resting so that they can employ the combined sensory abilities of all the individuals in the school to scan the environment and to detect potential danger. It is well known that dolphins can scan the water by echo location over much greater distances than would be possible by eyesight. (In echo location a dolphin projects high-frequency sounds in short pulses, much as bats do. The sounds bounce off objects, and the echoes give back information on the distance, size, shape and even texture of the object.) Norris and Dohl's hypothesis about combined sensory abilities, particularly during rest, relies in part on the ability of dolphins to get information from echo location. Presumably each dolphin



FIVE DOLPHIN SPECIES are portrayed to indicate their differences in shape and marking. They are the bottlenose dolphin (*Tursiops truncatus*), the spinner dolphin (*Stenella longirostris*), the South Atlantic dusky dolphin (*Lagenorhynchus obscurus*), the Indo-Pacific humpback dolphin (*Sousa plumbea*) and the rough-toothed dolphin (*Steno bredanensis*). All of them are drawn to the same scale. About 25 other species of dolphins are known. Porpoises, which are quite similar, have a blunter snout and are ordinarily shorter and fatter.

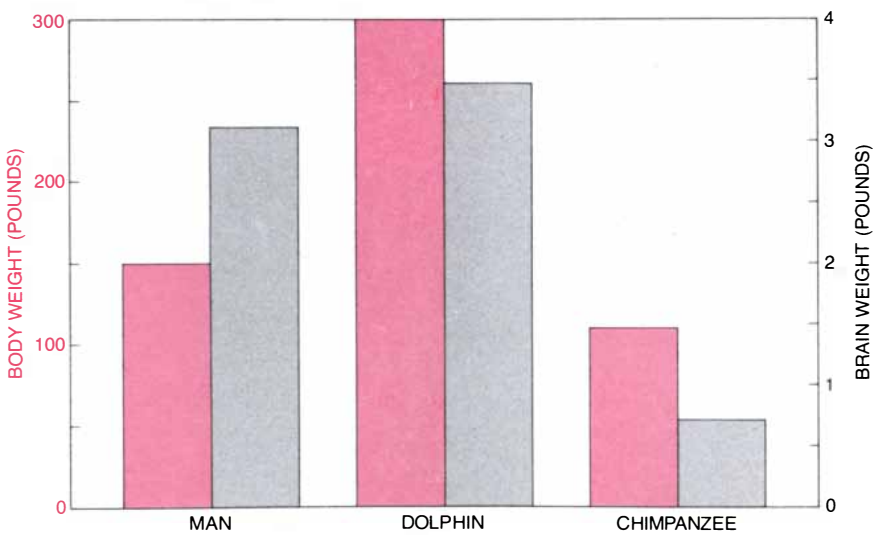


BRAIN OF DOLPHIN is depicted for the species *Tursiops truncatus*, the bottlenose dolphin. A dolphin brain weighs about 3.5 pounds (1.59 kilograms); a typical human brain weighs about three pounds (1.36 kilograms) but is larger than the dolphin's in proportion to body weight.

in a closely organized school can hear the echo-location sounds made by other members of the group. Therefore even though any given individual might not make many sounds, much information about the environment would be rapidly and efficiently disseminated to all. It is also probable that a resting group swims close to shore in order to be in shallow water that is not frequented by large deepwater sharks.

During periods of alertness the spacing of the spinner dolphins increases and the animals do a great deal of leaping. This activity may in part be play, as many people have suggested, but it may also represent a form of communication. A leap is usually followed by a loud slap or splash as the dolphin enters the water. Such sounds travel fairly long distances underwater and may signify the presence of the leaper to others. Indeed, groups of some dolphin species sometimes converge on an active, leaping school from a distance of several kilometers. Nighttime feeding in deep water is also attended by much leaping and loud splashing. At this time the members of a group are quite widely separated, and the assumption once again is that leaping may serve to communicate location and possibly information such as the number of other dolphins nearby and what they are doing.

Graham Saayman and C. K. Tayler of the Port Elizabeth Museum in South Africa studied Indian Ocean bottlenose dolphins (*Tursiops aduncus*), and my wife Melany Würsig and I studied South Atlantic dusky dolphins (*Lagenorhynchus obscurus*). Both species have habitats similar to the habitat of the Hawaiian spinner dolphin. The habitats are coastal-pelagic, meaning that all three populations can often be seen and studied from the shore but that they also move far from the shore, usually to feed. All three populations exhibited similar patterns of behavior and movement. It therefore seems reasonable to say that the habitat of these marine mammals is largely responsible for their way of life.



BRAIN WEIGHT AND BODY WEIGHT are compared for man, the dolphin and the chimpanzee. Man's brain weighs about 2 percent of his body weight, the chimpanzee's less than 1 percent and the dolphin's slightly more than 1 percent. Although rough ideas of comparable intelligence can often be obtained by comparing the ratio of brain weight to body weight, other factors such as the length of the body, the manner in which limbs are used and the complexity of the brain must also be considered. It is not possible to say on the basis of a simple comparison of brain weight to body weight that one kind of organism is more intelligent than another.

This assertion can be examined by studying dolphins in a different environment. Three such environments can be found: deep-ocean, coastal and freshwater. The most thoroughly examined populations have been coastal ones. Susan H. Shane of Texas A & M University observed Atlantic bottlenose dolphins off Texas; A. Blair Irvine, Randall S. Wells and Michael Scott of the University of Florida observed them off Florida; Melany Würsig and I observed them off Argentina, and Saayman and Tayler observed Indo-Pacific humpback dolphins (of the genus *Sousa*) off South Africa. Again major similarities were evident. They can be summarized by a description of our study of bottlenose dolphins off Argentina.

The bottlenose dolphin is a coastal species in many parts of its worldwide range. Hence it can be observed readily from the shore. For a period of 21 months my wife and I observed a school of bottlenose dolphins that passed close to the shore (always within a kilometer) in water less than 40 meters deep. We studied them by observations from coastal cliffs and from a small rubber boat, by underwater recordings of their sounds, by photographing their dorsal fins in order to recognize individuals and by tracking their movements with a surveyor's transit on the shore. (With a transit one can determine precisely the

We reinvented the 35mm camera so others couldn't catch up. They haven't.

When Olympus examined the design concepts of traditional 35mm SLR's, they recognized the need for an engineering revolution: achieve quality without bulk, sophistication without complication. The result is called Olympus OM-1. The state-of-the-art SLR embodying the latest technology in electronics, metallurgy and optics. So advanced that others, while attempting to copy its compactness, have failed to duplicate its sophisticated internal design. • OM-1 is lightweight, yet professionally-rugged. Compact, but with a viewfinder 70% brighter and 30% larger than conventional SLR's. And its special air dampers make shooting exceptionally quiet and vibration-free. All this, part of the most complete, compact SLR system, with almost 300 lenses and accessories to meet every photographic challenge. No wonder OM-1 is Number 1.

For the photographer demanding the last word in automatic exposure control, the one choice is the Olympus OM-2. A quantum leap ahead of traditional SLR design, incorporating the features of the OM-1 plus electronic sensor circuitry found in no other camera's light measurement system. It's called OTF ("Off-the-Film") light measurement, because it measures the light reflected off the film plane and the film itself, during actual exposure. The result is a range of capabilities found in no other camera in the world, including automatic electronic flash whose exposure is controlled by the OM-2's internal light sensors. And each frame is exposure-controlled with motor drive (5 fps!) and rapid winder (over 3 fps!). Learn more about the incomparable choice offered by the OM-1 and OM-2, detailed in our free brochure. Write OLYMPUS, Woodbury, New York 11797.

OLYMPUS

OLYMPUS OM-1 OLYMPUS OM-2

Free. The kit that has shown hundreds of companies new ways to control travel time and costs.

Chances are, last year you travelled less but spent more. It's a simple fact: business travel costs are getting out of hand.

But help is just around the corner.

It's the Beechcraft Business Flying Kit. And it's already helped hundreds of successful companies find new ways to put a lid on the skyrocketing cost of business travel.

This straight-forward and easy-to-read kit is designed to do one thing. Help you, step-by-step, realistically evaluate whether a company airplane can help solve your travel problems.

Inside the kit, you'll find up-front answers to the real questions you have about owning a company

airplane. How do you determine the need for one? Will it be a good investment? Who will fly it?

And this year, the kit has been up-dated and expanded with new information that makes objectively evaluating a company airplane even easier.

You'll also get a look at dozens of creative ways many companies use business airplanes, like the Beechcraft Bonanza shown here, to reduce their travel costs—while doing more business in more places.

Get some help with the vexing problems of business travel.

Send for the Beechcraft Business Flying Kit today.

Write on your company letterhead to: Beech Aircraft Corporation, Dept. Y, Wichita, Kansas 67201. Ask for our free Business Flying Kit, and please mention if you're a pilot. If you'd rather call, make it collect and ask for Dick Schowalter, Jr. 316-681-7072.

Member General Aviation Manufacturers Association

Beechcraft AVIATION CENTER

The Beechcraft Bonanza. A 4 to 5 seat airplane that can whisk you in style and comfort to almost anywhere you could ever want to go. All at speeds over 200 mph.

Know any friends who want to learn to fly? Tell them about the General Aviation Manufacturers Association **Take Off** Sweepstakes. They could win a $50,000 airplane just for earning their private pilot license. Have them call TOLL FREE, 24 hours a day, any day and ask for the BEECH "TAKEOFF" operator: USA 800-447-4700 (in Illinois, 800-322-4400); Canada 800-261-6362 (Toronto, 445-2231).

location, movement and speed of wild animals. The method was developed by Roger Payne of the New York Zoological Society and has proved to be useful in studies of the coastal movements of marine mammals.)

The school ranged in size from eight to 22 dolphins; the average was 15. Some individuals stayed together consistently; others were at times absent. Thus the population was made up of subunits of shifting membership.

Eventually our sightings were all of animals we had identified previously. We concluded that we knew all or nearly all 60 or so dolphins that at one time or another were part of the group. Although all the animals in the larger group interacted to some degree, we never saw the entire group together at one time. It is therefore likely that other subgroups, also composed of individuals known to us, traveled near other shores. Apparently they cover a large range; we once sighted a subgroup of known animals more than 300 kilometers south of our base. It is probable that the size of a subgroup represents an optimum balance among the number of animals needed for defense against predators, the number needed for efficient feeding and the number best suited for social interaction, reproduction and the survival of young.

In the fall, winter and spring the bottlenose dolphins seemed to feed on schools of anchovies at about midday in water from 15 to 35 meters deep. At such times they advanced as a spread-out school, each animal being separated from the next one by as much as 25 meters. After several minutes in this formation they began to dive and mill around in one area. From comparison with the behavior of dusky dolphins, which in this region feed almost exclusively on anchovies, we deduced that the bottlenose dolphins were herding schooling anchovies to the ocean surface and feeding on them there. The spread formation before feeding probably serves the purpose of acoustically scanning as large an area as possible in the search for anchovies.

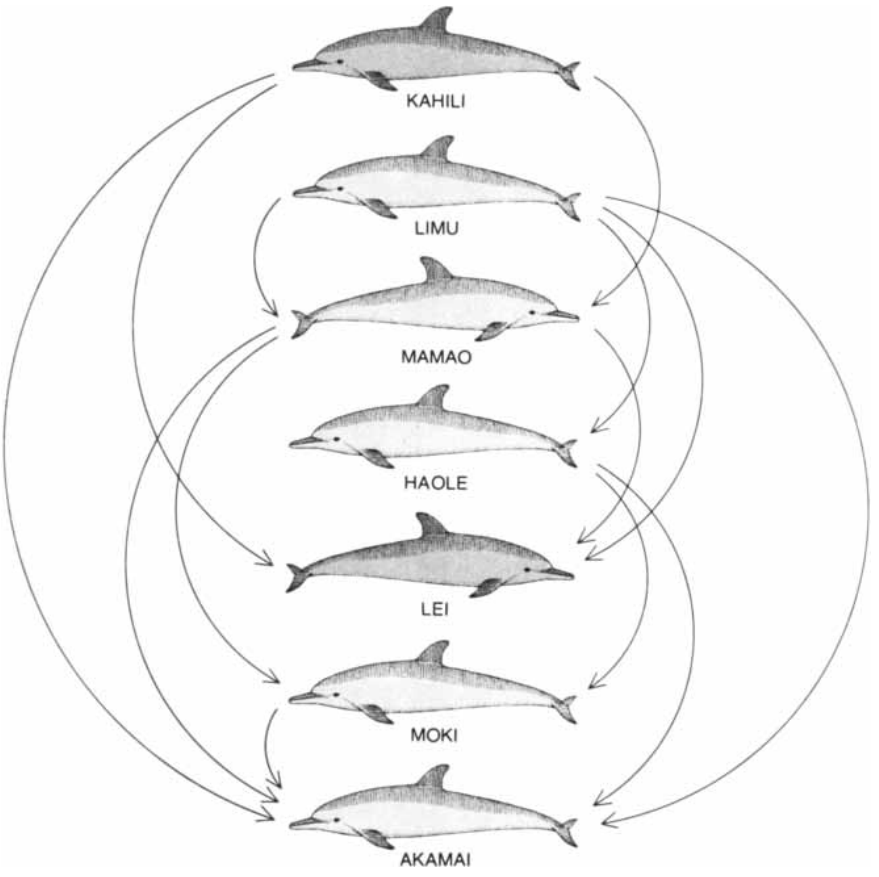
This cooperation among members of a school is quite different from what the dolphins do in the summer. Anchovies are not present in the waters off Argentina during the summer, and so the dolphins feed mainly on large solitary fishes living among rocks near the shore. At such times the dolphins move in water that is from two to six meters deep, and they spread out in a line that is longer than it is wide, with every animal at essentially the same depth and as close to the shore as possible. In this formation individual dolphins nose their way among the rocks and poke into crevices as they search for their prey.

Although the average size of subgroups was 15 dolphins, the average was lower (14) during the summer than it was during the winter (20). It can be argued that fewer dolphins are needed for successful individual feeding near the shore. Indeed, it is possible that the resource—large fishes inhabiting crevices—is limited and that a group of about 14 dolphins in one area at one time represents the upper limit of the area's carrying capacity. The limit is higher when dolphins are feeding cooperatively on schooling fish.

Other reasons for the seasonal fluctuation in the size of dolphin subgroups must also be considered. For example, one could argue that the size of subgroups increases because dolphins are more susceptible to killer-whale attacks when they feed farther from shore and that with more animals per subgroup they can protect themselves better. Perhaps the fluctuations reflect seasonal peaks in mating and calving. On the basis of the information now available, however, food seems to be an important determinant of the size (and presumably the composition) of subgroups of bottlenose dolphins.

What can be said about the internal structure of subgroups? The fact that individual dolphins shift about from one subgroup to another suggests that the animals have what is known among mammalogists as an open society. Among terrestrial mammals the African chimpanzee (*Pan troglodytes*) exhibits a quite similar system. Certain African ungulates (hoofed mammals) also have a degree of openness, with individuals moving frequently from subgroup to subgroup within a more rigidly defined herd. Among chimpanzees the variations appear to be in response to the availability of food; the animals search for food in small units but aggregate into larger units when the food has been found. Melany Würsig and I have found a similar situation among dusky dolphins, but it is not yet clear to what extent a patchy distribution of food might govern variations in the size of schools of bottlenose dolphins.

Notwithstanding the tendency to openness, groups of five or six dolphins were almost always observed traveling together. One such group included a notably large animal and a second adult that traveled with a calf during the entire 21-month study. I therefore believe the second adult was a female. The large



DOMINANCE HIERARCHY in a group of captive dolphins studied by Gregory Bateson at the Oceanic Institute of Hawaii is depicted. The group consisted of two spotted dolphins (*dark gray*) and five spinner dolphins (*light gray*). In this drawing males face to the left, females to the right. The arrows indicate threats (threatening entails showing teeth or lunging) made by one animal to another. The largest male, Kahili, and the male Limu were never threatened. Although Akamai, the smaller male dolphin at the bottom of the hierarchy, never threatened any other member of the group, he copulated with the female Lei on at least one occasion.

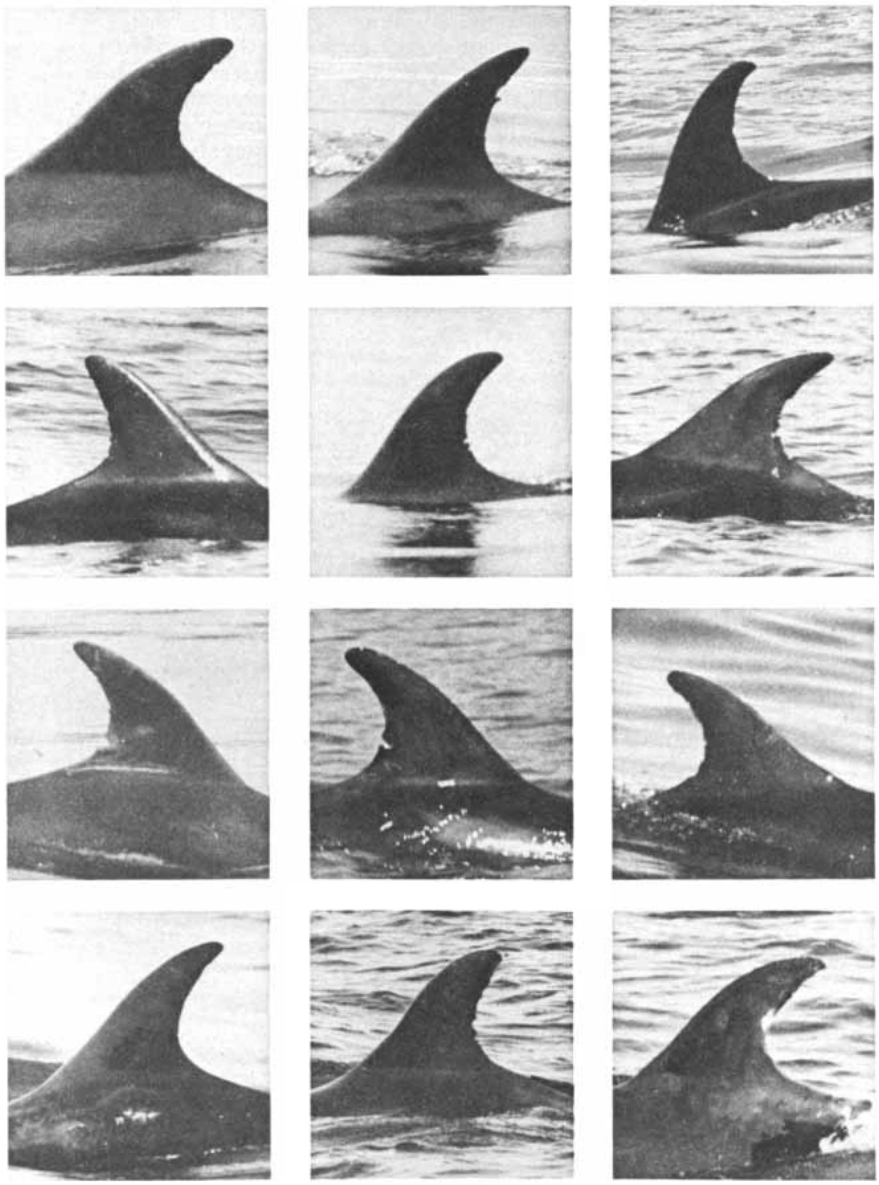
animal was the only one that consistently slapped its tail against the surface of the water when a boat approached. Experienced dolphin trainers say that tail slapping can signify anger, aggression or warning. It is tempting to say that the large dolphin was in some sense the leader of the group, but I do not have enough information about the social system and behavior of dolphins to substantiate the hypothesis.

Two other groups of five and six dolphins that we always found together consisted of adults of approximately the same size. It is possible (again I conjecture) that the groups were made up of nonbreeding members of the population, much as bachelor herds of elephants travel together.

The most detailed findings on the age and sex composition of dolphin schools have come from the work of Irvine, Wells and Scott on Florida bottlenose dolphins. They captured 47 dolphins and put tags and other identifying marks on them in order to recognize them later in their natural habitat. As they tagged the animals they were able to determine the sex and size of the individuals.

Once the dolphins had been released Irvine, Wells and Scott found that the home range of the resident herd covered about 85 square kilometers. Females and calves often traveled in groups that included only a few adult males or none. Such males tended to associate more often with calfless females than with mothers and young, and they rarely associated with subadult males. The subadult males were at times found in bachelor groups far from the other dolphins. Several females were seen with their calves for as long as 15 months. Hence a strong social bond exists between mother and calf, probably continuing long after weaning. No such long-term association between a male and a calf has been observed.

It should be emphasized that the social relations of dolphins are not clear-cut or immutable. They are highly variable. Nevertheless, a few major features are apparent. The bonds between mother and calf are strong; the bonds between male and female and male and calf appear to be less so. This comparison suggests that mating is somewhat promiscuous. Subadult males may be excluded from the normal social routine but subadult females are not, which suggests that adult males may copulate more with various females than females do with different males. Such a relation is indicative of a polygynous mating system, which is also common among terrestrial mammals.

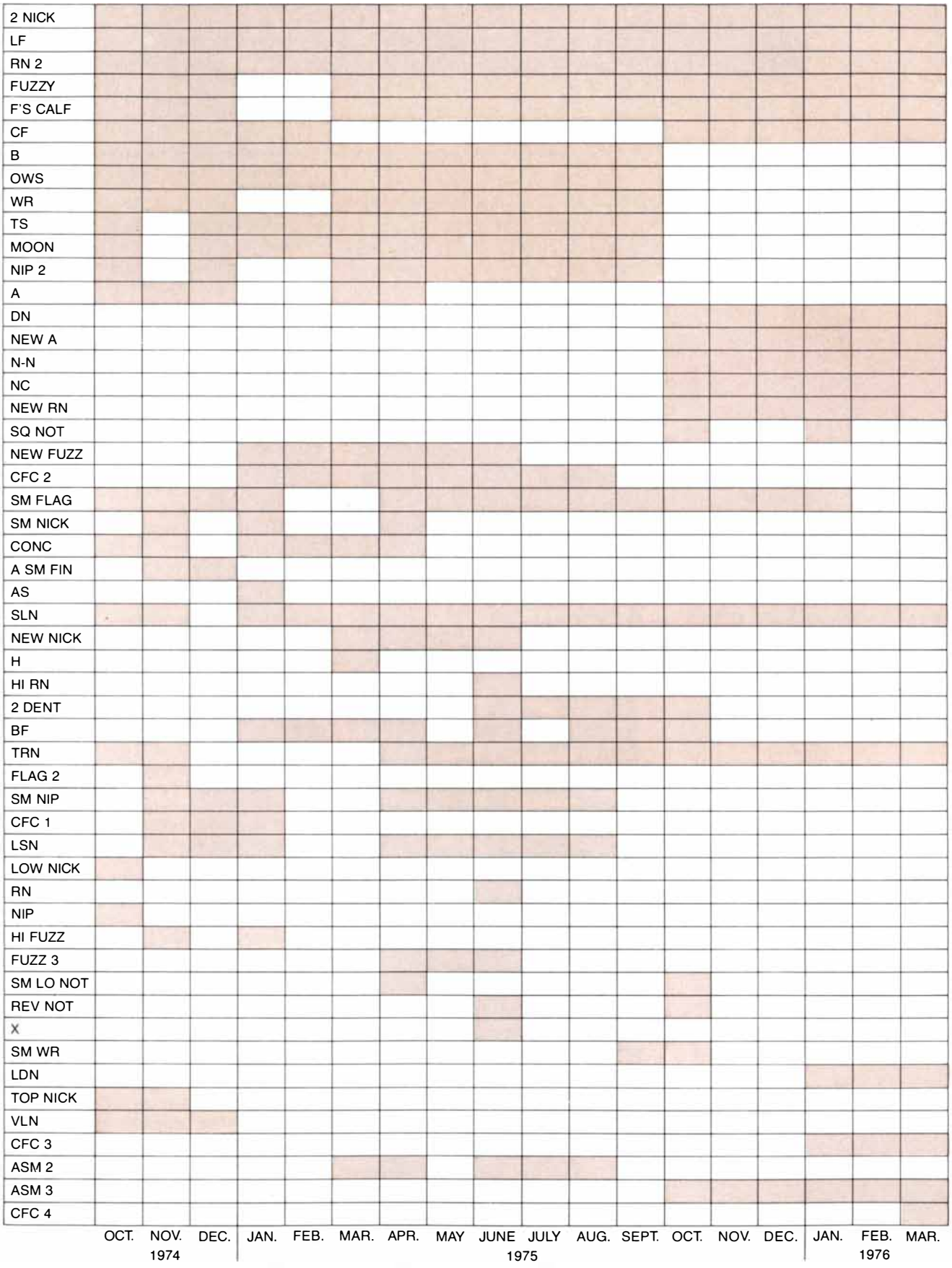


MEANS OF IDENTIFICATION of individual dolphins is provided by the pattern of nicks and scars on the trailing edge of the dorsal fin. These photographs represent a sampling of 12 bottlenose dolphins from a group of about 50. All individuals in the group could be identified.

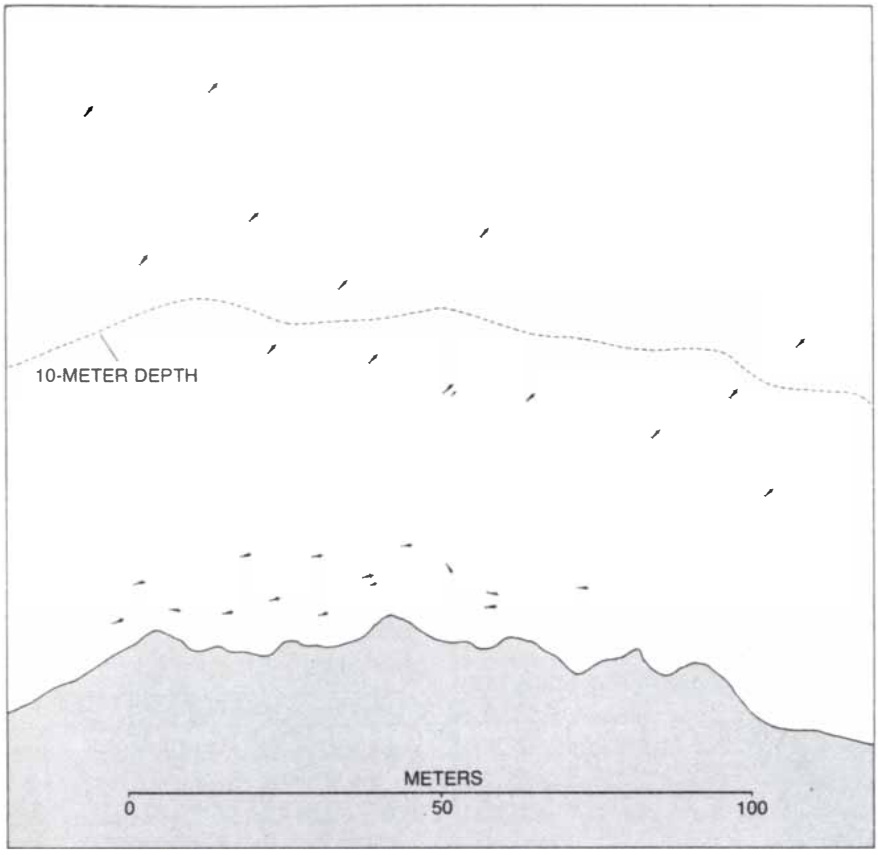
Less can be said about the behavior of dolphins than about their social relations, even though a considerable body of accounts of behavior has been built up from observations of dolphins in nature and in captivity. I believe not enough is yet known to support any firm and broad statements. Still, a few major examples of behavior can be cited. Bottlenose dolphins (and other species) appear to engage in courtship and copulation throughout the year, as is often indicated in the wild by belly-to-belly swimming. Yet bottlenose dolphins and some other species have a definite yearly calving peak (sometimes two peaks). Among the dolphins we observed off Argentina all the calves were born in the summer. This finding indicates that a physiological change in the male or the female causes conception to occur in a limited period. Such a change has been documented in seasonal increases in the weight of the testes in the males of several dolphin species.

Yearlong mating also implies that courtship may have more than a sexual connotation. Several investigators have suggested that such interactions may also serve to define and strengthen social hierarchies and bonds. The argument is reinforced by the frequent homosexual activity seen at least among captive dolphins. Future studies may show that "homosocial" might be a better term. A carry-over of sexual signals to dominance hierarchies is seen in many other mammalian groups.

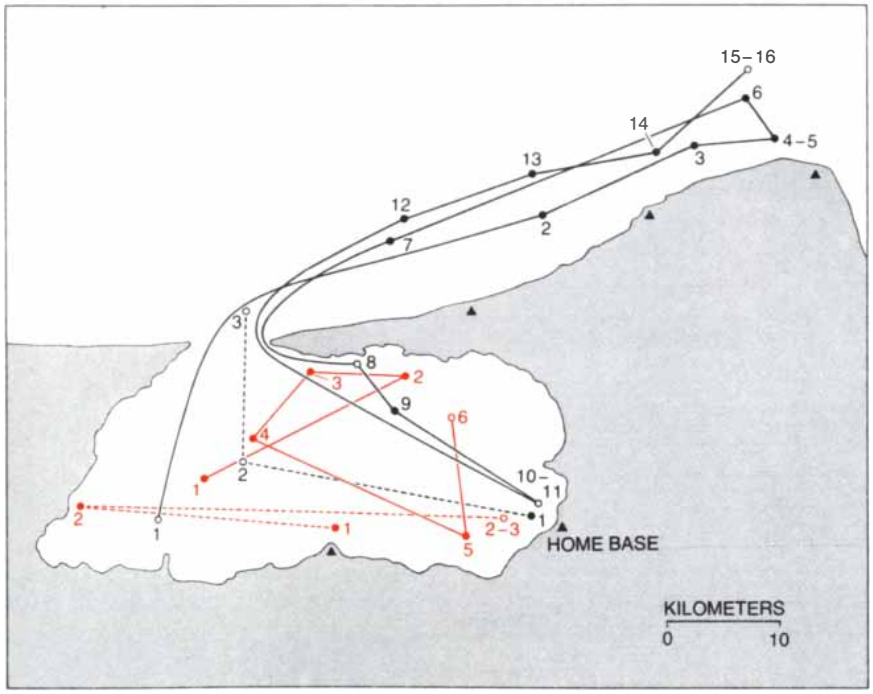
A second behavior found in almost all dolphin species is leaping. I have mentioned that it tends to occur most often when animals are widely separated and so may have a communicative function. Bottlenose dolphins off Argentina leaped far less than dusky dolphins in the same vicinity did, even when both species were hunting fish in essentially the same manner. The bottlenose dolphins, however, moved in one school



GROUP COHESIVENESS of dolphins is indicated by this chart of the presence and absence of 53 known bottlenose dolphins off the coast of Argentina over a period of 18 months. The designations at the left are abbreviated names the observers gave the dolphins. A bar opposite a designation indicates that the individual was seen near shore at least once during the corresponding month; a blank space means the animal was not seen during that month. A dolphin shown as having been seen may have been seen more than once in the month.



GROUP FORMATIONS of bottlenose dolphins off the coast of Argentina varied according to whether the animals were in shallow water close to shore or in deeper water farther out. In shallow water the dolphins were individually hunting rock-dwelling fishes, whereas in deep water they functioned as a group to find schools of anchovies and to herd them to the surface.



RADIO TRACKING of four dolphins off Argentina (in the Gulf of San José and the adjoining open ocean) produced these patterns of movement. Each dolphin is represented by a different track, and the number beside each circle shows the animal's position that number of days after it was fitted with a radio. Solid circles represent known positions, obtained either by triangulation from the shore points indicated by triangles or by approaching the dolphin in a boat, and open circles represent estimated positions. The animals' daily locations are given as of midday.

and so would have little cause to communicate with other members of the school by leaping. The dusky dolphins moved in as many as 30 small schools in one vicinity, and we often saw schools coalesce when leaping began. Leaps undoubtedly have other functions, such as helping to herd or catch prey, but the function of communication among members of a particular species may be important.

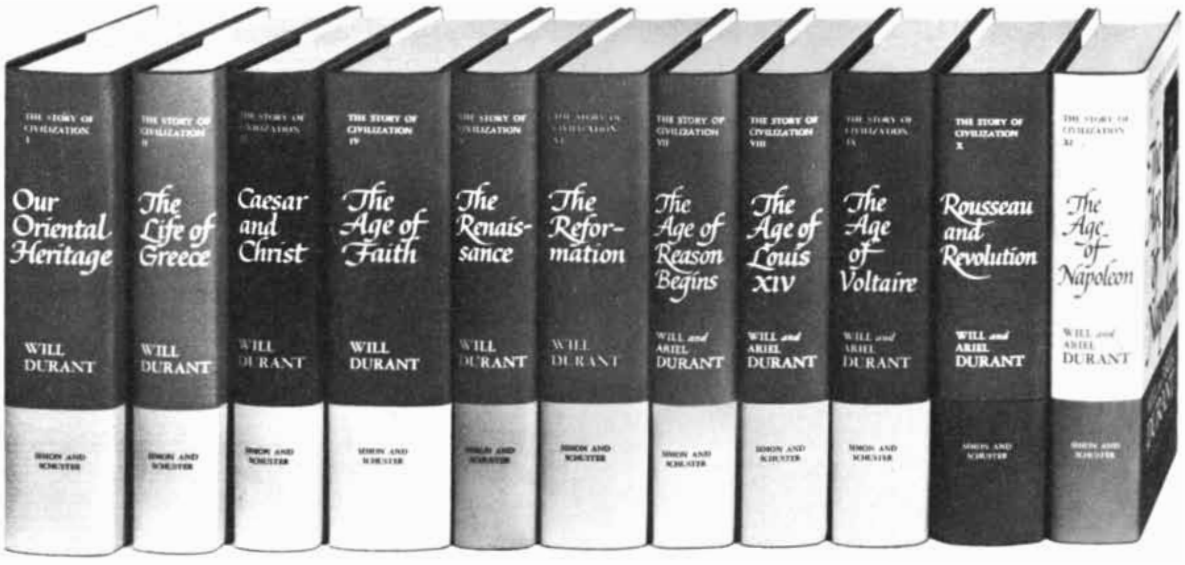
On the basis of the rather modest body of knowledge about dolphins, what can be said of their intelligence? Dolphins are certainly adept at learning complex tasks, as they demonstrate in their tricks in oceanariums, and they remember the tasks for years. They also have been shown to be capable of relatively abstract thinking. For example, Karen Pryor, working at the Oceanic Institute in Hawaii, trained rough-toothed dolphins (*Steno bredanensis*) to perform a new trick for a reward of fish. After several days of training they exhibited ever-different types of leaps and contortions, apparently "realizing" that the forms of behavior they had displayed previously would not be rewarded. Still, various trainers have pointed out that the same thing happens among dogs and other mammals and possibly even among pigeons, which implies that one does not have to invoke superintelligence to explain what the dolphins are doing.

Edward O. Wilson of Harvard University has suggested that the brain of the dolphin may be larger in relation to the size and weight of the body than the brains of most other mammals because of the same reputed imitative abilities that have made dolphins such a favorite with animal trainers and the public. The question then is: Why should an animal be such a superb imitator? R. J. Andrew of the University of Sussex has noted that vocal mimicry may be important to animals that often travel out of sight of one another. Individuals of a widely spaced group could then recognize other members because of an elaborate convergence of signals among the animals of one group or herd. This system, in the form of dialects, has been shown to operate among some primates and birds. It is plausible that the system evolved to an even greater degree among dolphins, which rely heavily on sound.

What about mimicry of movement? Wilson suggests that individual dolphins may imitate the members of the group that are most successful at catching fish and avoiding predation. Furthermore, it is advantageous for animals in social societies that cooperate to hunt food (as has been shown in at least some dolphin species) to know one another's movements well and for individuals to be able to take several roles in the herding of a school of fish. Wilson argues that imita-

Rediscover the past in The Story of Civilization.

"Our century has produced no more successful attempt to narrate the whole common story of mankind"—*Clifton Fadiman*



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.

XI. The Age of Napoleon. France's domination of European history, from the fevers of the French Revolution to Napoleon's defeat at Waterloo. A history of European civilization from 1789 to 1815.

Will and Ariel Durant's masterpiece is now offered as your first benefit of membership in the Book-of-the-Month Club.

The complete 11-volume set: **$20**

Publisher's price: **$241.10**

You simply agree to buy 4 books within a year.

Since the first volume of this remarkable work appeared forty-four years ago, it has been hailed both as a great literary achievement and an epic history of mankind. To read it is to enjoy the past made live with excitement and power. To own it is to have at your fingertips a reference work of undisputed authority. It is hard to imagine a more valuable set of books for your home library.

Beginning with this special opportunity to own *The Story of Civilization* at the extraordinary saving of more than 85% (including shipping and handling charges), you'll enjoy continuing benefits on the best and most important books published today. Every Club book is as good as, or better than, those sold in stores. As a Club member, you select from the best and you never settle for the altered or inferior editions that some book clubs send their members.

And when you remain a member after the short trial period, each book you buy earns Book-Dividend® credits. These entitle you to enormous savings on current art books, reference sets, literary classics, records, and children's books. Join now, and shop for books the civilized way—in America's Bookstore.

BOOK-OF-THE-MONTH CLUB®

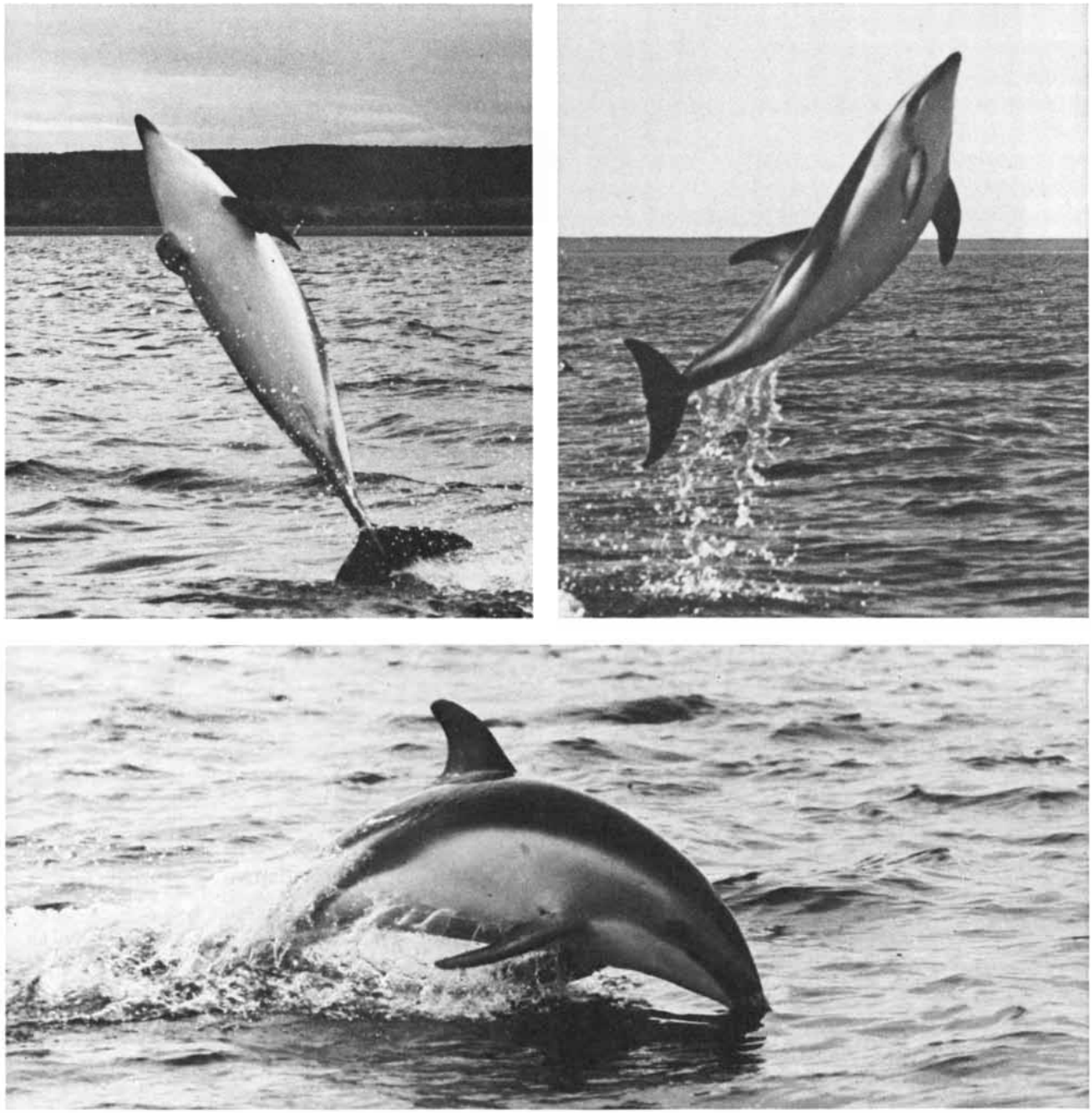
Camp Hill, Pennsylvania 17012

AMERICA'S BOOKSTORE®

tion alone is enough to explain the size of the dolphin's brain and that the social signals of dolphins are probably no more sophisticated than those of most mammals and birds. In my opinion not enough is known about the social signals of dolphins to provide a basis for such a statement. Norris views the imitative powers of dolphins as not being necessarily better than those of many other mammals with brains that are smaller and physically less complex. It seems futile at present to compare the intelligence of dolphins with that of other mammals simply because of the lack of appropriate information about dolphins and the great differences between their environments and those of terrestrial mammals.

The need for better information about dolphins turns one's mind to better means of obtaining it. One possibility is to try to habituate dolphins to observers to such an extent that they will go about their daily activities as if the observers were not present. George Schaller studied mountain gorillas in this way, and Jane Goodall similarly opened a new era of work on chimpanzees. They moved with the animals and sat patiently until the animals either accepted them or simply ignored them.

How might one follow a group of dolphins in the ocean? Perhaps it is not necessary. Jody Solow of the University of California at Santa Cruz recently learned to make a sound underwater that at times called individuals of a group of nearby Hawaiian spinner dolphins to her. Her achievement opens the possibility that an investigator could eventually recognize all the members of a group, learn their social patterns and interactions and gain a better idea of their natural behavior.



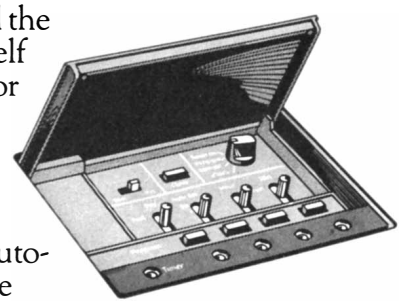
VARIETY OF LEAPS performed by dolphins is suggested by these photographs of three dusky dolphins (*Lagenorhynchus obscurus*) in the open ocean. Dolphins in captivity have been observed to increase the variation in their leaps when they are receiving rewards of food.

RCA announces SelectaVision 400.

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

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

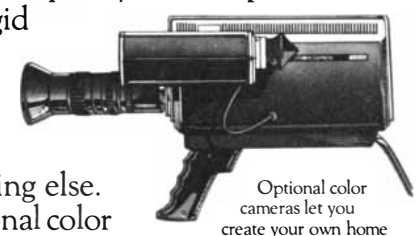
Now, simply by touching a few buttons, you program your selections into the timer of the incredible new SelectaVision 400. The rest is automatic: The 400 will turn itself on at kickoff time, silently record the game, then turn itself off. When it's time for your second selection, the 400 turns itself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



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

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

It's all there in the new 400.



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

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

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

Let RCA turn your television into

SelectaVision.™

For more information on the complete line of RCA SelectaVision Video Cassette Recorders, write to RCA Consumer Electronics, Dept., 27-212 SA, 600 North Sherman Dr., Indianapolis, Ind. 46201.

RCA

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

Ancient Catapults

Surviving Greek and Roman texts reveal the remarkable level of mathematical and engineering skill that went into the development of these early ballistic-missile launchers

by Werner Soedel and Vernard Foley

In 399 B.C. Dionysius the Elder, ruler of the Greek colony of Syracuse in Sicily, prepared his city for a long war with Carthage by undertaking a research-and-development program. Utilizing such now-familiar techniques as the assembly of large teams of specialists, the division of labor to break down the tasks into manageable units, and the provision of financial and psychological incentives, Dionysius clearly aimed from the outset at the production of novel weapons. Out of the program came quadriremes and possibly quinqueremes, ships with the equivalent of four or five banks of oars and so with more potential power behind their rams than the standard three-bank triremes. Dionysius' engineers also devised the first catapults.

These early machines probably fired arrows from a bow not much stronger than one a man could draw. By mechanizing the drawing and releasing of the arrow, however, the catapult inventors made possible the construction of much more powerful bows. These devices appear to have been built of composite materials, with a wood core surmounted by a tension layer of animal sinew in the front and a compression layer of horn in the back. Eventually the flexible bow reached the limits of its design, and it was superseded by catapults based on the torsion principle. In this approach tightly stretched bundles of elastic fibers were further strained by a rigid bow limb as the weapon was brought to a full draw. Horsehair or human hair could be used for the ropes that made up these bundles, but for superior performance animal sinew was preferred.

To mechanize the archer's motions the catapult engineers incorporated a number of important design features. The basic piece in the catapult was the stock, a compound beam that formed the main axis of the weapon. Along the top of the stock was a dovetail groove, in which another beam, the slider, could move back and forth. The slider carried on its top rear surface a claw-and-trigger arrangement for grasping and releasing the bowstring. In front of the claw on top of the slider was a trough in which the arrow lay and from which it was launched. In operation the slider was run forward until the claw could seize the bowstring. Then the slider was forced to the rear, taking the string with it, until the bow was fully drawn. In the earlier versions linear ratchets alongside the stock engaged pawls on the slider to resist the force of the bow. Later a circular ratchet at the rear of the stock was adopted. Forcing back the slider on the first catapults was probably done by hand, but before long the size and power of the machines called for a winch.

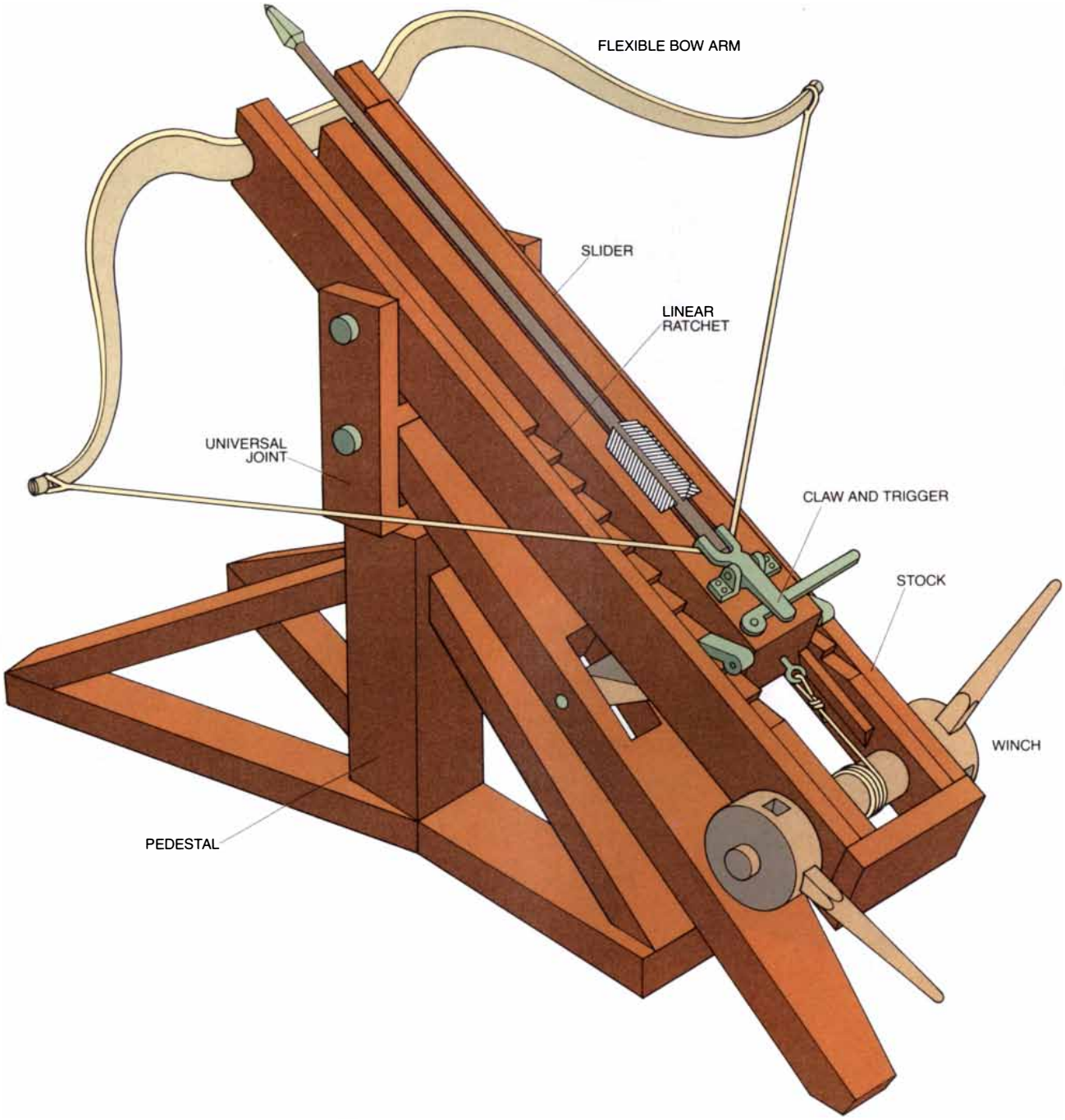
As soon as the catapult became too large to be fired from the shoulder it was placed on a pedestal. To facilitate aiming a special joint was devised to connect the stock with the pedestal. The solution to this problem anticipated the invention of the universal joint, which is usually attributed to either Girolamo Cardano or Robert Hooke, and hence to the 16th or the 17th century. Similarly, the sliding dovetail surfaces of the slider reappear everywhere in the construction of modern machine tools, and the use of a claw to replace the action of the human hand established a tradition of mechanical manipulation that has led to the robots under development today.

Most of this technological sophistication was passed over in silence by the more literary of the ancient Greeks and Romans. Even in more recent times classical scholars did not pay much attention to the surviving catapult texts. Not until the 20th century, when scholars who combined an engineering background with military experience began to decipher the ancient catapult treatises, was their importance made clear. The pioneering field studies of actual weapons constructed according to these texts shortly after the turn of the century by the German artillery officer Erwin Schramm stimulated a line of inquiry that has culminated in recent years with the definitive works of the British historian E. W. Marsden. This literature makes it plain that the Greeks were far from being as disdainful of close observation and exacting experiment as is usually supposed. Plato may have been contemptuous of the failure of objects in the real world to measure up to the ideal dimensions of geometry, and Aristotle may have based his biology in good measure on merely verbal descriptions of species, but within the community of ancient mechanical engineers methods of assessing nature of considerably more importance for the future were being developed.

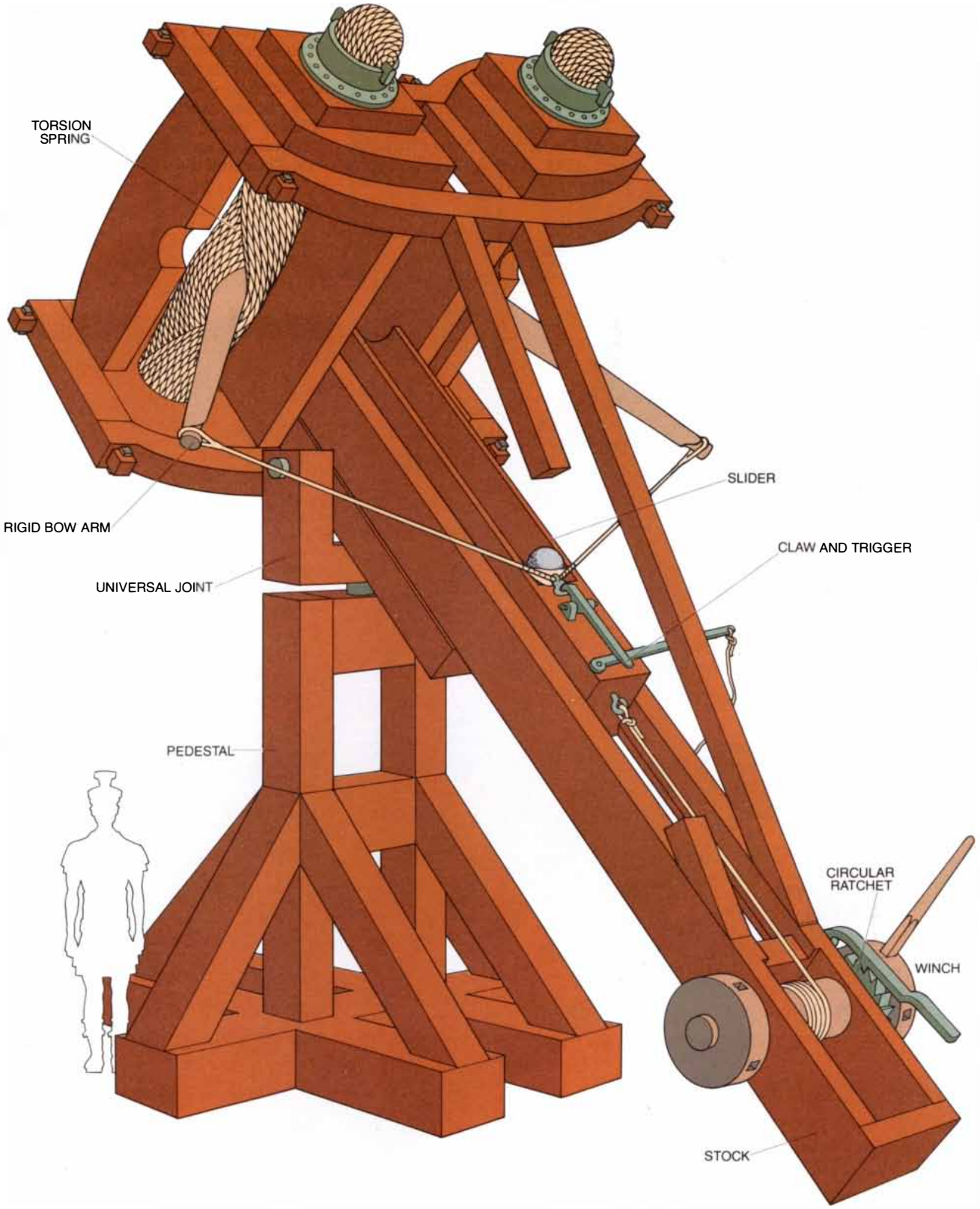
The replacement of the flexible bow by the torsion spring gave a great boost to catapult engineering. That advance was made roughly half a century after the invention of the catapult, and there is some evidence that Philip of Macedon, the father of Alexander the Great, was the ruler who subsidized this next phase of research and development. Certainly it was with the campaigns of Alexander that very powerful catapults first appeared. Indeed, there is reason to associate the rise of large empires with the advent of the catapult.

The flexible-bow catapult had been limited to comparatively small arrows or stones, the stones requiring a bowstring with a pouch at its midpoint. The arrows might have been as large as a light javelin and the stones of a size small enough to be hurled overhand. When the torsion principle was perfected, it became possible to fire a stone weighing as much as 78 kilograms. Indeed, the Roman military engineer Vitruvius gives dimensions for catapults firing stones as heavy as 162 kilograms, although such giant machines may never have been actually constructed. More typical machines fired balls weighing from 13 to 26 kilograms. Arrow shooters firing shafts nearly four meters long now appeared. Even with catapult missiles of a more human scale the archer or slinger found himself completely outranged. The longest recorded range for a catapult firing an arrow of the ordinary size, about 70 centimeters, was about 640 meters, and there is some reason to believe the claim was not inflated. The maximum range for an archer was about 450 meters.

Catapults were able to fire such projectiles with considerable accuracy. Their fire could easily be concentrated on a single spot with repeated hits, knocking away the protecting battlements on top of a city wall or detaching the armor on a mobile siege tower. It was possible to aim catapults during the day, when the fall of the missile could be observed, and then terrorize the enemy by firing at intervals into the same spot at night. At the siege of Avaricum in 52 B.C. Julius Caesar noted that his catapults had no trouble shooting down Gaulish warriors one after another as they stepped into a highly exposed position that was vital to the progress of the Roman attack. All these details point to a high degree of intrinsic accuracy. (Indeed, when one of Schramm's reconstructed catapults was test-fired in the presence of the Kaiser it reportedly split one of its arrows with a subsequent arrow, in the best Robin Hood style.) Clearly mechanization had far outstripped the capabilities of human archery. The result must have been dismay-



FLEXIBLE BOW was mounted at the end of a long wood framework enclosing a dovetail slider in this early arrow-firing catapult, based on a design originally devised by technicians working for Dionysius the Elder of Syracuse in the fourth century B.C. The movable slider, carrying the bowstring with it by means of a claw-and-trigger arrangement, was held to the rear of the stock against the force of the bow by a linear ratchet after being pulled back with the aid of a circular winch. The piece connecting the catapult to its pedestal appears to have been an ancient version of the universal joint. The bow itself probably consisted of three different materials glued together: a wood core, a front layer of animal sinew and a back layer of horn. Since sinew is strong in tension and horn in compression, such bows would have been much more powerful than the ordinary kind carved out of a single piece of wood. The arrow is roughly two meters long.



TORSION SPRINGS enabled the ancient catapult engineers to design much larger weapons, such as this Roman stone-throwing version, which launched a stone weighing one talent, or 26 kilograms. (A front view of a similar device appears on the cover.) A pouch woven into the center of the bowstring holds the stone, and a ring attached behind the pouch is grasped by the trigger claw. The washers at the ends of the torsion bundles in this particular model could be rotated and then pinned in place to adjust the tension before firing.

ing for those who found themselves facing such machines.

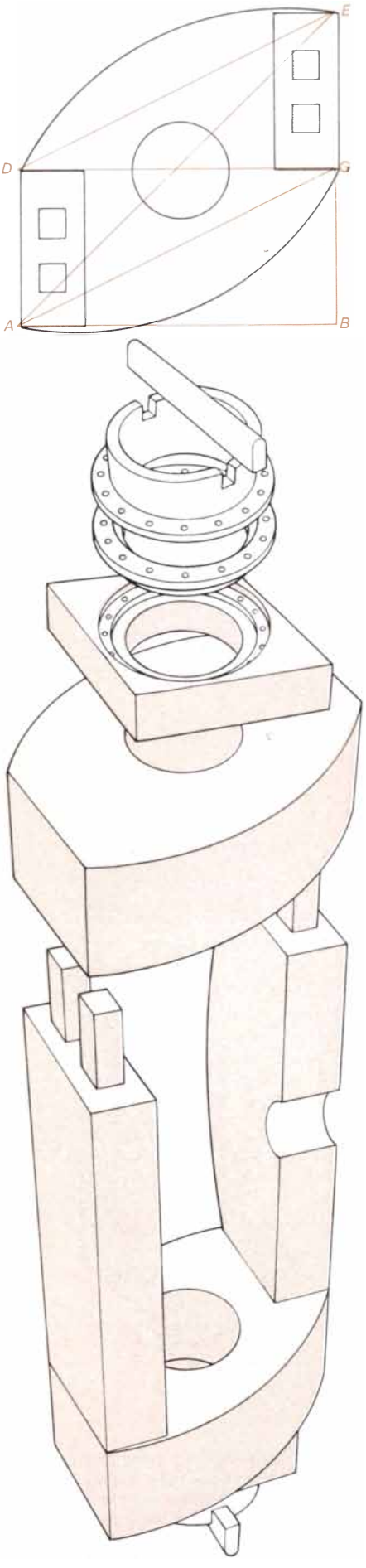
At the siege of Jerusalem in 63 B.C., Josephus, the commander of the Jewish forces defending the city, recorded that the head of a friend standing beside him on the wall was struck off completely by a Roman catapult ball. Even at ranges approaching 400 meters one of these balls could apparently smash through several ranks of soldiers before bouncing to a stop. At the same siege, according to Josephus, a pregnant woman was killed by a ball and the fetus was hurled 100 feet. The long, heavy arrows were equally effective. Advancing troops might be literally nailed to the ground by their descent. In fact, the word catapult is derived from the machine's penetrating power. Lightly armed Greek troops carried a shield called a *peltē,* and the prefix *kata* denotes downward motion. Thus a catapult is a device that can smash downward completely through a shield. At the siege of Gaza in 332 B.C. Alexander was wounded in the neck by a catapult arrow that had pierced both his shield and his breastplate. Archimedes' engines are known to have inspired terror during the siege of Syracuse by the Romans from 213 to 211 B.C. A typical reaction of the time was that of the Spartan general Archidamus, who watched a catapult being fired and then exclaimed prophetically: "Oh Hercules, human martial valor is of no use any more!"

In sum, catapults significantly affected the direction of warfare and with it the equilibrium of politics and society. Broadly speaking, they shifted the advantage in favor of the offense. Until the time of the catapult besiegers were almost always at a great disadvantage. The Trojan War supposedly lasted for 10 years, and the struggle between Athens and Sparta certainly lasted for a quarter century, notwithstanding the great superiority of the Spartan army. In those days most cities that fell did so because of treachery.

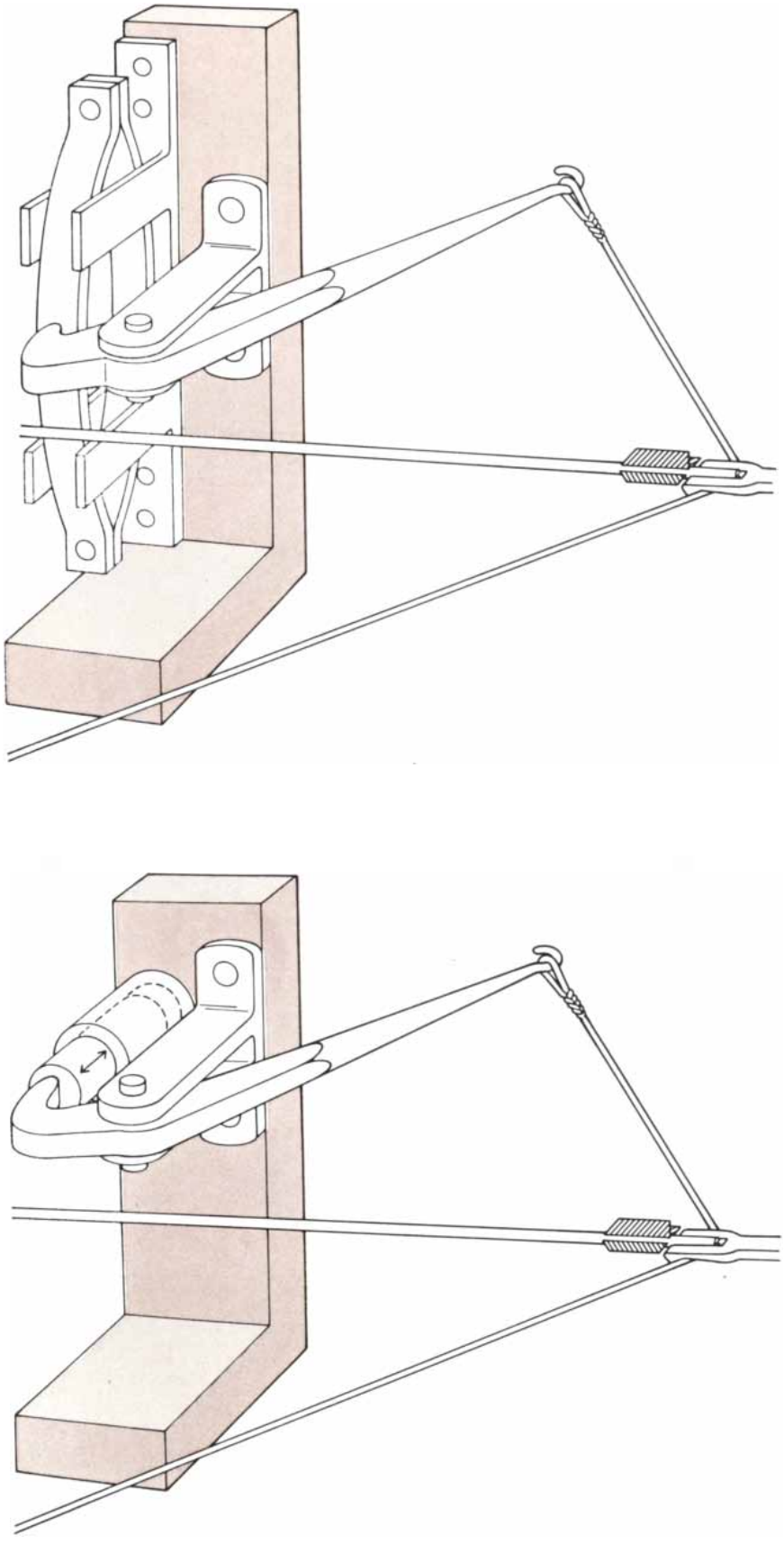
With the introduction of catapults, together with other war machines just coming into use in the West, sieges became more effective. First Dionysius of Syracuse and then Philip and Alexander of Macedon employed tall, mobile siege towers that could overlook a city's walls and pour a withering fire down on the battlements, while enormous rams, powered by as many as 1,800 men, pounded the wall from below. Catapults could not compete with such rams in power, but they were able to knock down walls that were not properly constructed. Philo of Byzantium, in an artillery manual written in about 200 B.C., stated that a wall had to be at least 4.62 meters thick to withstand catapult stones and that it was a good idea to keep the stone throwers at least 150 meters distant by means of ditches and other obstacles.

Even with proper walls the battlements projecting above them remained vulnerable. Because the battlements had to be kept thin to provide a good field of view it was easy to knock them away with stone balls, leaving no shelter on the walls for the defenders. With the returning fire neutralized, the rams and the crews undermining the wall could work with less interference.

The catapult played a key role in making urban life in the fourth century B.C. significantly more precarious. During his first five years in power Alexander captured five major cities and many smaller ones. A passage in the *Politics* of Aristotle (Alexander's tutor) reflects the change. Rational town planning, with straight streets intersecting to form quadrilateral city blocks, had just been popularized in Greece by the architect Hippodamus. Aristotle objected that at least part of every city should preserve the haphazard arrangement of earlier times to make it more difficult for invad-



DIMENSIONS OF A BOARD forming the top piece of one of the torsion-spring frames in a large stone-throwing catapult were specified by the catapult designers in terms of the dimensions of the vertical sides of the frame, which in turn were determined by the diameter of the cord bundle forming the torsion spring. The thickness of the top board is not known for certain, but it was probably equal to approximately the diameter of the cord bundle. The catapult builders appear to have proceeded by first laying out a rectangle with one side equal to the depth of the vertical frame piece and the other side equal to twice this length. They then drew the diagonal of the rectangle, from *A* to *G,* and from *D* they drew a line parallel to the diagonal. Line *BG* was next extended to intercept the parallel from *D* at *E.* The center point of the parallelogram *ADEG* was located and a circle was drawn around it equal to the size of the cord bundle. (In the finished piece the circle actually defined a hole housing the cord bundle.) The arcs *DE* and *AG* were then drawn, each with a radius equal to three times the diameter of the cord bundle. Finally the tenon holes were centered appropriately in the straight edges of the piece and sunk to about two-thirds of its thickness. Thus the catapult engineers had advanced their design procedures to the point where they incorporated automatic scaling methods in their instructions for catapult construction. Once the site and mission of the weapon had been selected the size of the projectile could be determined, and after that was specified the catapult formula would give the size of the torsion bundle that was needed. With the diameter of the bundle known, the construction manuals, incorporating decades of careful testing, would finally yield the sizes of the major parts of the machine as multiples of the diameter of the cord bundle. In the actual construction critical parts of the wood frame would be reinforced with ironwork.



TWO ALTERNATIVE DESIGNS for catapults were introduced by Ctesibius of Alexandria in the middle of the third century B.C. to replace the standard torsion-spring device. Both approaches incorporated rigid bow arms pivoted close to the inner ends. When the bow was drawn, the inner ends of the arms bent in such a way that they pressed either on bronze springs (*top*) or on pistons sliding in airtight cylinders (*bottom*). Neither scheme provided a force comparable to that of the torsion bow, however, and the attempted improvements came to nothing.

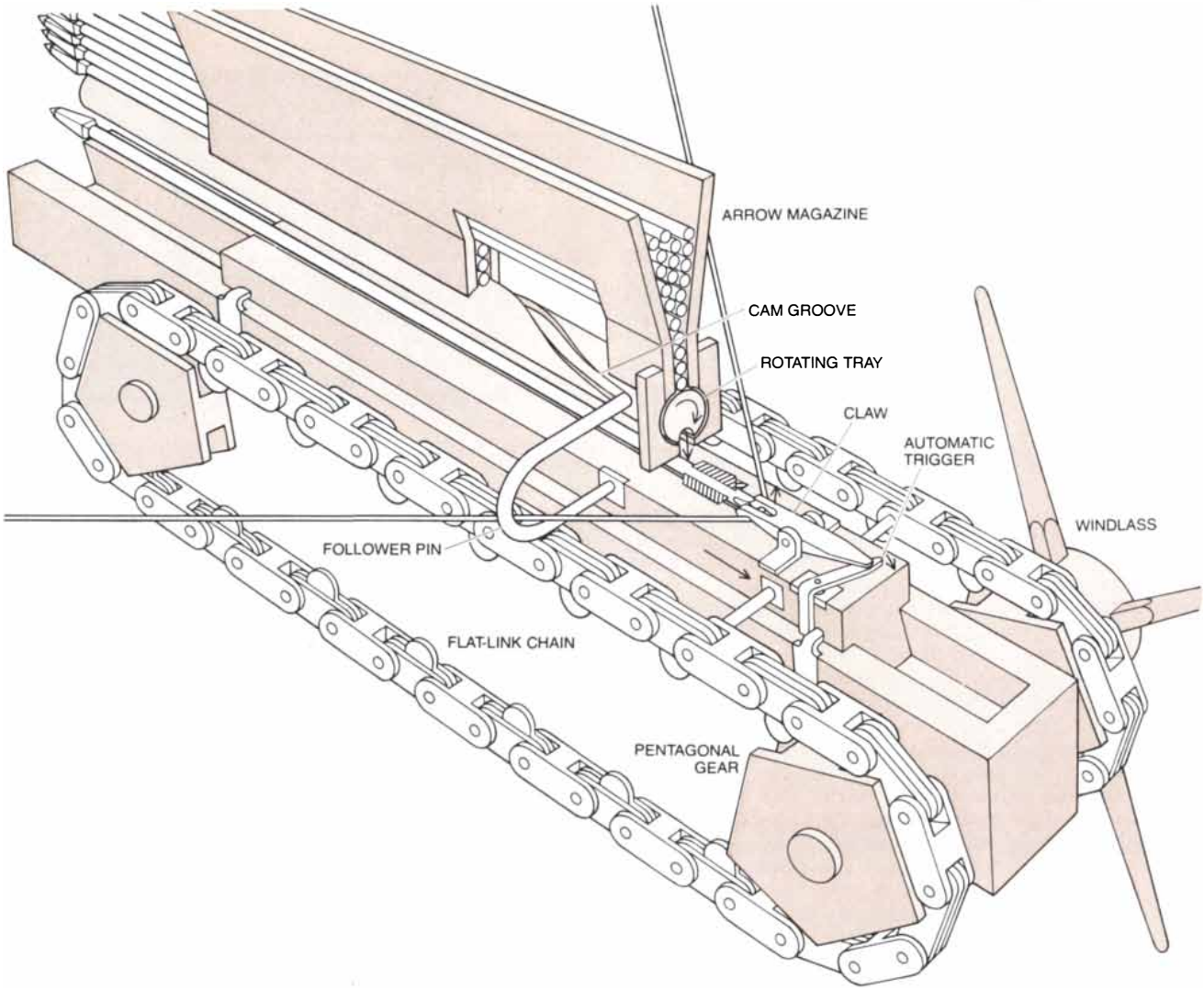
ers to fight their way in. Moreover, he wrote, the design of walls and their careful maintenance was particularly important at that time, "when . . . catapults and other engines for the siege of cities [have] attained such a high degree of precision."

Aristotle's counsel was not enough. The small democratic city-state was dying at the hands of the new technology. As its walls became vulnerable it was swallowed by large absolutist empires such as Alexander's. Democratic field warfare had stressed identically armed spearmen standing shoulder to shoulder in a line. Warriors had a standardized role, and one man could easily replace another in that role. This equality on the battlefield, together with the similar equality of the rowers in the battle fleet, had proved a strong prop of the democratic city system.

With the onset of specialized military engines the equality of arms was lost. Special mathematical and technical skills were necessary to build and maintain a catapult, and the risks involved in operating it were less than those of the rank and file. As a result hierarchies of specialists with particular functions and prerogatives began to appear. In time a political arrangement well suited to a level of technological development in which human muscular force played a significant part was found wanting as individual citizen-soldiers began to yield their priority to machines.

In the new arrangements the ancient engineer tended to benefit. In earlier times his status had not been high. Of all the Olympian gods only Hephaestus, who labored at the forge and made cunning works of metal (including robots), was portrayed as being dirty, ugly and lame. The sculptor Phidias, who played a superintending role in the construction of the Acropolis under Pericles, was accused of sacrilege for daring to incorporate his own portrait in his work. Later, however, matters improved for such artisans. The names of Dionysius' inventors are not known, but from the court of Philip and Alexander three are recorded: Polydias, Diades and Charias. In the following century catapult designers broke into print, and the later names and works, at least in part, of Hero, Philo, Bito, Vitruvius, Ctesibius and others survive. The tradition culminates with Archimedes, whose great contemporary fame rested chiefly on his war machines, not his mathematics. Recent research by A. G. Drachmann of the University of Copenhagen and by Derek J. de Solla Price of Yale University makes it increasingly unlikely that Plutarch's report of Archimedes' disdain for engineering is an accurate one. In Roman times catapult expertise may even have enabled its possessor on occasion to survive political purges.

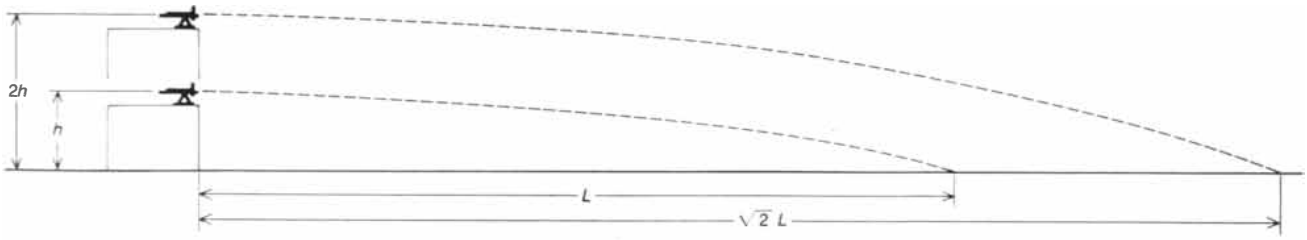
This rise in the status of the engineer



MOST COMPLEX CATAPULT invented in ancient times was a repeating weapon designed by Dionysius of Alexandria, who worked in the arsenal at Rhodes. As this detail drawing shows, arrows were fed by gravity from a magazine into the arrow trough by means of a revolving drum that was slotted to accept one shaft at a time. The revolution of the drum was controlled by a curved cam groove on its surface, which engaged a metal finger mounted on the slider. The motion of the slider was in turn produced by two flat-link chains on each side of the machine. According to the surviving text describing the repeater, the chains ran over five-sided prisms at each end of their loop. In the authors' view these prisms are assumed to have worked as inverted gears; in other words, the chain-link drive for the cocking and firing sequence relied on an engagement between lugs on the chain links and a pentagonal gear for accepting the lugs. The rear prism was turned by a winch, and the bowstring claw was locked and unlocked at the appropriate times by pegs mounted in the stock of the weapon, past which the slider moved. Hence by reciprocating the winch the device could fire arrows automatically until the magazine was empty.

rested on a strong demand for catapults. They became a part of every up-to-date fortress and siege train, and gradually they began to be deployed in the more mobile warfare of the battlefield. At sea they may have played a role in the naval arms race that led from the trireme, with its three banks of oars, to huge vessels with as many as 40 banks. Evidently the underlying assumption was that catapult fire could decimate the enemy boarding force while their ship was still too far away to grapple or ram. The larger the ship was, the more catapults it could carry and the stabler its firing platform was. This interpretation, then, sees the catapult superseding hand-to-hand warfare at sea as the cannon did 2,000 years later. Eventually the advent of new battle tactics, of armored ships called cataphracts and of Roman efforts to dominate the entire Mediterranean combined to reduce the size of warships once again. In a political parallel to land warfare, the influence of the citizen-rower was diminished in the process.

The effectiveness of the catapult led to efforts to improve its performance even beyond the introduction of the torsion bow. The engineer Ctesibius, for example, working in Alexandria in the middle of the third century B.C., attempted to supplant hair and sinew ropes, which were susceptible to breakage, rotting and changes in tension due to humidity or stretching. Both of his two alternative designs incorporated rigid arms pivoted close to their inner ends, which were bent in such a way that they pressed, when the bow was drawn, either on hammered bronze springs or on pistons sliding in airtight cylinders [*see illustrations on opposite page*]. Neither the compression of the bronze springs (which are of course inferior to steel springs for most purposes) nor the compression of the small amount of air the cylinders could contain, however, could provide a force comparable to that of the torsion bow. (In the process of researching his ideas Ctesibius discovered that "fire" would fly from the cylinder together with the piston he had forced into it with a hammer. The flame or smoke came perhaps from the ignition of the carpenter's glue he used as a sealant. If the ignition was caused by the



ONE EFFECT of the development of catapults was to force city dwellers to build higher walls in order to get longer ranges from their machines. Doubling the height of a wall, say, would not increase the range in proportion, but it could give the defenders an increased advantage. This development in the history of fortifications suggests that catapults were more of a threat than rams, since higher walls are less stable when pounded from below. Defensive catapults were probably fired more or less horizontally for the sake of accuracy, as is shown in this schematic diagram. Doubling the height (*h*) would accordingly multiply a catapult's range (*L*) by the square root of 2.

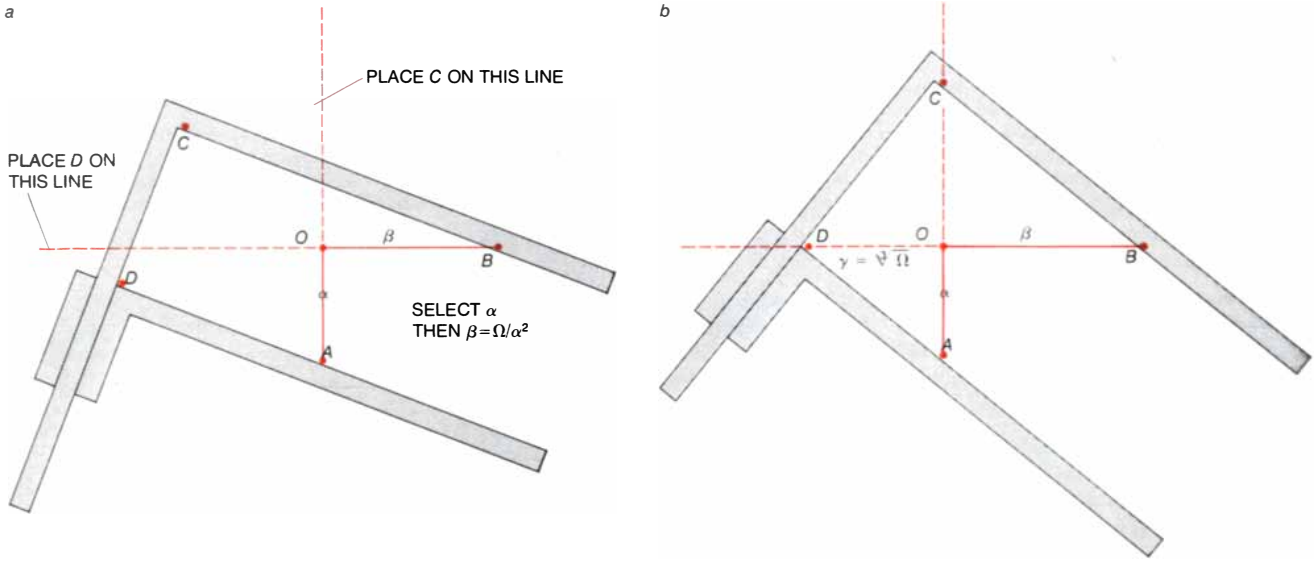
compression heating of the air, he can be viewed as the discoverer of the diesel effect.)

At about the same time Dionysius of Alexandria developed what was perhaps the most remarkable machine of its kind: a repeating catapult [*see illustration on preceding page*]. Arrows were loaded into a vertical, gravity-fed magazine and then transferred one at a time into the firing groove by a rotating tray whose motion was controlled by a cam-follower system actuated by the slider. In this system the follower reciprocated alongside the cam, which turned in response. No earlier instance of such a cam is known, and none as complex is known until the 16th century. A single windlass motion controlled the tray, the slider, the claw and the trigger, so that simply winding the windlass back and forth would automatically fire the machine until its magazine was empty. It is here that the flat-link chain, often attributed to Leonardo da Vinci, actually made its first appearance. The chain links presumably had extensions that meshed with an inverted gear; in other words, the teeth were internal, not external, much like those of a modern chain saw. (This interpretation rests in part on details in the surviving text and in part on the mechanical necessities of the situation.)

The repeating catapult failed to replace the standard one. It paid for its ease and speed of operation by having too short a range. Furthermore, its accuracy paradoxically worked against it. The device concentrated its shots so closely at its maximum range (about 200 meters) that it did not pay to open fire on even a small group of troops at that distance. (It was a model of one of these repeaters that split the arrow in Schramm's shooting exhibition for the Kaiser.) Commanders also feared that it would waste ammunition, a complaint that was raised again with the invention of repeating rifles two millenniums later.

Another reason for the failure of these interesting variations can be seen in the sophistication of the engineering efforts applied in the meantime to the common catapult. Its success made it imperative to achieve ranges at least as long as those of the enemy. This made it necessary to adjust the quantity of elastic fiber to the weight of the missile. Probably the designs were pushed not to the point of attaining the absolute maximum range but only to the point where escalating costs, declining convenience in handling or diminished accuracy due to downrange ballistic factors supervened.

One of the crucial steps in designing the torsion spring was establishing a ratio between the diameter and the length of the cylindrical bundle of elastic cords. If the cords were too short, they would develop high internal friction and might not have enough allowable elastic elongation to avoid breaking when the arms were pulled all the way back. If



CUBE-ROOT EXTRACTOR, a simple mechanical contrivance invented by an unknown Greek geometer in the third or fourth century B.C., made it possible to design a stone-throwing catapult to scale by solving a formula stating in effect that the diameter of the cord bundle in dactyls was equal to 1.1 times the cube root of 100 times the weight of the stone in minas. In order to find the cube root of some value Ω, for example, one first selected a line segment α and then obtained the value β such that $\beta = \Omega/\alpha^2$. The lines α and β were next plotted at right angles (*left*). By sliding the movable jaw of the device one was able to line it up so that point *C* would lie on the vertical extension of the line α and point *D* would lie on the horizontal extension of the line β (*right*). The solution was then given by γ, a value equal to the distance *OD*. Suppose, for example, γ equaled the cube root of 100. If one were to select α equal to 4, then β would equal 100/16, or 6.25. The cube-root extractor, after proper alignment, would yield approximately correct result for γ, namely 4.64.

they were too long, some of their elasticity would remain unused as the arms were pulled to the limits imposed by the framework. All the surviving catapult specifications imply that an optimum cylindrical configuration was indeed reached, and it was not departed from except in special circumstances, such as the exclusively short-range machines Archimedes built at Syracuse.

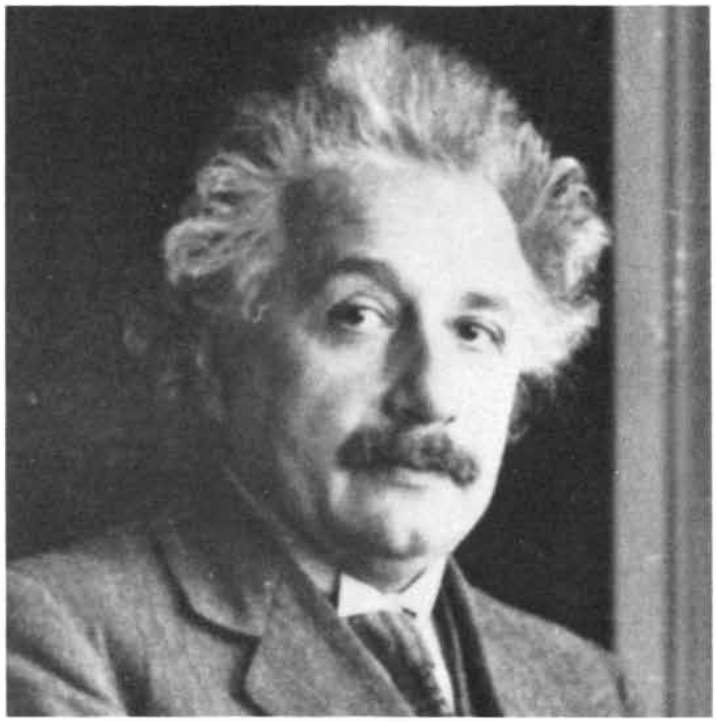
This optimization of the cord bundle was completed by roughly 270 B.C., perhaps by the group of Greek engineers working for the Ptolemaic dynasty in Egypt. There and at Rhodes the experiments of the catapult researchers were, according to Philo, "heavily subsidized because they had ambitious kings who fostered craftsmanship." This phase of the investigations culminated in quantified results of a distinctly modern kind.

The results were summarized in two formulas. For an arrow shooter the diameter of the cord bundle was set simply at 1/9 the arrow length. The more complex stone-thrower formula stated, in modern terms, that the diameter of the cord bundle in dactyls (about 19.8 millimeters) is equal to 1.1 times the cube root of 100 times the weight of the ball in minas (about 437 grams).

The stone-thrower formula has two remarkable features. First, it gives a true and accurate solution for optimal design. To see why, first assume (as is indeed reasonable) that the catapult engineers wanted to maximize the performance of their machines. Accordingly they had to maximize the kinetic energy of their projectiles. To do this they had to maximize the potential energy stored in the torsion springs. Modern elasticity theory applied to the design of these springs tells us that the stored energy available for firing will be proportional to the amount of the initial tension given the bundle in stringing it through the frame, the additional tension caused by the pre-twisting of the bundle, the square of the angle indicating the amount of additional twisting by the pulling back of the bow arm, and the cube of the bundle's diameter. The cubing of the bundle's diameter means that to express the diameter of the spring bundle in terms of the mass of the projectile one would have to extract a cube root.

Note that to arrive at this result one must employ the concepts of kinetic and potential energy, which were not brought into a meaningful relation until the 18th century and the work of Leonhard Euler and Daniel Bernoulli. Also needed is elasticity theory, which had been begun by Hooke and Robert Boyle about half a century earlier. Finally, one must employ the principles of ballistics, which were not clarified until the work of Francesco Cavalieri and Galileo Galilei in about 1630. That the ancient catapult engineers were able to arrive at a

He committed himself to science, peace



and the atom bomb.

He was the high priest of modern science. A genius who was both mythologized and misunderstood. NOVA explores the conflicts, ironies and historical forces that shaped the life and thought of Albert Einstein... and changed the destiny of the world. Watch NOVA's *Einstein.*

From Einstein to the Amazon

In the weeks ahead, NOVA looks at a new breed of athlete that's using science to improve performance... India's radical approach to economic development... a revolutionary treatment for pain, depression and schizophrenia... a reclusive billionaire's 3½ million acre Amazon jungle.

NOVA takes you to worlds you've never seen.

Series of Surprises

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

Take NOVA to a new world every week.

NOVA

Thursday March 15 at 8:00*

*Check Local Listings

PBS

A COMPANY CALLED TRW

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

formula that stands up in the light of these much later developments is truly impressive.

It is the utilization of a cube root that constitutes the second remarkable feature of the stone-thrower formula, because it was written at a time when Greek mathematics was not yet capable of dealing fully with third-degree equations. In about 460 B.C. Hippocrates of Chios (not the famed physician) had stated that a cube might be accurately doubled in volume if two lines defining two mean proportionals between two given lines could be found. In the following century only a beginning had been made toward solving the problem. Archytas of Tarentum and Eudoxus of Cnidus had devised elegant theoretical solutions, but they were three-dimensional, very awkward physically and of almost no use in performing calculations. There the matter stood until the advent of the torsion bow.

Most of the next group of solvers of the cube-root problem had either a direct or an indirect connection with catapults. Menaechmus, according to tradition, was a tutor of Alexander the Great's, and therefore he was present at the time and place when the torsion bow first came into prominence. His solution involved intersecting conic sections, a concept he seems to have discovered. Unfortunately there is no evidence surviving on whether he was led to consider conics by the problem of extracting cube roots for the design of catapults.

The next solver of the cube-root problem was Eratosthenes, a friend of Archimedes' and a native of Alexandria, which was then a center of catapult research. Eratosthenes stated explicitly that the catapult was the chief practical reason for working on cube-root problems. We can assume he was interested in engineering problems, since Archimedes dedicated his book *On Method* to him. In this work approximate solutions to mathematical problems are roughed out initially by a practical mechanical-engineering approach. For example, sections of bodies are weighed to determine the ratios of their volumes. Eratosthenes' solution relied on a mechanical contrivance with sliding parts, somewhat similar to the one in the bottom illustration on page 156.

All of the next group of cube-root investigators, including Philo of Byzantium, Archimedes of Syracuse and Hero of Alexandria, were famous for their work on catapults. It is interesting to note that the largest stone thrower on record, a three-talent (78-kilogram) machine, was built by Archimedes. A machine of this caliber would have required cube-root extraction, because there are no natural roots in the quantities needed. Archimedes was also forced to depart from normal catapult proportions in building his special short-range machines. Their effectiveness testifies to his skill as a mathematical engineer.

Taken as a group these early students of the cube-root problem stand apart from the mainstream of Greek mathematics. Instead of limiting themselves to the straightedge and the compass, they devised simple mechanical contrivances that enabled them to generate conic sections and even curves of higher orders. Some aids, such as the simple "slide rule" shown in the bottom illustration on page 156, anticipate the proportional compasses and gunner's sectors of the late 16th century. The work of this group, however, was neglected until the Renaissance, when mathematical growth resumed at about the point where they had left off. Descartes's *La géométrie,* for instance, begins with procedures and devices much like theirs.

It would appear, therefore, that the catapult engineers conducted experiments that forced them into a domain that traditional mathematical procedures had not yet penetrated. It is fairly easy even today to fit third-degree data to a second-degree curve if the data are bad or the investigator is unscrupulous. Hence one must feel a good deal of respect for these ancient investigators. They must have repeated their catapult-firing tests many times, kept very accurate records and interpreted their results with a high degree of precision. The introductory passages of Philo's *Belopoeica* lay great stress on the experimental procedures and achievements of the early catapult engineers, and from the vantage of modern engineering theory the accuracy of his account seems to be fully borne out.

Having arrived at an optimal volume and configuration for the torsion-spring bundle, the catapult engineers continued their experiments until they had optimized the dimensions for the remaining major pieces of the machine. If the bow arms were too short, the cocking force required would be excessive, the travel of the bowstring would be limited and its energy-transfer capabilities would be curtailed. If the bow arms were too long, they would retard the action of the springs by their increased mass or make the weapon too bulky. Once the length of the bow arms was determined, the length of the slider and the stock could be determined from the travel of the bowstring, and so on for the rest of the machine.

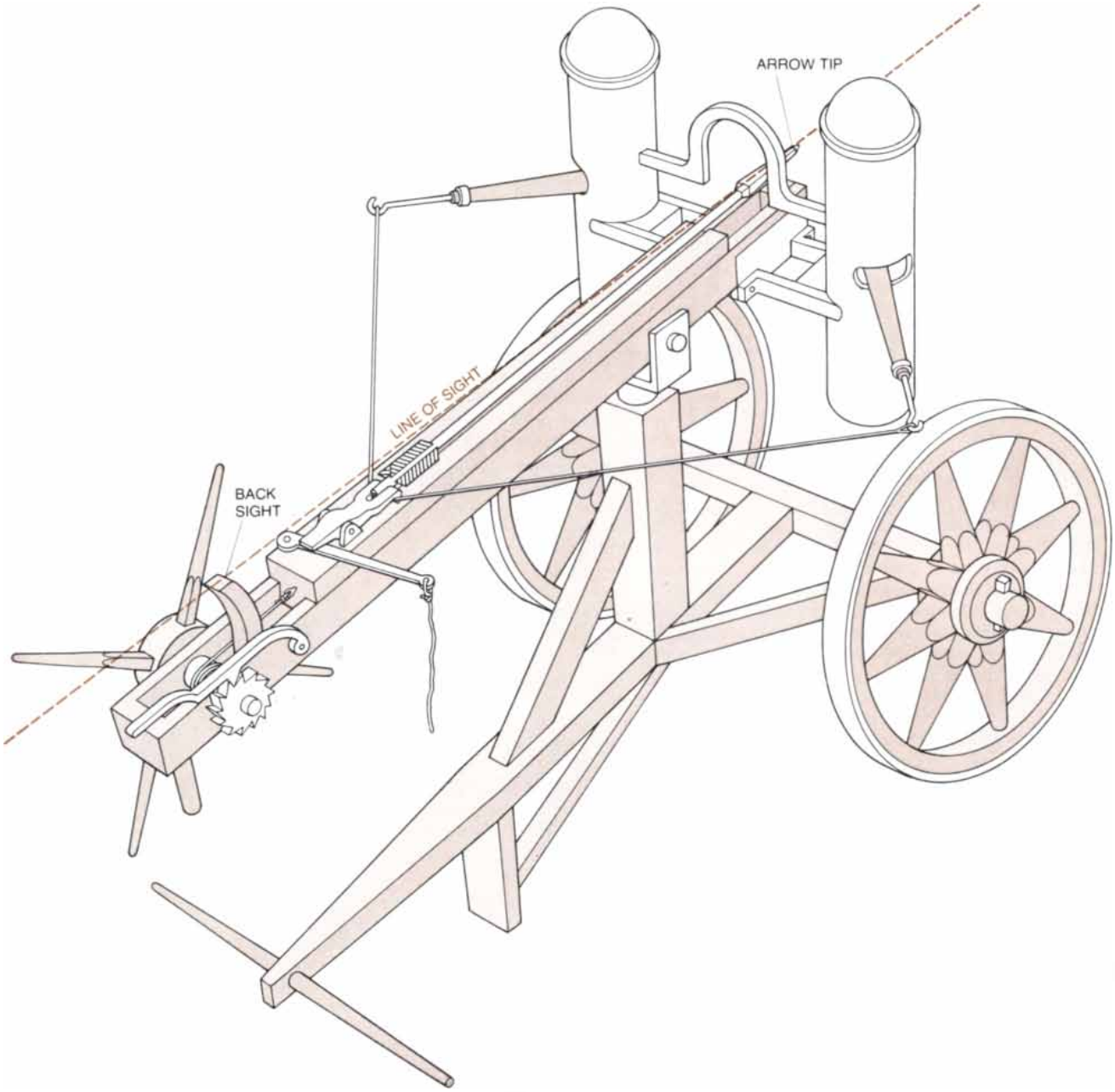
Eventually the catapult engineers wrote their texts in such a way that the dimensions of the major parts were given as multiples of the diameter of the spring. Once this diameter had been calculated for the size of the projectile desired the rest of the machine was automatically brought to the proper scale. The surviving texts that contain this information testify to a level of engineering rationality that was not achieved again until the time of the Industrial Revolution.

The last major improvement in cata-

pult design came in later Roman times, when the basic material of the frame was changed from wood to iron. This innovation made possible a reduction in size, an increase in stress levels and a greater freedom of travel for the bow arms. The catapult could now be mounted on wheels and moved easily from one part of a battlefield to another in response to the shifting fortunes of the fight. The new open frame also simplified aiming, which with the wood construction of the earlier machines had been limited, particularly for close, moving targets.

This advanced catapult design came too late for the expansive period of Roman civilization, but it played a role in stabilizing the boundaries of the Empire and in helping to prevent their erosion. As the decline of the Empire proceeded, however, the technical skills necessary to build and maintain such sophisticated machines appear to have become scarcer. A new, simpler machine, called the onager, with only one spring and one arm, which terminated in a spoon and was used for throwing stones, now came increasingly into prominence. It would provide the heavy artillery of the Middle Ages until the appearance of the trebuchet, whose even simpler construction was gravity-powered.

The scientific design of complex machines, with deliberate experimental adjustment of the dimensions of the components, did not appear again in Western civilization until the 18th century. During the ancient period the changes in which the catapult played a key part prefigured in striking ways issues that would appear again in the relations between science and technology on the one hand and warfare and society on the other.



LATE DEVELOPMENT in the history of catapult technology was the Roman cheiroballistra, a comparatively small wheeled machine dating from about the first century A.D. The weapon's iron frame gave it enough mobility for battlefield use, and its open structure made it easier to locate moving targets. The weapon was probably aimed by lining up the tip of the arrow with an appropriately elevated back sight mounted on the rear of the stock. The mathematical relations between the force of the bow, the displacement of the slider and the angle of elevation were such that by using a line of sight passing through the target, the arrow tip and the back sight, the shooter could automatically achieve the correct trajectory simply by estimating the distance and winching the slider backward for the appropriate number of ratchet clicks. The cheiroballistra was succeeded in turn by a simpler, one-armed stone thrower, called the onager.

How to tell original prints from cheap reproductions.

Buying original prints is an excellent way for the novice to make a modest investment in beautiful works of art.

The buyer had better beware, however.

As more and more attention is being given to art as an investment, more and more dealers are trying to pass off cheap reproductions as fine art.

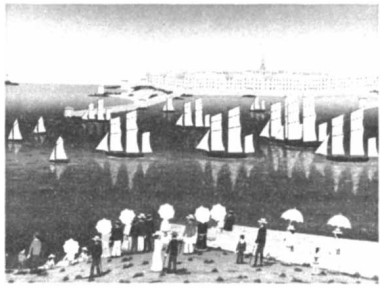
At the Original Print Collectors Group, we sell only fine original works to a growing number of people who are enthusiastic about buying good art at moderate prices.

One of our primary concerns is to educate our customers, so that they'll be able to discriminate and get true value for their money. A few pointers, then about original prints:

Original prints, whatever the process used to create them, are "hand-pulled" under the artist's supervision, or by the artist himself. This means that any work produced photographically or by another purely mechanical process is not an *original* print.

Bernard Charoy's *Sophie*, original lithograph. Signed limited edition of 150.

If you examine a print under a magnifying glass and discover a regular pattern of small dots, you will know immediately that it is not an original work.



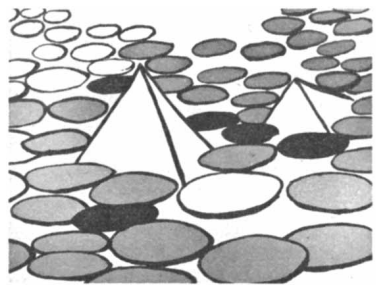
Fanch Ledan's *Un Dimanche à St. Malo*, original lithograph. Signed limited edition of 250.

If the print is a serigraph (produced by a silkscreen process), the ink will seem rather like paint and appear to be sitting upon (rather than absorbed into) the paper.

Etchings and engravings will produce a physical impression on the paper itself.

Contemporary original prints will always be signed in pencil by the artist himself, to indicate personal supervision and approval of the individual print.

Original prints are always produced in limited editions, at the most several hundred copies. If thousands of copies are available, then the work is almost certainly *not* original. (In most cases, to assure that the integrity of their work is not violated, artists will destroy the plate after the edition has been produced.)



Alexander Calder's *Circle of Pyramids*, original lithograph. Signed limited edition of 100.

The print will have a figure on its border, indicating the number of prints pulled and the number of the individual print in the sequence.

For example, the figure 50/100 means that the edition has been limited to 100, and that this is the 50th print pulled.

Is the edition limited and is the print numbered? Did the artist personally sign the individual work?

There's a lot more to be learned about buying original prints, more than we can tell you on this page.

But if you'll send for our free color brochure, we'll be happy to tell you more, and show you the quality and variety of the prints we offer.

They include limited edition etchings, serigraphs and lithographs by both contemporary and traditional artists with international reputations. Some examples: Calder, Chagall, Dali, Delacroix, Picasso and Renoir.

If you decide to buy a print, we'll frame it handsomely and send it to you with a full money back guarantee. If you're not satisfied for any reason, you simply return your print and get a full refund.

Rolf Rafflewski's *Pont Neuf*, original lithograph. Signed limited edition of 225.

We'll also send you a brochure, every two months, featuring new selections of prints, along with an informative newsletter.

If you're interested in investing in art, invest a 15¢ stamp in your art education and send for our brochure today.

Original print collectors group, Ltd.

120 East 56th Street, Dept. SA-17, New York, N.Y. 10022

☐ PLEASE SEND ME YOUR FREE COLOR BROCHURE, LATEST NEWSLETTER AND OTHER INFORMATION.

Name

Address

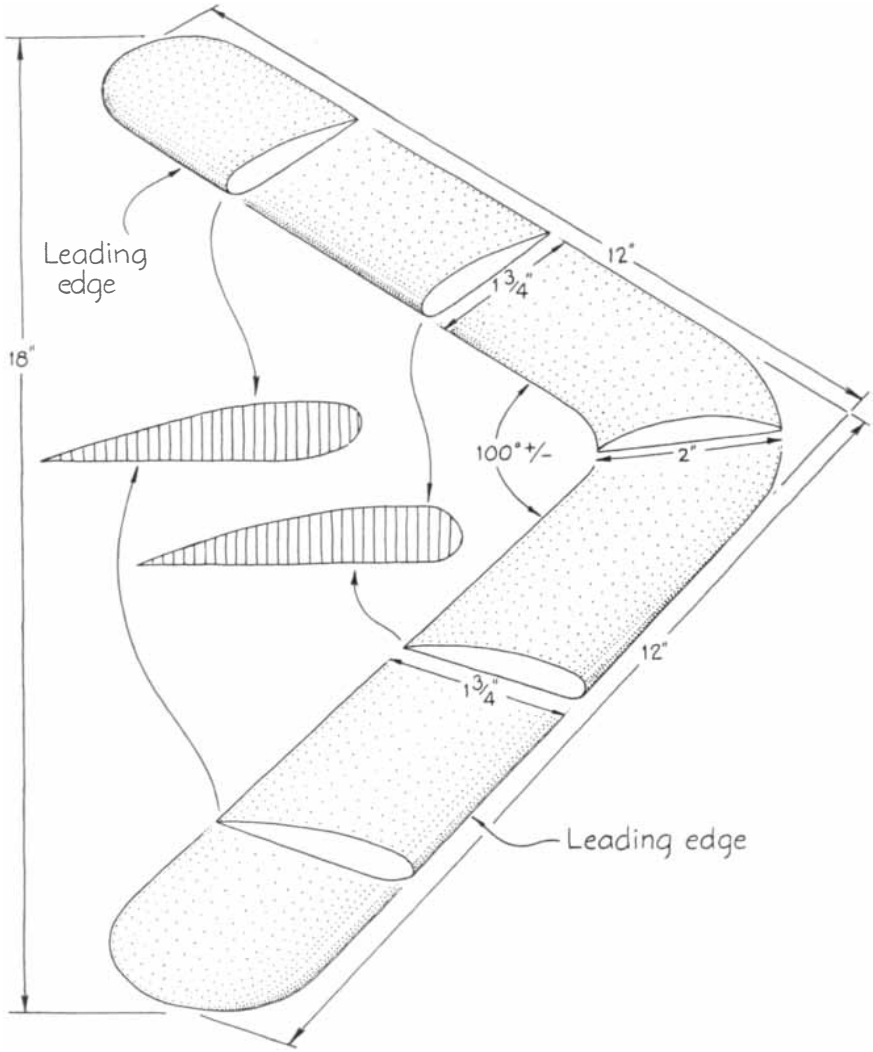
City State Zip

THE AMATEUR SCIENTIST

Boomerangs! How to make them and also how they fly

by Jearl Walker

A boomerang is surely one of the oddest devices ever to serve as a weapon or a plaything. It was apparently invented by accident, notably in Australia by the native people but also independently in many other places. If you throw an ordinary stick, it falls to the ground not far away, but a boomerang can travel as much as 200 meters (round trip) and can be aimed so skillfully by an expert thrower that game or an enemy can be hit. The boomerang probably originated as a weapon designed for accurate straight flight, but most people find the returning boomerang, which was mainly just a plaything for the Australians, more interesting. Ironically the straight-flying boomerang is probably more complicated aerodynamically than the returning boomerang. As ancient as both devices are, an amateur investigator can still do a great deal to advance the understanding of their flight.



Herb Smith's "Gem" design for a boomerang

Although good returning boomerangs are occasionally available in sporting-goods stores, most commercial returning boomerangs are mass-produced and fly poorly. Indeed, many of them do not even return. The plastic boomerang made by Wham-O is one of the best types available in toy stores. Many other types, all of them excellent, are available from Ruhe-Rangs, Box 7324, Benjamin Franklin Station, Washington, D.C. 20044, thanks to Benjamin Ruhe, a boomerang enthusiast formerly with the Smithsonian Institution. During his service there Ruhe helped to organize the Annual Smithsonian Open Boomerang Tournament. The tournament, held in late spring or early summer, is great fun for the 100 or so contestants who enter. This year it is scheduled for June 9 on the mall in Washington.

If you would like to experiment with boomerangs, you need to be able to construct your own. Only then can you make the variations necessary to determine what factors influence a boomerang's flight. The best material from which to cut the basic stock of a boomerang is Baltic birch in a marine or aircraft plywood between 1/4 and 3/8 inch thick, with five or more laminations. This type of plywood is resistant to wear and water; it is also dense, which makes for a boomerang that is heavy for its size.

Cut a cardboard pattern for a boomerang of whatever shape you want. An example is the returning boomerang designed by Herb Smith and shown in the illustration at the left. (If you are left-handed, you will have to make a left-handed boomerang, which is the mirror image of the one drawn.) Place the pattern on the plywood and mark off the outline of the boomerang in pencil.

With a coping saw or a handsaw cut out the boomerang blank. The edges and the top must now be shaped in the general form shown in the illustration. (Little is done to the bottom other than putting a slope on what will be the leading edge.) The leading edge must be blunt, whereas the trailing edge must be sharper, with the top surface sloping down to meet the unaltered, flat bottom surface. The arms must have the cross-sectional shape of a classic airfoil, because they must provide lift much as the classic airfoil does.

Clamp the blank in a vise and cut and shape the edges and the top with a rasp that has a curved surface. Smooth out the grooves left by the rasp and finish shaping the wood by rubbing the sur-

All those looking for a better First Class service can now rest assured.

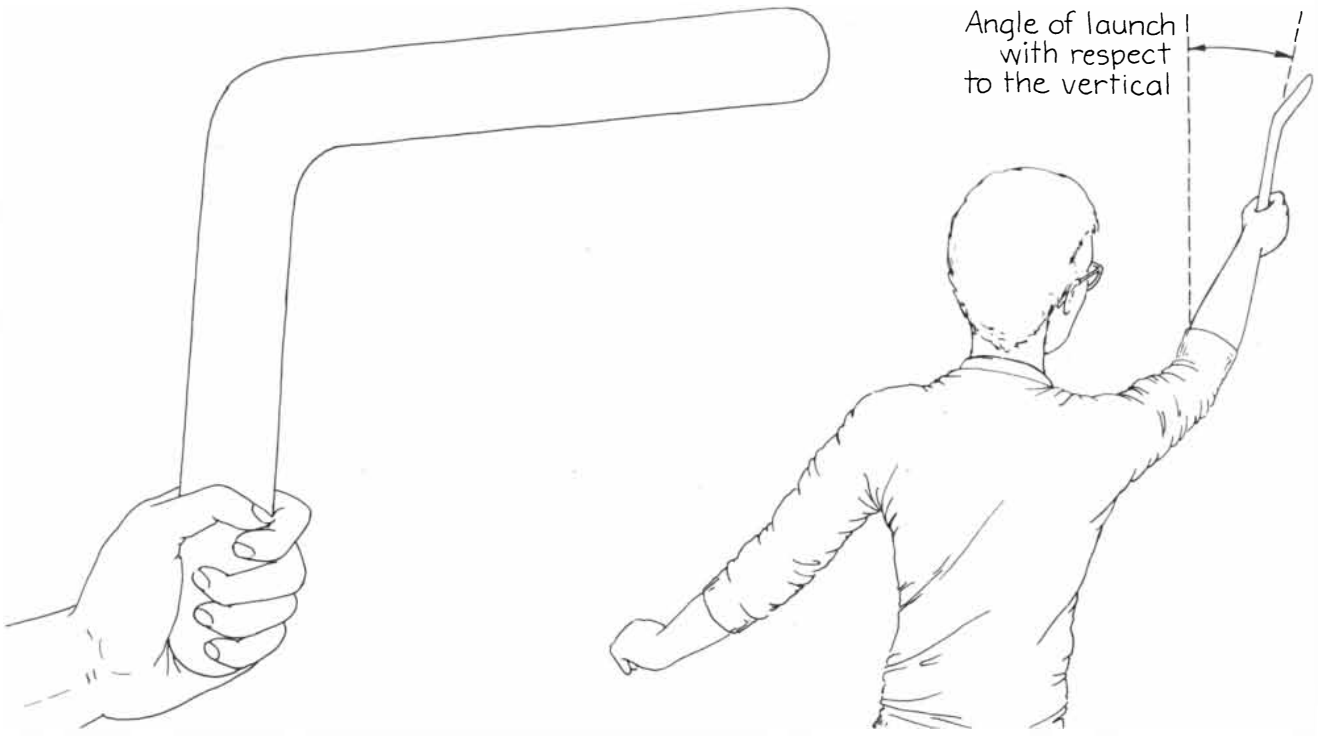
JAL's first class is first class service the way you have dreamed it should be.

Consider our new and exclusive Sky Sleeper Service. Sky Sleeper is a big, comfortable bed in the sky. The service is available for an additional charge* on our specially fitted 747 Garden Jets that fly across the Pacific.

There's more. There are French, German and American wines and continental cuisine and Japanese delicacies that look as good as they taste. And a hostess in kimono whose courtesy and hospitality make you feel as if you're in Japan already. Welcome to JAL's first class—first in its class.

*$90 Honolulu—Tokyo
$120 San Francisco or Los Angeles—Tokyo
C$130 Vancouver—Tokyo

We never forget how important you are.
JAPAN AIR LINES

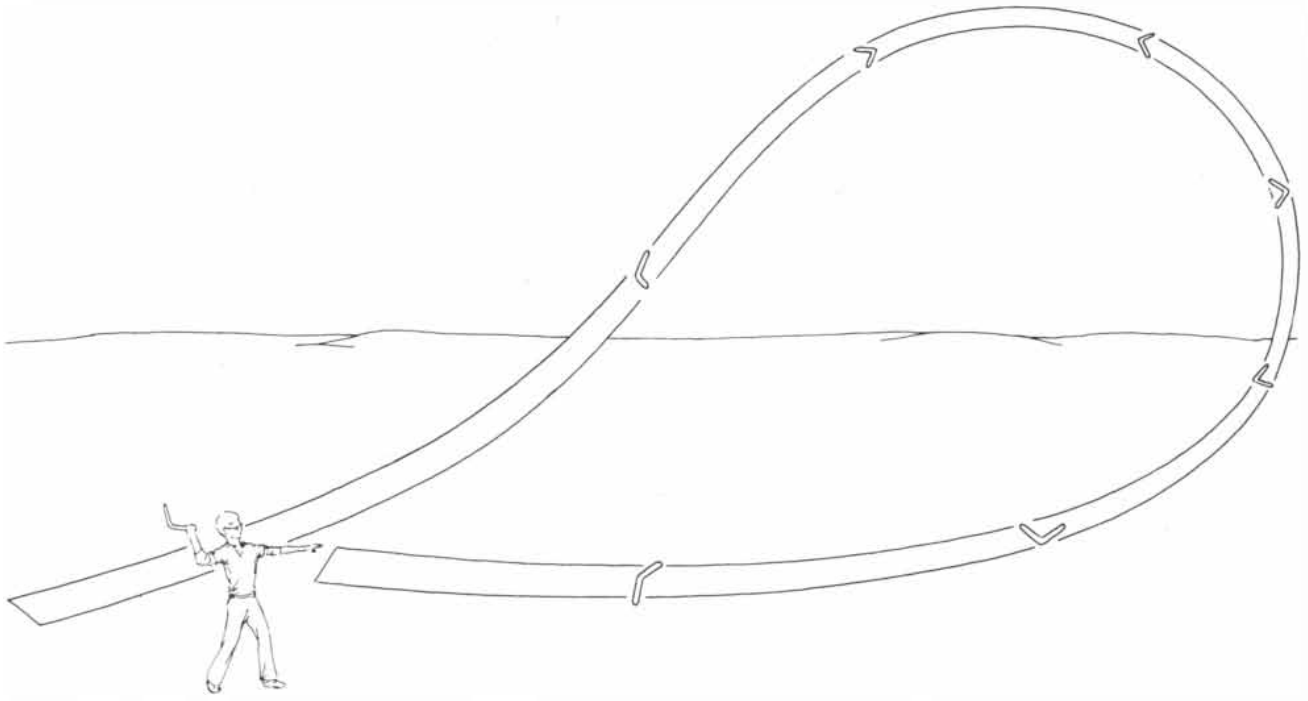


Preparing to throw a boomerang

faces with a piece of coarse sandpaper wrapped around a piece of soft wood. Before you finish off the surface with a finer sandpaper you should make a test flight with the boomerang so that you can tune it by more shaping with the rasp or with coarse sandpaper. Tuning the boomerang means that you throw it, cut or sand it more and then throw it again until it flies the way you want it to.

A boomerang cannot be thrown well in a strong wind. If there is a light wind, face toward it, turn 45 degrees to the right and throw the boomerang in that direction. Hold the boomerang vertically by the tip of one of the arms (which one usually does not matter) with the flat side away from you. Reach behind your head with the boomerang and then throw it toward the horizon, snapping it forward when your arm is fully stretched forward. Do not try to throw hard at first. It is the snap that counts, not the strength of the throw. The boomerang stays up in the air because of the spin the snap imparts to it.

The proper orientation of the boomerang (the plane in which it will be spinning) will vary according to the wind conditions and the type of boomer-



A typical flight path

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading because it's easy to learn . . . lasts a lifetime . . . applies to everything you read . . . and is the only fully accredited course with the option of college credits.

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and re-read words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions — then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading" — but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven-step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark made by the 75,000 slow readers who completed the *Speed Learning Program* was: "Why didn't someone teach me this a long time ago?" They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

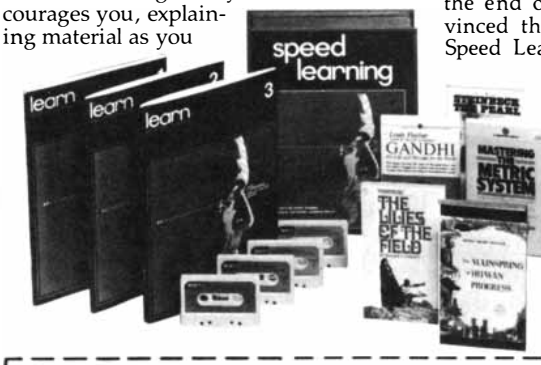
The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think — a radical departure from anything you have ever seen or heard about. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning* gives you what speed reading *can't*.

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you — no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week*, *Wall Street Journal*, *Family Circle*, *N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.



Executives, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 10 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 10 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

Note: Many companies and government agencies have tuition assistance plans for employees providing full or partial payment for college credit programs.

In most cases, the entire cost of your Speed Learning Program is Tax Deductible.

EARN PROFESSIONAL & COLLEGE CREDITS

Speed Learning is approved for credit by the following professional and educational institutions:

- **Foundation for Accounting Education** 20 CPE Credit Hours
- **American Management Association** 1.9 Continuing Education Units
- **National Society of Public Accountants** 20 Continuing Education Hours
- **College Credit** 3 credits from the National College of Business

Details and registration forms included with each program.

OFFERED INTERNATIONALLY BY

- **Institute of Electrical & Electronics Engineers**
- **American Chemical Society**
- **National Association of Life Underwriters**

learn INCORPORATED 113 Gaither Drive, Mount Laurel, N.J. 08054

3ISAJ

Please send me the Speed Learning Program at $89.95 plus $3.00 for handling and insured delivery.
Please check the method of payment below:

☐ Check or money order enclosed for items ordered. New Jersey Residents add 5% sales tax.
☐ Please charge my credit card under the regular payment terms: ☐ Bank Americard
☐ Master Charge Interbank No. _____ ☐ American Express ☐ Diners Club

Card No. _____ Exp. Date _____

I understand that if after 10 days I am not delighted in every way, I may return the materials and obtain a full refund with no questions asked.

Name _____

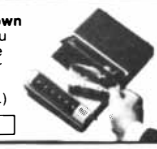
Address _____

City _____ State _____ Zip _____

x Signature _____

If you don't already own a cassette player, you may order this Deluxe Cassette Recorder for only $49.95. (Includes handling and delivery.)

Check here to order ☐



Outside U.S.A. $99.95 + $3 surface mail — airmail extra

On Photographing a Monster

It *hates* being photographed. It glares balefully at the camera. It sulks, fidgets, pouts.

It is that most fearsome of creatures, a perfectly normal four-year-old child.

The photographer, wise in the ways of distracting difficult subjects, abandons his camera. He wanders about the studio, talking and gesturing. The monster, no longer feeling threatened by the evil eye of the lens, relaxes and becomes once more the sweet child its mother knows.

From time to time, the eye of the camera blinks. There is a quiet whirring of an electric motor, and the camera blinks again. Yet the photographer is nowhere near it.

An Ingenious Little Dial.

The camera is a Hasselblad, the motor-drive 500EL/M.

A dial on the camera is set at "SR." This prereleases the entire reflex system. At the end of a 20-foot extension cord, the photographer holds a release mechanism. When triggered, it operates the leaf shutter without a moment's delay, thus reducing the risk of "eye blink" from the subject and delivering a higher number of useable photographs.

There are four other settings on this dial: Single-frame in the normal mode. Continuous, automatic firing for as long as the release is held down. A single frame in the prereleased, speeded-up mode for one frame only. And continuous, automatic firing in the speeded-up mode at the rate of one frame every 8/10th of a second.

Hasselblad Versus 35mm.

Why is this photographer not using one of the many excellent motor-drive 35's for this assignment?

One compelling reason might be the size of the Hasselblad negative. Each is 2¼ inches square, *almost four times the area of a 35mm frame.* (See box, below right, for actual size.)

To make an 11 x 14 print of the child, one would have to enlarge a 35mm negative 11.6 times. A comparable Hasselblad print requires only a 7-times enlargement. The ability to hold contrast is phenomenal.

Then there is the camera's leaf shutter. Unlike the shutter of a 35mm camera, it is fully synchronized for every type of flash at every speed up to 1/500th of a second.

Hasselblad Versus Others.

But why a Hasselblad when there are less-expensive alternatives? Why did eight out of ten top professional photographers surveyed name Hasselblad as the medium-format camera used in their work?

Consider this: The Hasselblad System makes available no less than 4 cameras, 21 lenses, 8 viewfinders, 9 film magazines (including Polaroid® backs) that give you a choice of anywhere from 1 to 500 exposures, and a staggering array of accessories.

A total of 332 pieces.

No other system that exists today in the medium-format field can muster more than a fraction of that number of pieces.

For that matter, no other system offers three different formats—6 x 6cm, 6 x 4.5cm, and a 4.5 x 4.5cm superslide format for showing in any 35mm projector.

*A lavish brochure on the 500EL/M and other Hasselblad cameras is available free if you write: Braun North America, **Dept. SA3H, 55 Cambridge Parkway, Cambridge, Mass. 02142**— a division of The Gillette Company and exclusive marketer of Hasselblad in the U.S.*

It is actually possible to custom-build for yourself over 100 different Hasselblad cameras for almost every imaginable photographic situation—from the most simple of family snapshots to the most sophisticated of professional work.

Is it any wonder that the space program came to Hasselblad for the equipment which was to record man's first voyage to the moon?

HASSELBLAD®

The Motor-Drive Hasselblad 500EL/M, made in Sweden. The standard lens is a multi-coated f2.8 80mm Zeiss Planar.

ang. To achieve a good flight you might have to throw the boomerang with its plane nearly vertical. Under other circumstances you will have to shift the plane (tilting the top of the boomerang away from you) by as much as 45 degrees. The greater the tilt of the plane in which the boomerang spins is, the more upward lift the boomerang will initially have. If you give it too much lift, it climbs too fast and then plummets to the ground with such force that it may break.

In a good flight a returning boomerang travels horizontally around an imaginary sphere. On returning it probably will hover or even loop a bit before it drops to the ground at your feet. If you are lucky, it might make one or two additional circles (smaller than the first circle) before it falls. Although you launch the boomerang with its spin plane almost vertical, it probably will return with the plane nearly horizontal. I shall explain below why the plane must turn over in this way if the boomerang is to complete the trip.

If your boomerang consistently lands to your right in a light wind, try throwing it a bit more to the left of the wind. Similarly, if it is landing to your left, try throwing it more to the right of the wind. If it lands behind you, try throwing it with less force. If that does not work, throw it somewhat above the horizon with the spin plane tilted less from the vertical. If the day is windless and you are not getting a full return, tilt the spin plane farther from the vertical in order to gain more lift during the flight.

Be careful not to injure people or damage things with your boomerang. It can be quite a weapon. Throw it only in a large, open space. If people are present, make sure they know what you are doing so that they can be prepared to dodge.

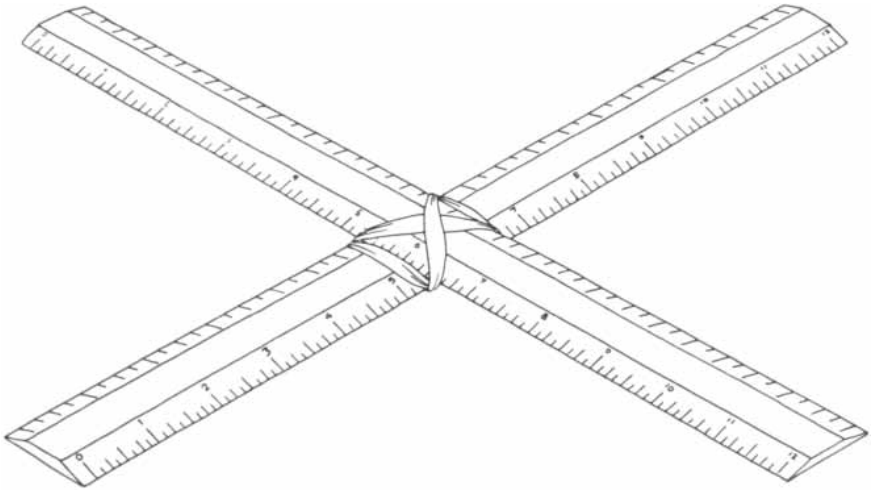
The successful tuning of a boomerang involves both experience and luck. In general if you make the top surface more curved, the boomerang will have more lift, which means that it will return in a tighter circle. Flattening the top surface or curving the bottom surface will give the boomerang less lift because the cross-sectional shape of the arms is then less like a classic airfoil. If the spin decreases too fast, so that the boomerang falls to the ground in mid-flight, the reason may be that excessive air drag on the arms is robbing them of their spin. Some surface roughening might be beneficial to the flight, but any large grooves left by the rasp will surely create additional air drag that will shorten the flight time.

Instead of shaping the arms carefully you might prefer to twist them so that during a flight the leading edge on each arm is tilted to deflect the passing air to the right, giving the boomerang a lift to the left. This type of lift is easy to visualize. When you put your hand out of a window of a moving car and turn it through various angles, you can feel the lift. To twist a boomerang heat it gradually in an oven at 400 degrees Fahrenheit and then (with gloves on, of course) carefully twist the arms until the wood is cool. If you twist too much, heat the boomerang again and twist the arms back a bit.

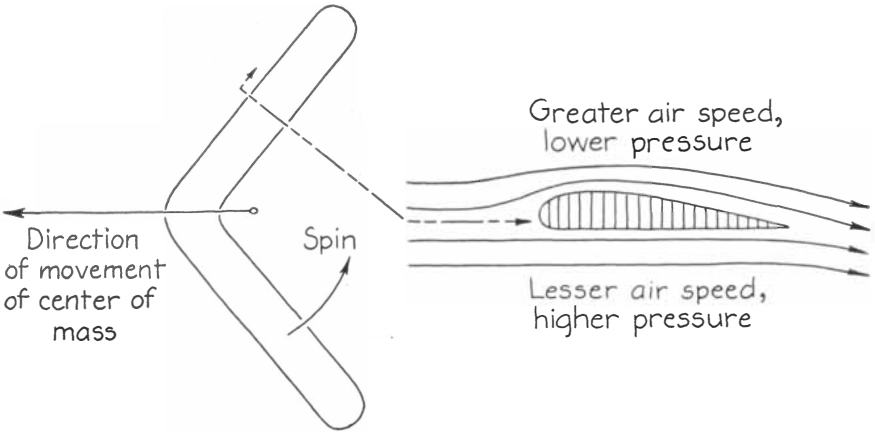
Once you have a properly flying boomerang you might want to finish it off with a cellulose covering and some decorative designs. Smith's excellent booklet, which is listed in the bibliography for this issue [*page 174*], explains how to do this kind of finishing and also gives a number of boomerang designs.

If your boomerang breaks, do not throw away the pieces. Glue them together with epoxy, clamp them until they dry and then file and sand the surface back into the desired shape. Although the boomerang will not be as strong as it was before, its flight path might be altered in an interesting way by the small change in its distribution of mass resulting from the break.

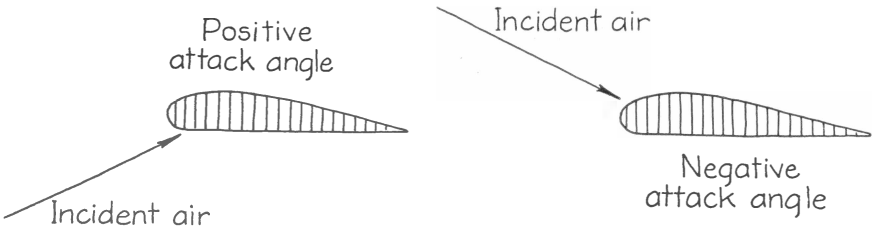
A boomerang does not have to be limited to two arms. Indeed, one of the easiest boomerangs to build is a four-blade design consisting of two rulers crossed and fastened at the center. The right



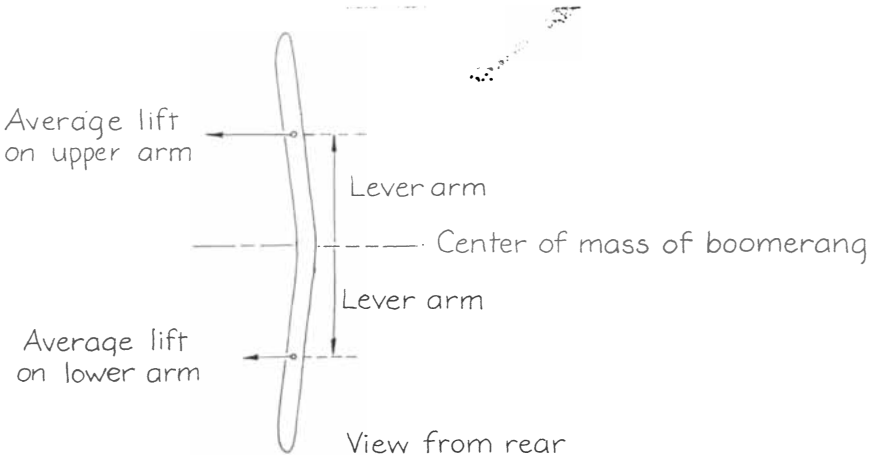
A four-blade design made with plastic rulers and rubber bands



Pattern of airflow over the arm of a flying boomerang



Positive and negative attack angles



Average lift on the upper and lower arms

kind of ruler has a curved top surface and a fairly flat bottom surface. You can attach two such crossed rulers either by wrapping a strong rubber band around them or by putting a bolt and nut through the center hole they usually have. Throw this type of boomerang the same way that you would a two-armed one. Take care to avoid being cut by sharp edges, and never use rulers that have metal edges.

A simple cross boomerang can be fashioned from a cardboard square about five inches on a side. Cut out a boomerang with three or four blades and twist them slightly so that the boomerang is not quite all in the same plane. Adding weight to the arms of a boomerang increases its range. With the cardboard boomerang it is easy to add weight by attaching paper clips to the end of the arms. This boomerang can be demonstrated in a classroom. If it has too much range for a classroom, you can decrease the range by increasing the twist on the arms so that the boomerang is less in the same plane or by bending the arms along a center line through their length. With the latter technique the arms have an exaggerated airfoil shape: at least one side is sharply convex but the other is not. As usual, throw the boomerang with the convex side toward you. By changing the arms from being almost flat to being more like an airfoil, you increase the lift on the boomerang; its path will be a tighter circle.

When you begin to throw your wood boomerang well, you might be tempted to catch it. The result may be a sharp blow to your fingers. If you are determined to catch a boomerang, hold your hands flat and slap them together to trap the boomerang as it hovers above the ground, still spinning, in the last stage of its flight. Keep your fingers away from the turning blades.

The explanation of the return of a boomerang lies primarily in the cross-sectional shape of the arms and the fact that the boomerang spins. Without these two features a boomerang would behave like any other thrown stick. The cross-sectional shape gives the boomerang aerodynamic lift similar to the lift generated by some airplane wings. The spinning gives the boomerang stability. Through a bit of fortunate rotational mechanics the spinning also causes the axis about which the boomerang spins to rotate in much the same way that the spin axis of a top rotates about the vertical. The lift and the stability keep the boomerang up, and the rotation of the spin axis brings it back to the thrower.

Aerodynamic lift can be explained with a simple model of a classic airplane wing similar to the one I described in this department for February, 1978, to explain the lift of a kite. The classic airfoil has a flat bottom, a blunt front, a sharp rear and a convex top. Air passes around a wing faster along the top of the wing than along the bottom. The reason can be seen by visualizing the passing air as being of two kinds. One kind flows around the wing with no rotation in the stream and with the same speed on the top and the bottom of the wing. The other is a circulation cell: it flows to the rear over the top of the wing and to the front over the bottom. Such a circulation is created by a real wing because the air's viscosity and its adhesion to the surface of the wing force it into this pattern as it flows to the rear off the curved top surface.

In the superposition of the two idealized airstreams the two velocities add above the wing and subtract below it, with the result that the real air speed is greater above the wing than it is below it. The difference is important to the lift because the air pressure in the stream is inversely related to the speed of the stream. Hence the air pressure is less above the wing than below it, and the wing gets a push upward. (A real airplane wing can have a more complicated airflow pattern than this simple model implies. Moreover, when an airplane is traveling at high speed, some of the lift may come from the impact of the passing air on the underside of a wing that is inclined slightly upward in order to deflect the air downward.)

If the classic airfoil is inclined to the airstream in such a way that the airstream is more incident on the curved top side, the lift is of course less. Such an arrangement is termed a negative angle of attack. In a simple model the reduction of lift is due to the downward push the incident stream exerts on the top surface. One might also argue that lift is partially lost because the tendency for the air to circle about the wing is lessened and the speed of the air on the top side of the airfoil differs less from the speed on the bottom side.

Conversely, if the airfoil is inclined so that the airstream is incident somewhat more on the flat bottom side than on the top, a situation that would be called a positive angle of attack, the lift increases because of the upward push from the airstream on the bottom side. The air drag also increases. If the angle is too large, the disadvantages of increased air drag outweigh the advantages of lift. The attack angle of the arms of the boomerang as they turn through the air is important to its flight.

Boomerang arms can have a variety of cross-sectional shapes, but most of them are similar in cross section to the classic airfoil. Usually this shape includes a blunt edge that turns into the air as the boomerang spins and a sharper edge that trails during the turn. One side is usually flat and the other convex. Variations on this basic form are numerous, however, and little systematic work seems to have been done on determining which shapes are best aerodynamically. Some boomerangs are actually flat on both sides but with their arms twisted so that the air is deflected as the arms turn through the wind.

The lift on a boomerang differs in a major way from the lift on the classic airplane wing. In the first stage of a flight the boomerang's "lift" is mostly horizontal, with only enough upward force to balance the weight of the device. Since the boomerang is spinning mostly about a horizontal axis, the curved sides of the arms spin in a plane that is almost vertical and the lift is almost horizontal. For the sake of simplicity in what follows I shall ignore the weight of the boomerang. I shall also assume a boomerang that is thrown outward by a right-handed thrower so that the plane of spinning is initially exactly vertical. The lift will be to the thrower's left, so that the boomerang begins to move to the left as it continues to spin in the vertical plane.

If this were the entire story, the boomerang would never come back. To see why it turns around and returns you must understand what else the lift does to the boomerang. In particular it is nec-

SCIENCE/SCOPE

The planet Jupiter will be scrutinized as never before by a spacecraft to be launched from the Space Shuttle in January 1982. NASA's Project Galileo will consist of an orbiter and a Hughes-built probe that will journey together for 42 months to the massive planet. Scientists hope the mission will provide new insight into the origin and evolution of the solar system, as well as reveal why Jupiter mysteriously emits more energy than it receives from the sun.

The probe, after separating from the orbiter, will plunge deeply into the hostile Jovian atmosphere. Its six instruments will take temperature and pressure readings, analyze cloud and atmospheric composition, detect for lightning, and glean other data. The probe will be designed to withstand the force of Jupiter's intense atmospheric pressure for about one hour. The orbiter will relay probe data to earth and will continue its mission and circle Jupiter for 20 months, taking pictures and analyzing the atmospheres of the red-eyed planet and its four largest moons.

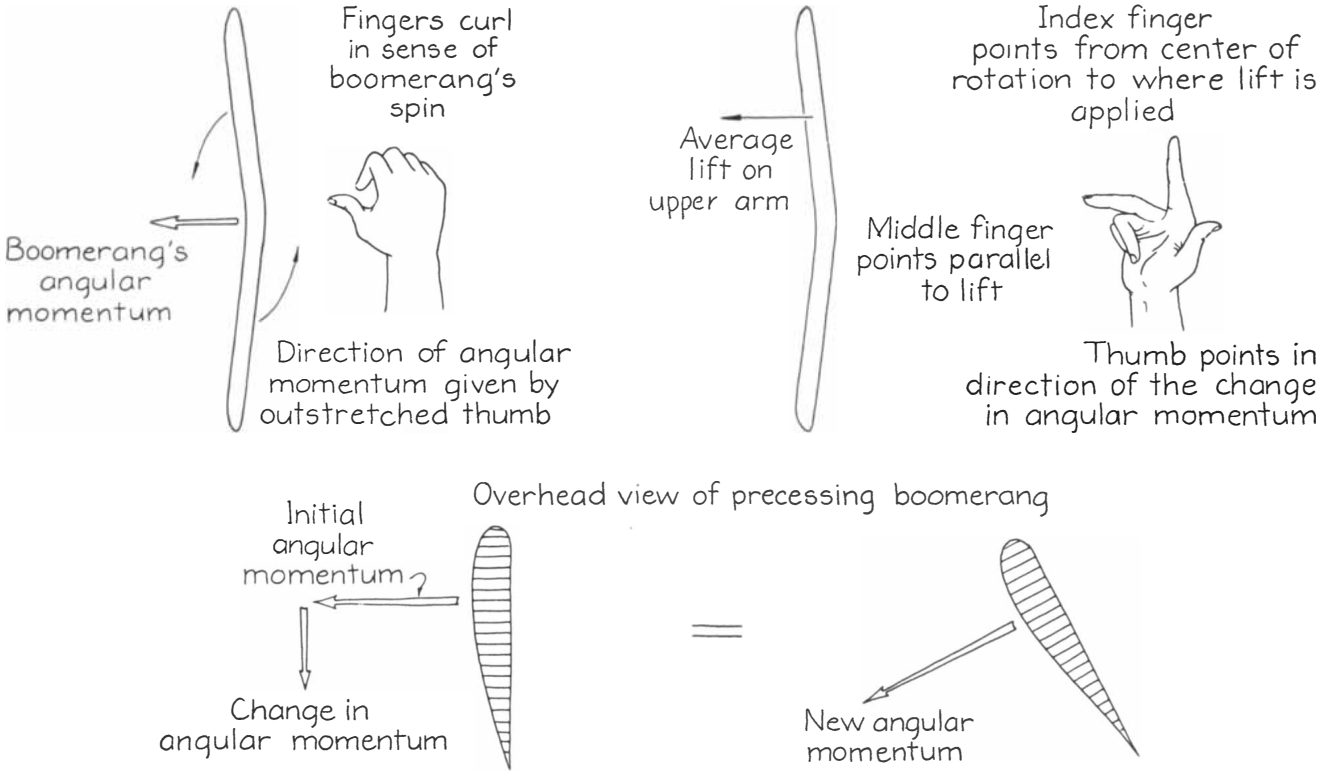
Using signals transmitted in sequential bursts lasting 1/30,000th of a second, a Hughes multiplexing system designed for jet aircraft is able to control more than a dozen channels for passenger entertainment and service over a single coaxial cable. Because the signals are too rapid to be discerned separately by the human ear, the channels for high-fidelity stereo music, movie sound tracks, and passenger address announcements all seem to operate simultaneously. Other channels regulated by the system include reading lights and flight attendant call buttons.

By eliminating the need for separate wires to each channel, the system is able to save several hundred pounds of weight. It also offers better reliability, lower operating costs, and improved passenger services. Hughes systems have been installed on every McDonnell Douglas DC-10 in the world today, and now six international carriers have specified the equipment for their Boeing 747s.

Hughes Radar Systems Group has career opportunities for engineers, scientists, and programmers. We design and manufacture many of today's most complex airborne and spaceborne radar electronics systems, including data links, electronic warfare systems, and display systems. These projects incorporate microelectronics, microprocessors, solid state microwave devices, and other advanced technologies. We need Systems Analysts, Microwave Specialists (antenna, receivers, transmitters), Circuit Designers (analog, digital & RF/IF), Scientific Programmers, Mechanical Designers, Product Design Engineers, and others. Rush your resume to Engineering Employment, Hughes Radar Systems Group, 2060 Imperial Hwy., El Segundo, CA 90245.

A contract to develop and build 100 satellite earth terminals has been awarded to Hughes by Satellite Business Systems. SBS, a partnership formed by Aetna Life & Casualty, Comsat General, and IBM, will receive the communications industry's first four-year warranty. The terminals will operate with the system's three satellites, also being built by Hughes to service U.S. businesses, government agencies, and public service organizations.

The unmanned terminals will be the first to use Ku-band frequencies (12 to 14 GHz) on a large scale in a U.S. domestic satellite. This makes possible small terminals located on customer premises, thus eliminating the need for a microwave link. With these highly-reliable redundant terminals, SBS data communications customers will benefit from a very high availability rate. The first engineering model is scheduled for delivery in nine months. All 100 terminals should be delivered by mid-1982.



How to determine angular momentum and a change in it

essary to know how the torque due to the lift on the boomerang causes a precession of the spin plane.

Imagine that one of the boomerang's arms has spun to its highest possible position and the other arm is almost in its lowest possible position. (I am discussing the basic banana-shaped boomerang.) The upper arm is turning in the same direction in which the center of the boomerang is moving, whereas the lower arm is moving opposite to the motion of the center. The air passing the upper arm is moving faster (in relation to the arm) than the air passing the lower arm. Therefore more lift is generated on the upper arm than on the lower one. The part of the boomerang higher in the spin will always experience a greater lift and hence a greater push to the side than the part lower in the spin.

My first thought was that the difference in horizontal lift (more lift on the upper arm than on the lower one) would cause the spin plane of the boomerang to tilt, thereby angling the lift downward (a disastrous effect). What actually happens, however, is that the difference in lift causes a rotation of the plane about a vertical axis. It is this rotation of the spin plane, commonly called precession, that brings the boomerang back.

To understand what causes the rotation you must examine the torque created by the lift. Take the center of the boomerang as the axis about which it is spinning. (Actually the center of mass around which a two-armed boomerang spins is likely to be well off center, but that does not alter the outcome of the argument.) Take the average lift on the upper arm as being directed horizontally outward from the center of the arm. Similarly, take the average lift on the lower arm as being also directed horizontally outward from the center of the arm. The torque created by one of these lifts, as measured from the center of the boomerang, is the product of the lift and the distance to where the lift is applied, that is, half the length of an arm. Since the upper arm has the greater lift, it also has the greater torque.

If the boomerang were not spinning, this difference in torques would merely make the plane of the boomerang tilt over. Since the upper arm has the greater torque, the plane would tilt counterclockwise as seen by the person who has just thrown the boomerang. The fact that the boomerang is spinning, however, makes a big difference, because it then has angular momentum and the tendency to tilt the spin plane results in a rotation of the spin plane about the vertical axis.

Angular momentum is the product of the boomerang's rate of spin and a function involving the mass and the mass distribution of the device. For an example in another setting imagine yourself attempting to turn a merry-go-round holding several children. The force you apply tangent to the rim multiplied by the radius of the merry-go-round is the torque you are supplying. When you begin, the torque causes an angular acceleration of the merry-go-round; the spin increases from zero to some final value. How would you arrange the children in order to achieve a given angular acceleration with the least force? Intuitively you would place them near the center. Their mass is the same, of course, but their mass distribution with respect to the center of rotation is different. When the mass is nearer the center, the merry-go-round is easier to turn. The mass and its distribution are taken into account by the function known as the moment of inertia. The greater the mass or the farther from the center it is placed, the greater the moment of inertia and the greater the force you will have to supply in order to achieve a given angular acceleration.

Once the merry-go-round is spinning and you are no longer pushing on the rim, the apparatus has a certain angular momentum because it has spin and a moment of inertia. Angular momentum is usually represented by a vector pointing perpendicularly to the plane in which the object is spinning. Here the vector would be vertical. The direction (up or down) is chosen by convention as being the direction of the thumb on the right hand when the hand is held in a hitchhiker's pose with the fingers curled in the direction of the spin of the object.

The only way you could change the size or the direction of such a vector would be to apply another torque to the object. With a merry-go-round you

Three Lively Introductions to the Excitement of Science and Mathematics

New for College Courses in Basic Physics

Introductory Physics

Mashuri L. Warren, San Francisco State University and Lawrence Berkeley Laboratory
Consulting Editor: Edwin F. Taylor, Massachusetts Institute of Technology

Teachers and students will welcome this volume's concrete approach. *Introductory Physics* comprehensively covers the essential topics in basic physics, with an organization easily adapted to a wide range of courses for beginning physics students, students in the biomedical sciences and other students in the natural sciences.

Basic concepts in *Introductory Physics* are introduced with careful definitions, worked examples, and applications. Extended examples challenge the better student, and each chapter features a review of important principles and sources for further reading. Recognizing students' lack of full familiarity with mathematics, Mashuri L. Warren introduces mathematical concepts gradually, reintroducing techniques as needed.

Dr. Warren's approach emphasizes experimental physics: the basis of measurement, units of measurement, the evidence underlying important ideas. *Introductory Physics* explains the important interplay between theory and experiment, drawing on examples from engineering, medicine, industry, biology, transportation, and sports.

February 1979, 700 pages (approx.), 628 illustrations, cloth: $16.00 (tent.)

New for Precalculus Courses

Algebra and Trigonometry with Analytic Geometry

Arthur B. Simon, California State University, Hayward

"This book has three purposes: to prepare students for calculus, to illustrate the strong relationship between algebra and geometry, and to present some interesting applications of mathematics."
—from the Preface

Arthur B. Simon achieves the goals he sets forth by making sure students get intense, active practice. Included in this book are more than 2,000 exercises, graded in difficulty, many with answers provided.

Students will appreciate the book's straightforward tone and helpful organization. Each chapter begins with an outline alerting students to important points, and most sections conclude with a convenient review. Each new concept is introduced intuitively and illustrated with several examples before an accurate formal definition is presented.

Teachers will appreciate the flexibility of this text, which is designed to fit the structure of a variety of courses. The Preface suggests outlines for both semester and quarter courses of different emphasis and duration.

February 1979, 550 pages (approx.), 299 illustrations, cloth: $14.50 (tent.)

New in High School and Liberal Arts Mathematics

Elementary Algebra

Harold R. Jacobs

Elementary Algebra is a complete, day-by-day algebra course. Each lesson is suitable for one or two days' work and includes four sets of exercises that can be tailored to fit students' needs. Set I reviews earlier ideas and techniques; Sets II and III offer parallel exercises, with the answers to Set II in the back of the book; and Set IV provides more challenging exercises.

Elementary Algebra offers a complete instructional package, including a detailed Teacher's Guide, Test Masters, and Transparency Masters. Like Harold Jacobs' other acclaimed texts, *Elementary Algebra* not only explains ideas with extraordinary clarity but also captures the interest and imagination of students, who will find themselves drawn into mathematics even if they have resisted it in the past.

February 1979, 872 pages, 690 illustrations, cloth: $12.00
Teacher's Guide, Transparency Masters, and Test Masters.

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104

could push on the rim again. (A convention for choosing how to draw a vector representing the change in angular momentum involves pointing the index finger of the right hand from the center of the rotation toward the place where the force is applied and pointing the middle finger in the direction of the applied force. If you make your thumb on that hand perpendicular to both fingers, it automatically points in the direction of the change in the angular momentum. The new angular-momentum vector is the combination of the old one and the vector representing the change.) With a merry-go-round that you had resumed pushing tangent to the rim the new vector would still be vertical but would be larger or smaller depending on whether your aim was to make the merry-go-round turn faster or slower.

A boomerang that is spinning has two torques acting on the arms, one created by the average lift on the upper arm and one created by the average lift on the lower arm. Since the lift on the upper arm is greater, it determines what happens to the angular momentum, and so I shall ignore the lift on the lower arm. (The argument would not change even if I included the smaller lift.) Imagine that the boomerang is receding from you just after you have thrown it with your right hand. It is spinning in a vertical plane and has an angular-momentum vector pointing to your left. The average lift on the upper arm creates a torque that will change the direction of the vector as the boomerang continues to fly away.

To determine how the vector changes use your right hand, orienting the fingers and the thumb properly. With your index finger pointing from the center of the boomerang to the center of the upper arm and your middle finger pointing to your left in order to be in the direction of the lift on that arm, your outstretched thumb points toward you. Therefore the vector representing the change in angular momentum points toward you. Mentally combining the change vector and the original vector is best done from an overhead point of view. The change vector is perpendicular to the original one and gives a new vector rotated from the old one toward you. The size of the angular momentum is unaltered because the change vector is perpendicular to the old one. Only the direction of the angular momentum is changed, and it is rotated about a vertical axis to point more toward you.

This type of rotation of an angular-momentum vector is precession; it is seen when the axis of a top precesses about the vertical. Another common example of precession is seen in the turning of a motorcycle. The wheels of a motorcycle spin fast enough and have moments of inertia sufficiently large to make their angular momentum large. To turn the motorcycle you cannot just turn the handlebars, as you would when riding a bicycle. Instead you make the motorcycle lean in the direction of the turn. The torques then experienced by the motorcycle cause the angular-momentum vectors of the wheels to precess, turning the motorcycle as a whole.

During the precession of the spin plane of a boomerang the boomerang continues to travel along a path with a certain speed but is continuously deflected by the horizontal lift it experiences. The resulting path approximates a large circle. In a successful boomerang flight the spin plane will precess at the same rate at which the device circles in its path. Its angle of attack remains somewhat positive. This match is necessary in order to keep the arms at the proper attack angle.

Suppose the boomerang precesses too slowly. Then as it travels along its circular path its spin plane rotates about a vertical axis at a rate lower than the rate at which the boomerang as a whole travels along its path. When the spin plane lags behind, the attack angle becomes increasingly negative and the boomerang loses lift.

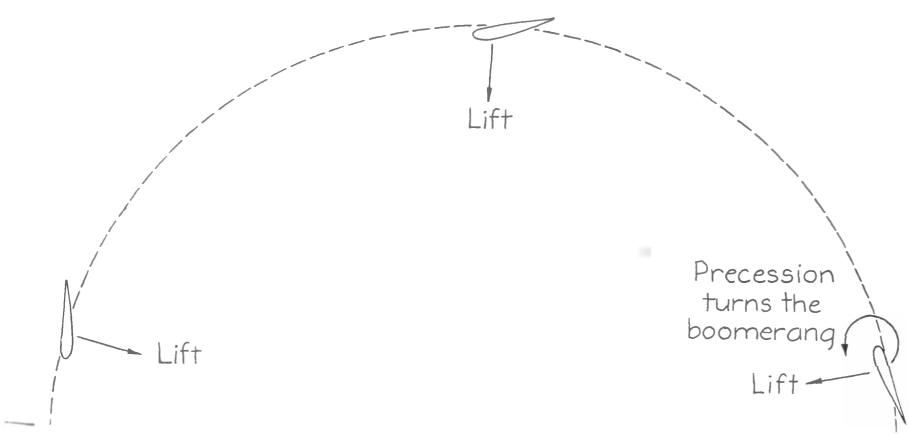
If the spin plane precesses too quickly, it turns about a vertical axis faster than the boomerang as a whole travels along the large circular path. As a result the attack angle becomes increasingly positive until the spin plane is perpendicular to the oncoming airstream. The air drag would surely ruin the flight by then.

The match between the precessional rate and the rate at which the boomerang travels along the large circular path is not critical and is in fact somewhat automatic, since both rates depend on the lift. Throw your boomerang, sand down and reshape the arms and throw it again until you come near the match and the boomerang returns. I know of no sure way to remedy a persistently unsuccessful boomerang.

The circular path of the boomerang is independent of the speed with which you throw it. Only the moment of inertia and the cross-sectional shape of the boomerang determine the radius of the path. With a given boomerang you will therefore achieve the same large circle (for the same throw of the boomerang in the vertical plane I have been assuming) regardless of how hard you throw the device (provided, of course, you throw it hard enough so that it has sufficient speed to complete its journey). If you want to change the size of the circle, you must ordinarily choose a different boomerang with a different moment of inertia or cross-sectional shape. Next month, however, I shall explain how you can also add ballast to the arms in order to increase their moment of inertia. This technique is used by boomerang throwers intent on breaking distance records.

A Frisbee flies in a quite similar way. It has a curved top surface and is launched with a flick of the wrist to give it spin. The Frisbee gains lift by virtue of the impact of the air or by the difference in air speed across the top and bottom surfaces. A Frisbee properly thrown in an almost vertical plane will return to the thrower the way a boomerang does. Usually, however, a Frisbee is launched to curve slightly to another person, so that the thrower orients the spin plane to provide just enough horizontal lift to achieve the curve.

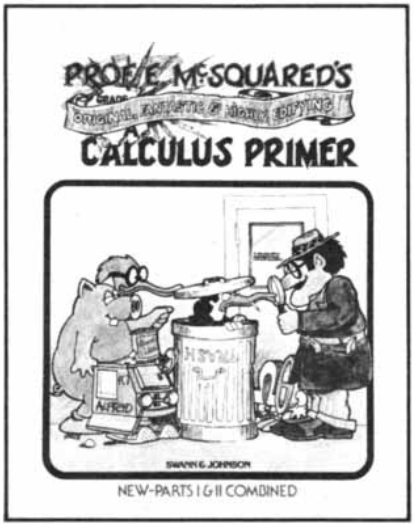
Both a boomerang and a Frisbee can be skipped across the ground without destroying the flight. Imagine a Frisbee skimming just above the ground with its leading edge tipped slightly downward. That edge then strikes the ground. The force from the ground at the contact point puts a torque on the Frisbee and changes the angular momentum, but because the change vector is almost perpendicular to the original angular-momentum vector the new angular-momentum vector is just a rotation of the original one. The angular momentum does not change appreciably in size, only in direction. Therefore the spin of the Frisbee is not much slowed; the device is merely reoriented and then goes sailing off in a new direction.



How precession keeps the attack angle positive

BOOKS OF FUNDAMENTAL IMPORTANCE

Published by William Kaufmann, Inc.



Professor E. McSquared's original, fantastic and highly edifying Calculus Primer

HOWARD SWANN and JOHN JOHNSON. 214 p., illus., 1977.

Through brilliant use of cartoons and ingenious analogy, the Professor and his friends introduce calculus with unbelievable energy, humor and clarity. Ideal for self-study. "Incredible . . . throughout the authors have lavishly spread their humor, poking fun at the 'gobbledygook' and mystique of mathematics whilst putting across the ideas of analysis . . . Highly recommended."—***London Times Educational Supplement.*** "Perhaps the most original primer on calculus yet devised."—***New York Public Library.***

$7.95 paperback **A**

Stalking the Wild Taboo

Second Edition

GARRETT HARDIN 290 p., 1978.

In this collection of his classic essays (available now for the first time in paperback), scientist-philosopher Hardin challenges some cultural taboos and provides fresh insights into contemporary issues such as abortion, nuclear power, earthquake prediction, aid to undeveloped countries, the values of wilderness and many other topics. The first edition was termed a "splendid collection" by **Martin Gardner** of ***Scientific American***, and ***Library Journal*** wrote: "Laced with common sense and the author's fine humor . . . this book can hardly fail to stimulate." This second edition includes much new material from Hardin's public lectures and from his articles in *Psychology Today*, *Skeptic*, *Science* and other well-known periodicals.

$4.95 paperback B **$11.95 hardcover C**

The California Water Atlas

WILLIAM L. KAHRL, Project Director and General Editor. 16" x 18", 132 p., illus. (including 41 color plates), June 1979.

The first comprehensive atlas of California water resources, their history and uses, prepared by the Governor's Office of Planning and Research in cooperation with the California Department of Water Resources. The Water Atlas project has been guided by an advisory group of the most prominent figures in the fields of hydrology, engineering, history, book design, environmental protection and water law.

The Water Atlas features a detailed narrative text written by topic specialists, illustrated with photos and diagrams and combined with 41 striking six-color maps and charts prepared by expert cartographers from California State University, Northridge. The mapmakers have employed the most modern cartographic techniques yet developed for conveying complex information with accuracy and immediate visual impact. The Water Atlas is the *only* single-volume point of access to understanding the complex California water story—an unprecedented model for water resource assessment and planning for the entire country.

$37.50 hardcover (individually cartoned) **D**

Prepublication price (requires prepayment of cash and shipping charges with order) **through June 15, 1979: $32.50.** Books will be shipped upon publication.

***Water Atlas* purchasers please note:** Because of the extraordinary size of the *Water Atlas*, special shipping charges apply. Please add $2.50 for the first copy ordered and $1.00 for every copy thereafter.

Draw!

A Visual Approach to Learning, Thinking and Communicating

KURT HANKS and LARRY BELLISTON. 10" x 9". 242 p., illus., 1977.

Over 900 drawings, cartoons, sketches and photographs, and brief, easy-to-grasp discussions of the essential tools, methods and techniques of modern drawing and design. "A veritable symposium on the visual media . . . Hanks and Belliston are really showing us how to turn on an unused channel of thinking; how can you refuse such an offer?"—***Human Behavior Magazine.*** "A very clever book designed to destroy myths about artistic ability . . . the point is that anyone can and should draw."—***Booklist*, 11/15/77.**

$19.75 hardcover E **$9.75 paperback F**

Into Print

A Practical Guide to Writing, Illustrating and Publishing

MARY HILL, U.S. Geological Survey, and WENDALL COCHRAN, Editor of ***Geotimes.*** 175 p., illus. 1977.

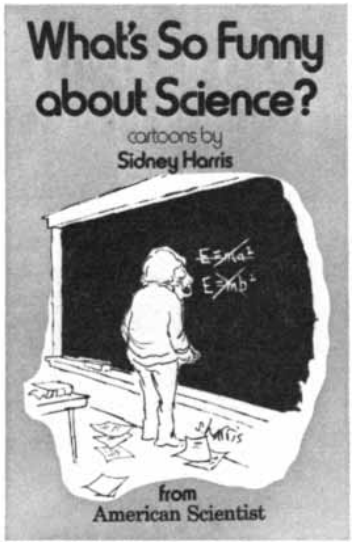
Two experienced non-fiction editors have written a practical, easy-to-use handbook for those who work with publishers, printers and artists on books, papers, speeches and other presentations. Richly illustrated, with extensive examples and references, including advice on self-publishing. "They touch all the bases."—***Publishers Weekly.***

$ 6.95 paperback G **$12.00 hardcover H**

Einstein Centennial Cards

Three S. Harris cartoons—two of each in a packet of six postcards—commemorating Albert Einstein and $E = mc^2$. The book cover below shows one of the cartoons.

$1.00 set of six cards **I**



What's So Funny about Science?

Cartoons from *American Scientist*

SIDNEY HARRIS, 120 p., 1977.

A collection of over 100 cartoons on a remarkable range of scientific and technical subjects, bringing out what is human and humorous in sophisticated scientific inquiry. S. Harris's inimitable cartoons have appeared in *Time*, *The New Yorker*, *Playboy* and many other publications. "For years we were aware that Albert Einstein and Werner Von Braun weren't tickling our funnybones. Now we know what's so funny about science. It's Sidney Harris."—***Bob and Ray.*** "Laugh at S. Harris's cartoons and you demonstrate two things: 1) you've got a sense of humor, and 2) you're intelligent."—***Isaac Asimov.***

$3.95 paperback J **$7.95 hardcover K**

William Kaufmann, Inc.
One First Street • Los Altos, CA 94022

Please send me the books I've circled below. Enclosed is full payment plus $1.00 for shipping. (Please note higher shipping rates for copies of *The California Water Atlas*.) I understand that if I am not completely satisfied, I may return the order within 30 days for a full refund. **A B C D E F G H I J K**

Name _____

Address _____

City _____

State _____ Zip _____

For all purchases, California residents please add sales tax.

sciam

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

TIME. C. D. Broad in *Encyclopaedia of Religion and Ethics,* edited by James Hastings. Charles Scribner's Sons, 1922.

TRAVELERS IN TIME: STRANGE TALES OF MAN'S JOURNEYINGS INTO THE PAST AND THE FUTURE. Edited by Philip Van Doren Stern. Doubleday & Co., Inc., 1947.

SYMBOLISM AND BELIEF. Edwyn Bevan. Kennikat Press, Inc., 1968.

THE PROFESSIONALIZATION OF THE U.S. LABOR FORCE

CORPORATE LIB: WOMEN'S CHALLENGE TO MANAGEMENT. Edited by Eli Ginzberg and Alice M. Yohalem. The Johns Hopkins University Press, 1973.

THE MANPOWER CONNECTION: EDUCATION AND WORK. Eli Ginzberg. Harvard University Press, 1975.

THE OVEREDUCATED AMERICAN. Richard B. Freeman. Academic Press, 1976.

APOLLO OBJECTS

BASIC PHYSICS OF THE SOLAR SYSTEM. V. M. Blanco and S. W. McCuskey. Addison-Wesley Publishing Co., 1961.

THE DISCOVERY OF ICARUS. Robert S. Richardson in *Scientific American,* Vol. 212, No. 4, pages 106–115; April, 1965.

METEORITES AND THE ORIGIN OF PLANETS. J. A. Wood. McGraw-Hill Book Company, 1968.

METEORITES: CLASSIFICATION AND PROPERTIES. John T. Watson. Springer-Verlag, 1974.

SPACE GEOLOGY: AN INTRODUCTION. Elbert A. King. John Wiley & Sons, Inc., 1976.

THE ASTEROIDS. Clark R. Chapman, James G. Williams and William K. Hartmann in *Annual Review of Astronomy and Astrophysics,* Vol. 16, pages 33–75; 1978.

CANCER METASTASIS

MECHANISMS OF CANCER INVASION AND METASTASIS. Isaiah J. Fidler in *Cancer: A Comprehensive Treatise,* Vol. 4, edited by Frederick F. Becker. Plenum Press, 1975.

A PATHOBIOLOGIC OVERVIEW OF METASTASIS. L. Weiss in *Seminars in Oncology,* Vol. 4, No. 1, pages 5–19; March, 1977.

EXPERIMENTAL TUMOR METASTASIS: CHARACTERISTICS AND ORGAN SPECIFICITY. Garth L. Nicolson in *Bioscience,* Vol. 28, No. 7, pages 441–447; July, 1978.

SPECIFICITY OF ARREST, SURVIVAL, AND GROWTH OF SELECTED METASTATIC VARIANT CELL LINES. Garth L. Nicolson and Kenneth W. Brunson in *Cancer Research,* Vol. 38, No. 11, Part 2, pages 4105–4111; November, 1978.

THE SPECTRUM OF ATOMIC HYDROGEN

THE SPECTRUM OF ATOMIC HYDROGEN. G. W. Series. Oxford University Press, 1957.

ATOMIC SPECTRA. W. R. Hindmarsh. Pergamon Press, 1967.

OPTICAL RESOLUTION OF THE LAMB SHIFT IN ATOMIC HYDROGEN BY LASER SATURATION SPECTROSCOPY. T. W. Hänsch, I. S. Shahin and A. L. Schawlow in *Nature Physical Science,* Vol. 235, No. 56, pages 63–65; January 24, 1972.

AN EXPERIMENTAL DETERMINATION OF THE RYDBERG CONSTANT. B. P. Kibble, W. R. C. Rowley, R. E. Shawyer and G. W. Series in *Journal of Physics B,* Vol. 6, pages 1079–1089; June, 1973.

HYDROGEN 1*S*-2*S* ISOTOPE SHIFT AND 1*S* LAMB SHIFT MEASURED BY LASER SPECTROSCOPY. S. A. Lee, R. Wallenstein and T. W. Hänsch in *Physical Review Letters,* Vol. 35, No. 19, pages 1262–1266; November 10, 1975.

THE PERCEPTION OF SURFACE BLACKS AND WHITES

EYE MOVEMENTS AND VISION. A. L. Yarbus. Translated by B. Haigh. Plenum Press, 1967.

DISCOUNTING THE BACKGROUND—THE MISSING LINK IN THE EXPLANATION OF CHROMATIC INDUCTION. J. Walraven in *Vision Research,* Vol. 16, pages 289–295; 1976.

PERCEIVED LIGHTNESS DEPENDS ON PERCEIVED SPATIAL ARRANGEMENT. Alan L. Gilchrist in *Science,* Vol. 195, No. 4274, pages 185–187; January 14, 1977.

THE RETINEX THEORY OF COLOR VISION. Edwin H. Land in *Scientific American,* Vol. 237, No. 6, pages 108–128; December, 1977.

ION CHANNELS IN THE NERVE-CELL MEMBRANE

CHARGE MOVEMENT ASSOCIATED WITH THE OPENING AND CLOSING OF THE ACTIVATION GATES OF THE NA CHANNELS. Clay M. Armstrong and Francesco Bezanilla in *The Journal of General Physiology,* Vol. 63, pages 533–552; 1974.

KINETICS AND STEADY-STATE PROPERTIES OF THE CHARGED SYSTEM CONTROLLING SODIUM CONDUCTANCE IN THE SQUID GIANT AXON. R. D. Keynes and E. Rojas in *The Journal of Physiology,* Vol. 239, pages 393–434; 1974.

ORGANIZATION OF THE IONIC CHANNELS IN NERVE MEMBRANES. Richard D. Keynes in *The Nervous System, Vol. 1: The Basic Neurosciences,* edited by Donald B. Tower. Raven Press, Publishers, 1975.

IONIC CHANNELS IN EXCITABLE MEMBRANES: CURRENT PROBLEMS AND BIOPHYSICAL APPROACHES. Bertil Hille in *Biophysical Journal,* Vol. 22, pages 283–294; 1978.

DOLPHINS

MAMMALS OF THE SEA: BIOLOGY AND MEDICINE. Edited by Sam H. Ridgway. Charles C. Thomas, Publisher, 1972.

OCCURRENCE AND GROUP ORGANIZATION OF ATLANTIC BOTTLENOSE PORPOISES (*TURSIOPS TRUNCATUS*) IN AN ARGENTINE BAY. Bernd Würsig in *The Biological Bulletin,* Vol. 154, No. 2, pages 348–359; 1978.

ANCIENT CATAPULTS

DIE ANTIKEN GESCHÜTZE DER SAALBURG. Erwin Schramm. Weidmannsche Buchandlung, 1918.

THE MECHANICAL TECHNOLOGY OF GREEK AND ROMAN ANTIQUITY. A. G. Drachmann. E. Munksgaard, 1963.

GREEK CATAPULTS AND CATAPULT TECHNOLOGY: SCIENCE TECHNOLOGY AND WAR IN THE ANCIENT WORLD. Barton C. Hacker in *Technology and Culture,* Vol. 9, No. 1, pages 34–50; January, 1968.

GREEK AND ROMAN ARTILLERY. E. W. Marsden. Oxford at the Clarendon Press, 1969–71.

GEARS FROM THE GREEKS: THE ANTIKYTHERA MECHANISM—A CALENDAR COMPUTER. Derek de Solla Price in *Transactions of the American Philosophical Society, New Series,* Vol. 64, Part 7; 1974.

THE AMATEUR SCIENTIST

BOOMERANGS: MAKING AND THROWING THEM. Herb A. Smith. Gemstar Publications, Arun Sports, Littlehampton, Sussex, 1975.

THE PHYSICS OF FRISBEE FLIGHT. Jay Shelton in *Frisbee: A Practitioner's Manual and Definitive Treatise,* edited by Stancil E. D. Johnson. Workman Publishing Co., 1975.

MANY HAPPY RETURNS: THE ART AND SPORT OF BOOMERANGING. Benjamin Ruhe. The Viking Press, 1977.

Panasonic gives you fast relief from 8 major telephone headaches.

How to dial with a damaged digit.

The ham on rye and long distance sandwich.

But I'm up to my elbows.

Tell him to talk a little louder.

The number's here somewhere.

Now he calls!

(311) 555-8349? Or is it 8439?

The hand-in-the-cookie-jar emergency.

The Panasonic speaker phone and dialer.

For those times when your love affair with the telephone is more pain than pleasure, Panasonic gives you instant relief with two Easa-Phone™ accessories.*

With the Panasonic speaker phone you may never pick up a phone again. Because you won't have to. The touch of a finger (or an elbow, or a nose) turns your phone into a loudspeaker and receiver. Even with two hands full you can talk to your factory, salesmen, wife, or mother-in-law. Our deluxe model, the KX-T1030 (shown below), has a quartz LED clock and timer—especially important for long-distance calls.

And with the Panasonic automatic dialer, you can teach your telephone to dial its own calls. Our two models hold up to 32 different 16-digit numbers. The deluxe model, KX-T1220 (shown below) can hold up to 16 30-digit numbers. And it will electronically display the number you're calling. To get that number, just push a button. If the number is busy or doesn't answer, the dialer will keep re-dialing the number it just dialed up to 16 times.
No forgetting. No wrong numbers. What a relief! Especially for your dialing finger.

Both Easa-Phone accessories work with a variety of phones including multi-line business phones. Both are handsome in simulated woodgrain to look great on your wall or desk. Prices? Definitely affordable.

So when your telephone gives you a pain, take 2 Panasonics and get relief fast.

*It may be necessary for the phone company to install a new modular jack. All models are registered with the FCC.

Panasonic.
just slightly ahead of our time.

"CERTAINLY IT'S THE BEST EUROPE IN A LONG WHILE."

*The sedan that Road Test Magazine is talking about does not sell for $20,000. It does not sell for $15,000. It sells for a base price of $8995.**

This pleasantly surprising contradiction between the test results of the Audi 5000 and the price has been pointed out by several automotive publications.

Read these excerpts from the Road Test report on the Audi 5000 that appeared in the January, 1978 issue:

"At freeway speeds the ride is serene and smooth, there is no mechanical noise, and no wind noise. You get bored, turn on the radio, and just drive—until you glance down at the gauges and notice that you're making 85 mph. It's sneaky; it'll fool you if you don't pay attention. It'll also nip in and out of traffic like a much smaller car, thanks to the way it responds to its controls..."

"It may just be the ultimate sports sedan."

"...not only does the 5000 have a 5-cylindered mill, but it's got one that sips gas like a four,** makes horse-

*Suggested retail price $8995 P.O.E. Transportation, local taxes and dealer delivery charges additional. Fully equipped "S" model $10,575.

SEDAN TO COME FROM

—ROAD TEST MAGAZINE, ON THE AUDI 5000

power like a six and feels nearly as smooth as an eight."

"...the price...isn't near as much as some of the 5000's closest competitors sell for. That, of course, is the name of the game, and Audi is playing it to the hilt with development of solid concept and good attention to detail. They're going to sell every one of these cars they can screw together."

"...the 5000 has an incredible heater/air conditioner that could heat half your house, or could freeze your—um, toes off, depending on which button is pushed."

Test drive the 1979 edition of the Audi 5000 at your Porsche + Audi dealer and see if it isn't so.

**26 mpg highway, 16 mpg city, std. transmission. (EPA estimates. Actual mileage may vary based on how and where you drive, car's condition, and optional equipment.)

PORSCHE+AUDI

Nothing Even Comes Close

Just a reminder that the most impressive thing about Chivas Regal is what's in the bottle, not what's on it.

12 YEARS OLD WORLDWIDE · BLENDED SCOTCH WHISKY · 86 PROOF · GENERAL WINE & SPIRITS CO., NEW YORK, N.Y.