

SCIENTIFIC AMERICAN



ANCIENT OARED WARSHIPS

\$2.00

April 1981

TIME: Now
PLACE: Anyroad USA
ACTION: Creation of
America's New
Personal
Sport Coupe...



1982

FORD EXP

Introducing Ford EXP. Personal transport carefully created to carry you into the future . . . a new way to express yourself.

EXP's a beautiful package of world-class technology. Carefully instrumented . . . with handling and road command that come from front-wheel drive and four-wheel independent suspension. And fuel efficiency that comes from EXP's advanced 1.6 liter CVH engine and a fine-tuned aerodynamic shape.

And the sporty EXP's a 2 PLUS . . . practical as well as personal. Room for two plus surprisingly generous room for your gear . . . from sports equipment to loads of luggage. There's even a luggage hideaway. And EXP has a dazzling array of standard features:

- Console ■ Contoured Bucket Seats ■ All-Season Steel-Belted Radial Tires ■ Power Front Disc Brakes ■ Self-Adjusting Brakes and Clutch ■ Electric Rear Window Defroster ■ Intermittent Windshield Wipers ■ Styled Steel Wheels ■ Front Stabilizer Bar ■ Remote Control Sport Mirror ■ Rectangular Halogen Headlamps ■ Electro-Drive Cooling Fan ■ Carpeted Cargo Area ■ AM Radio with Dual Speakers (may be deleted for credit) ■ Visor Vanity Mirror ■ Tachometer ■ Temperature Gauge ■ Trip Mileage Counter ■ Door Ajar Warning Light ■ Low Washer Fluid Warning Light ■ New Manual Transaxle — Overdrive in 4th Gear (automatic optional)

Ford EXP. Personal. Practical. Technologically advanced. And built with Ford's commitment to quality.

Tomorrow is here
...from the
World of Ford.

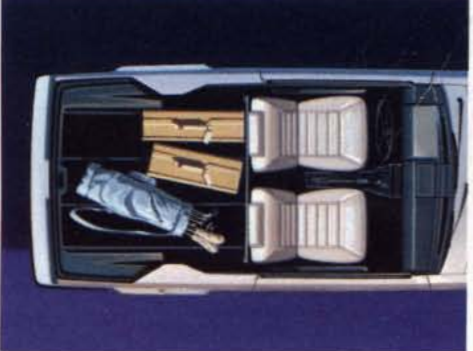
FORD EXP



Four-wheel independent suspension.



World-class technology: front-wheel drive.



Two-plus interior design: room for two plus a lot of room for a lot of living.



Projected Ford estimates for comparison. Your mileage may differ depending on speed, distance and weather. Actual highway mileage lower. See your Ford Dealer for actual EPA ratings.

Comfortably contoured and carefully instrumented for your driving pleasure.





The Chivas Regal of Liqueurs.

Chivas Brothers make the world's finest Scotch. And the same renowned Chivas quality and tradition goes into its equally-regal Lochan Ora Liqueur. With a unique taste all its own, Lochan Ora is to be lingered over—sipped slowly, savored fully. It's expensive but, then, good taste invariably is.

Lochan Ora.
The Imported Liqueur from Chivas Brothers.

70 PROOF • IMPORTED BY GENERAL WINE & SPIRITS CO., N.Y., N.Y.

ARTICLES

- 41 CATASTROPHIC RELEASES OF RADIOACTIVITY, by Steven A. Fetter and Kosta Tsipis** The worst reactor accident could not be as bad as a single nuclear-weapon explosion.
- 48 A UNIFIED THEORY OF ELEMENTARY PARTICLES AND FORCES, by Howard Georgi** At high energy there may be one kind of elementary particle and one major force.
- 64 SPEECH RECOGNITION BY COMPUTER, by Stephen E. Levinson and Mark Y. Liberman** Designing machines that can listen is much more difficult than making them speak.
- 88 THE ORIGIN OF GENETIC INFORMATION, by Manfred Eigen, William Gardiner, Peter Schuster and Ruthild Winkler-Oswatitsch** In the beginning it was carried by RNA.
- 120 THE SHELLS OF NOVAS, by Robert E. Williams** They are blown off by a white-dwarf star when a companion star spills fresh hydrogen fuel on it.
- 132 FILTER-FEEDING INSECTS, by Richard W. Merritt and J. Bruce Wallace** Insects of three orders hatch underwater and gather food with nets, brushes and other filters.
- 148 ANCIENT OARED WARSHIPS, by Vernard Foley and Werner Soedel** They evolved not only into the trireme but also into twin-hulled monsters with 4,000 oarsmen.
- 164 LITHIUM AND MANIA, by Daniel C. Tosteson** How is it that a substance as simple as an ion can affect a phenomenon as complex as mood?

DEPARTMENTS

- 6 LETTERS**
- 10D 50 AND 100 YEARS AGO**
- 15 THE AUTHORS**
- 18 MATHEMATICAL GAMES**
- 28 BOOKS**
- 78 SCIENCE AND THE CITIZEN**
- 176 THE AMATEUR SCIENTIST**
- 186 BIBLIOGRAPHY**

BOARD OF EDITORS Gerard Piel (Publisher), Dennis Flanagan (Editor), Brian P. Hayes (Associate Editor), Philip Morrison (Book Editor), Francis Bello, Peter G. Brown, Michael Feirtag, Paul W. Hoffman, Jonathan B. Piel, John Purcell, James T. Rogers, Armand Schwab, Jr., Joseph Wisnovsky

ART DEPARTMENT Samuel L. Howard (Art Director), Steven R. Black (Assistant Art Director), Ilil Arbel, Edward Bell

PRODUCTION DEPARTMENT Richard Sasso (Production Manager), Carol Hansen and Leo J. Petruzzi (Assistants to the Production Manager), Carol Eisler (Production Associate), Paul A. Harvey, Jr. (Assistant Production Manager), Carol Albert, Karen O'Connor, Martin O. K. Paul, Julio E. Xavier

COPY DEPARTMENT Sally Porter Jenks (Copy Chief), Nancy Ellen Bates, Mary Knight, Dorothy R. Patterson

GENERAL MANAGER George S. Conn

ADVERTISING DIRECTOR C. John Kirby

CIRCULATION MANAGER William H. Yokel

SECRETARY Arlene Wright



Pella Window and Door Ideas for older homes



Free Offer from Pella.

A book on renovation, remodeling and replacement with Pella wood windows and doors

Lots of handy facts and colorful photographs to help you make the right window choices. Everything you need to know about window styles and placement, plus information on how Pella windows conserve energy and reduce maintenance. Information on different styles of sliding glass doors and wood folding doors too. Send for this free 16-page brochure today.

Please send me a free copy of *Pella Window and Door Ideas for Older Homes*. I plan to remodel replace windows.

Name _____
 Address _____
 City _____
 State _____ Zip _____

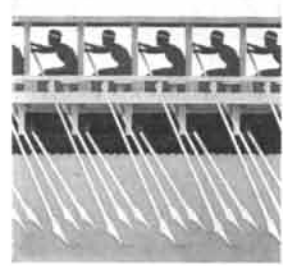
Telephone _____
 Mail to: Pella Windows & Doors, Dept. C17D1, 100 Main St., Pella, Iowa 50219

Also available throughout Canada
 This coupon answered within 24 hours



© 1981 Rolscreen Co

Pella. The significant difference in windows.



THE COVER

The painting on the cover is an idealized view of a section of the starboard side of a trireme, the standard warship deployed by the city-states of Greece during the fifth and fourth centuries B.C. (see "Ancient Oared Warships," by Vernard Foley and Werner Soedel, page 148). The ship illustrated carried a total of 170 rowers on three levels, one man to an oar. Only the rowers in the top rank are visible here; their oars pivoted on an outrigger, which extended about 60 centimeters from the side of the hull. The other two ranks of rowers sat inboard of the top rowers and lower in the hull. The oars of the middle rowers pivoted on the gunwale, or top plank of the hull. The oars of the bottom rowers penetrated the hull through oar ports, which were fitted with leather gaskets to keep out water. The Greek triremes of this period were designed primarily for ramming and carried only a small complement of marines on the deck above the top rowers. The hull was sealed with pitch, making the ship predominantly black.

THE ILLUSTRATIONS

Cover painting by George V. Kelvin

Page	Source	Page	Source
18-26	Ilil Arbel	133	Douglas A. M. Craig, University of Alberta (<i>top right and left and bottom right</i>); J. Bruce Wallace, University of Georgia (<i>bottom left</i>)
42-47	Alan D. Iselin		
49-63	Gabor Kiss		
65	Stephen E. Levinson and Mark Y. Liberman	134-135	Tom Prentiss
66-73	Jerome Kuhl	136	J. Bruce Wallace, University of Georgia
74	Stephen E. Levinson and Mark Y. Liberman	141-142	Tom Prentiss
75-76	Jerome Kuhl	144	Douglas A. M. Craig, University of Alberta (<i>top</i>); J. Bruce Wallace, University of Georgia (<i>bottom</i>)
89-118	Allen Beechel	149	Lionel Casson, New York University
121	Walken Graphics	150-161	George V. Kelvin
122	Steward Observatory	162	Lionel Casson, New York University
123	Yerkes Observatory (<i>top and middle</i>), Lick Observatory (<i>bottom</i>)	165-174	Albert E. Miller
124-128	Walken Graphics	177	R. G. Olsson and E. T. Turkdogan
131	Lick Observatory	178-184	Michael Goodman

A bird watcher's guide to Amoco's environmental efforts.



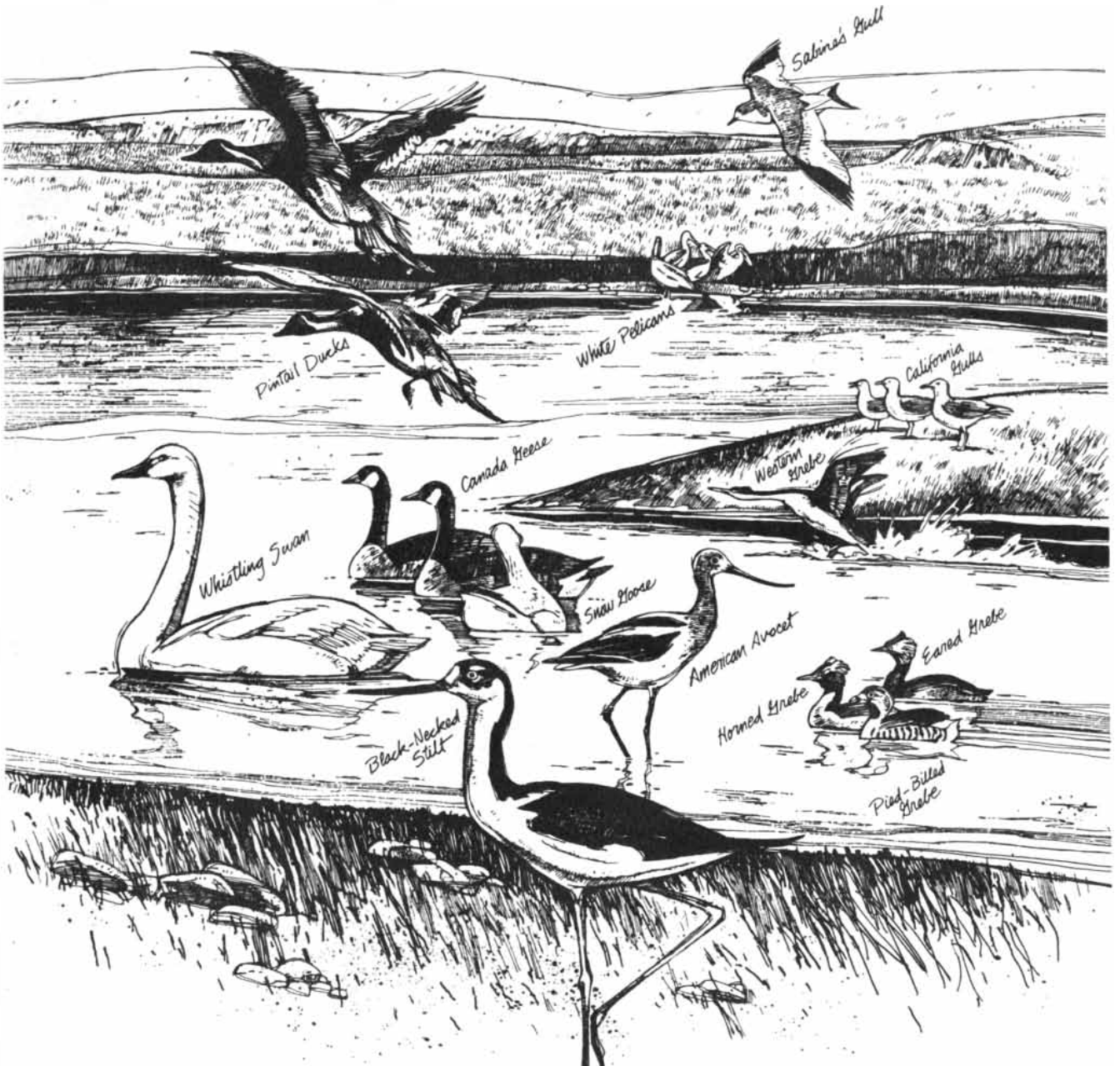
The harmonious coexistence of the energy industry and the natural environment is being proved at Amoco's Casper, Wyoming refinery. Biological management has turned a waste water pond into the state's most prolific habitat for birds and other wildlife. The refinery's fresh water effluent has decreased the pond's alkali content from 20,000 parts per million to 7,000. This remarkable improvement in water quality has created an ecosystem that supports more ducks and geese than anywhere else in the state.

The Audubon Society monitors the pond and now catalogues a wide variety of bird species—including grebes, geese, eagles, swans, and the rare Sabine's Gull—not found in the area before construction of Amoco's refinery.



This Wyoming wildlife habitat is a real example of how Amoco can supply energy to help meet America's needs while preserving the environment.

You expect more from a leader.



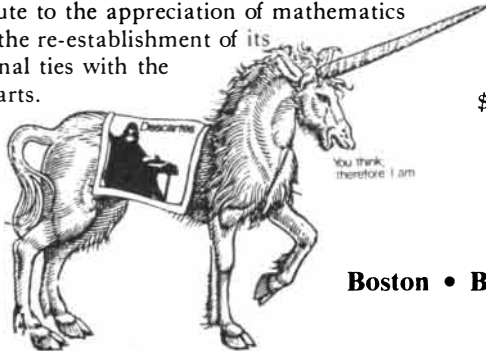
LETTERS

Like unicorns, mathematical objects are illusive.

The Mathematical Experience

Philip J. Davis, Reuben Hersh

The experience of mathematics – the science of numbers, quantities, shapes, relationships, and structures – has been with us since the beginning of civilization. Yet, mathematics and mathematicians have been continually suspect, but hardly understood. *The Mathematical Experience*, written by two distinguished mathematicians, helps to alleviate this lack of understanding with answers that are always informative, often amusing. The authors confront the reader with a personal view of the science and examine the internal and external factors that determine its structure and applications. *The Mathematical Experience* will greatly contribute to the appreciation of mathematics and to the re-establishment of its traditional ties with the liberal arts.



March 1981
460 pages
\$24.00; Hardcover

Birkhäuser
Boston • Basel • Stuttgart

At selected B. Dalton Bookstores —
or call toll free 1-800-228-2022; In Alaska: 1-907-276-3242;
In Nebraska: 1-402-571-4900; In Carolina (Puerto Rico):
1-809-752-1275 during store hours only.

B. Dalton

America's Favorite Bookseller

Sirs:

The kind of structural analysis engaged in, and the resulting questions raised by, Douglas R. Hofstadter in his amusing and intriguing article concerning self-referential sentences ["Metamagical Themas," *SCIENTIFIC AMERICAN*, January] need not lead inevitably to bafflement of the reader.

Help is at hand from the "laggard science" psychology, but only from that carefully defined quarter of psychology known as behavior analysis, which was progenerated by the famous Harvard psychologist B. F. Skinner almost 50 years ago.

In examining the implications of linguistic analyses such as Hofstadter's for the serious student of verbal behavior Skinner comments in his book *About Behaviorism* (pages 98-99) as follows:

"Perhaps there is no harm in playing with sentences in this way or in analyzing the kinds of transformations which do or do not make sentences acceptable to the ordinary reader, but it is still a waste of time, particularly when the sentences thus generated could not have been emitted as verbal behavior. A classical example is a paradox, such as 'This sentence is false,' which appears to be true if false and false if true. The important thing to consider is that no one could ever have emitted the sentence as

!Speak Spanish like a diplomat!

What sort of people need to learn a foreign language as quickly and effectively as possible? *Foreign Service personnel*, that's who. Members of America's diplomatic corps are assigned to U.S. embassies abroad, where they must be able to converse fluently in every situation.

Now you can learn to speak Spanish just as these diplomatic personnel do with the Foreign Service Institute's Programmatic Spanish Course.

The U.S. Department of State has spent tens of thousands of dollars developing this course. It's by far the most effective way to learn Spanish at your own convenience and at your own pace.

The Programmatic Spanish Course consists of a series of tape cassettes and an accompanying textbook. You simply follow the spoken and written instructions, listening and repeating. By the end of the course, you'll find yourself learning and speaking entirely in Spanish!

This course turns your cassette player into a "teaching machine." With its unique "programmatic" learning method, you set your own pace—testing yourself, correcting errors, reinforcing accurate responses.

The FSI's Programmatic Spanish Course comes in two volumes. You may order one or both courses:

- Volume I, Basic. (11 cassettes, 16 hours), instructor's manual and 464-page text, \$115
- Volume II, Advanced. (8 cassettes, 11½ hours), instructor's manual and 614 page text, \$98 (New York residents add sales tax.)

Your cassettes are shipped to you in handsome library binders.

TO ORDER, JUST CLIP THIS AD and mail with your name and address, and a check or money order. Or, charge to your credit card (American Express, VISA, Master Charge, Diners Club) by enclosing card number, expiration date, and your signature.

The Foreign Service Institute's Spanish course is unconditionally guaranteed. Try it for three weeks. If you're not convinced it's the fastest, easiest, most painless way to learn Spanish, return it and we'll refund every penny you paid! Order today!

Many other FSI language courses also available. Write us.

Audio-Forum
Dept. G-12
145 East 49th St.
New York, N.Y. 10017
(212)753-1783



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

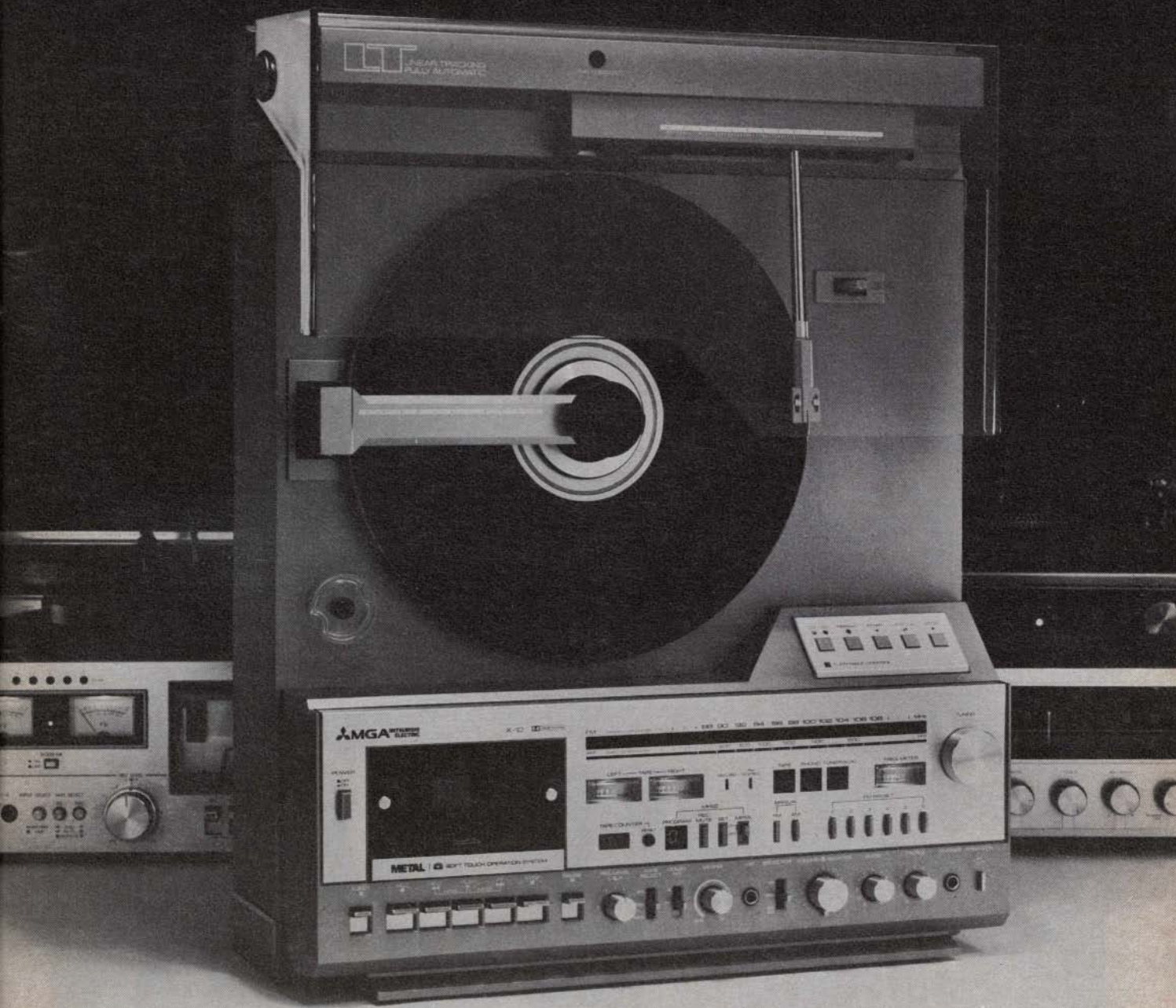
Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 60 cents.

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 _____



ONE SYSTEM STANDS ABOVE THE REST. VERTICALLY. SIMPLY. MITSUBISHI.

This is the system destined to turn the audio world on its ear. If you find that hard to believe, take another look. What you see is the InterPlay X-10.[™] The world's first *vertical format* complete audio system. No trick photography here. The way you see it is the way you play it. Vertically.

Play a record. Play a cassette. Play the tuner. Any way you play it, it's Mitsubishi.

The vertical *linear-track* turntable allows the X-10 to sit on a small bookshelf. It needs only 10" of shelf depth.

And we've integrated an AM/FM stereo tuner and cassette deck. Hook up the speakers, plug it in, and turn it on. The InterPlay X-10 is stereo at its upright best.

For your nearest dealer, call toll-free (800) 447-2882 or in Illinois, (800) 322-4400 and ask for operator X-10.

MGA/MITSUBISHI

MITSUBISHI ELECTRIC SALES AMERICA, INC. 3010 E. Victoria St., Compton, Ca. 90221



Shown here with optional CX-10 Cabinet and SX-10 Loudspeakers



**NEVER HAS
BEING FISCALLY
RESPONSIBLE
BEEN SO
MUCH FUN.**



It should now be obvious, even to the most insensitive observer, that, in America, the cavalier consumer has gone the way of the dinosaur.

Indeed, today, even those in the most heady income brackets are soberly peering under hoods, comparing fuel efficiency figures and scanning specification sheets with a vengeance.

Why then, in the face of all this circumspection, has the demand for a car as costly as the BMW 320i exceeded even the most optimistic predictions?

Perhaps because while the 320i appeals to people who know money, it also provides the sort of exhilarating performance that prompted one automotive journalist to proclaim it, "The quintessential sports sedan."

WHY CARS WITH SIMILAR MECHANICAL PARTS FAIL TO PERFORM LIKE A BMW.

Two decades ago, when the engineers at BMW suggested that only brilliant, relevant engineering could justify the price of an expensive sedan, considerable scoffing was heard abroad in the automotive community.

Now, of course, these same automakers are caught up in a revolution—cutting cylinders, redesigning suspension systems, converting to diesel engines, tinkering with all manner of electronic devices in a desperate attempt to bring their lagging technology in line with the realities of the 1980's.

The flaw in this frenetic technological scramble, as any automotive

expert will attest, is that evolution is always preferable to revolution.

And so the fact that the BMW 320i is the product of twenty years of development and millions of miles of testing gives the car and its driver a rather notable advantage.

Literally decades ahead of those who—while offering a litany of mechanical odds and ends—claim to perform "...nearly as well as the BMW."

A QUANTUM LEAP IN PERFORMANCE WITHOUT A CORRESPONDING DIP IN EFFICIENCY.

To cope with the very real needs of society and the equally pressing requirements of the gas pump, the automakers of the world have exhibited a tendency to sacrifice horsepower and favorable gearing for economy.

Not so the BMW 320i. Its performance remains uncompromised.

Take a spine-tingling curve and the perfectly balanced BMW suspension system—independent on all four wheels—flexes with a resiliency that makes one feel as if the car were slotted into the road.

Press the accelerator and the four-cylinder, hemi-head, overhead-cam engine responds in a manner that harks back to the days before environmental controls. (The secret is an emissions system ingeniously designed around a three-way catalytic converter with a Lambda sensor that electronically corrects the fuel/air mixture to

maintain a precise stoichiometric ratio.

Fuel efficient? Astonishingly enough, with five-speed standard transmission, the 320i delivers **[25]** EPA estimated mpg, 36 estimated highway mileage and, based on these figures, an estimated mpg range of **[383]** miles and a highway range of 551 miles.

(Naturally our fuel efficiency figures are for comparison purposes only. Your actual mileage and range may vary, depending on speed, weather and trip length. Your actual highway mileage and highway range will most likely be lower.)

A FINANCIER'S DREAM: A CAR THAT APPRECIATES.

As incredible as this statement sounds, the fact still remains that the October 1980 National Automobile Dealers Association Official Used-Car Guide shows the 1979 BMW 320i retained an extraordinary 105.3% of its original price on the used-car market—and an average over the past five years of 94%.

All of which explains why the editor-publisher of Car and Driver magazine—a gentleman not renowned for extravagant praise—recently wrote, "The BMW 320i is the sort of car that enthusiasts turn into legend."

If you find the notion of owning such a car intriguing, we suggest you phone your nearest BMW dealer and arrange a thorough test drive at your convenience.



THE ULTIMATE DRIVING MACHINE.

BMW, MUNICH, GERMANY

THE FIRST TRULY WATER RESISTANT ALARM/CHRONO DIGITAL



Casio's Mariner

You Don't Have Time To Worry About Your Watch

It's so convenient you just put it on and forget it! Never worry about the Mariner getting wet, because it's water resistant to 300 ft. Swim, ski and live the active life with this durable timepiece. Even the buttons have been engineered to prevent the loss of water resistancy during virtually any activity you undertake. Includes lithium batteries that won't need changing for approx. 5 years.

Use the Chronograph Without Losing Sight of the Time

The convenient and unique 12 digit LCD display allows you to view the time and your choice of the alarm, chronograph, countdown alarm, or calendar. Imagine a daily alarm, hourly chime, chronograph track accurate to 1/100th second, an 11 hr. countdown alarm, and self adjusting calendar at your command. Plus one touch allows you to change between 12 & 24 hour military/European time. All this backed by the manufacturer's one year warranty.

Buy with Confidence

We've become a leading media merchandiser by offering advanced technology products, prompt delivery, and a 30-day, no-risk, money-back guarantee. If this product isn't everything you expect, return it in new condition in 30 days or less for a complete refund.

Order Today

Charge customers call toll free 24 hours a day. Order product #91 for watch in tough black case and band, only \$39.95. Order product #93 for watch in stainless steel case and tough black band, only \$59.95.

(800) 854-3831 In California (800) 432-7451

DWS marketing international

350-A Fischer Avenue Dept. U01.
Costa Mesa, California. 92626.
(714) 540-4444

Call our toll-free number for quickest service or send your check money order, or credit card info. Please allow time for checks to clear. We accept VISA, M.C., A.E., D.C. charge cards. Add \$2.50 for insured postage & handling for first product & \$1.00 for each additional product. Calif. residents add 6% sales tax.

©DWS marketing international 1981

verbal behavior. A sentence must be in existence before a speaker can say, 'This sentence is false,' and the response itself will not serve, since it did not exist until it was emitted. What the logician or linguist calls a sentence is not necessarily verbal behavior in any sense which calls for a behavioral analysis."

As Skinner pointed out long ago, verbal behavior results from contingencies of reinforcement arranged by verbal communities, and it is these contingencies that must be analyzed if we are to identify the variables that control verbal behavior. Until we grasp the full import of Skinner's position, which goes beyond structure to answer *why* we behave as we do verbally or nonverbally, we shall continue to fall back on prescientific formulations that are about as useful in understanding these phenomena as Hofstadter's quaint metaphorical speculation: "Such a sentence would seem to be arrogantly proclaiming itself to be an animate agent."

GEORGE BRABNER, JR.

College of Education
University of Delaware
Newark, Del.

Sirs:

I assume that the quote from B. F. Skinner reflects Professor Brabner's own sentiments about the likelihood of self-referential utterances. I am always baffled by people who doubt the likelihood of self-reference and paradox. Verbal behavior comes in many flavors. Humor, particularly self-referential humor, is one of the most pervasive flavors of verbal behavior in this century. One has only to watch the Muppets or Monty Python on television to see dense and intricate webs of self-reference. Even advertisements excel in self-reference.

In art René Magritte, Pablo Picasso, M. C. Escher, John Cage and dozens of others have played with the level distinction between *that which represents* and *that which is represented*. The "artistic behavior" that results includes much self-reference and many confusing and sometimes exhilaratingly paradoxical tangles. Would Professor Brabner say that no one could ever have "emitted" such works as "artistic behavior"? Are paintings of paintings not examples of "artistic behavior"? Where is the borderline?

Ordinary language, as I pointed out in my column, is filled with self-reference, usually a little milder seeming than the very sharply pointed paradoxes Professor Brabner objects to. "Mouth," "word" and so on are all self-referential. Language is inherently filled with the potential of sharp turns on which it may snag itself.

Many scholarly papers begin with a sentence about "the purpose of this pa-

per." Newspapers report on their own activities, conceivably on their own inaccuracies. People say, "I'm tired of this conversation." Arguments evolve about arguments, and can get confusingly and painfully self-involved. Has Professor Brabner never thought of "verbal behavior" in this light? It is likely that in hunting woolly mammoths no one found it extraordinarily useful to shout, "This sentence is false." Civilization, however, has come a long way since those days, and the primitive purposes of language have by now been almost buried under an avalanche of more complex purposes.

Part of human nature is to be introspective, to probe. Part of our "verbal behavior" deliberately, often playfully, explores the boundaries between conceptual levels of systems. All of this has its root in the struggle to survive, in the fact that our brains have become so flexible that much of their time is spent in dealing with their own activities, consciously or unconsciously. It is simply a consequence of representational power—as Kurt Gödel showed—that systems of increasing complexity become increasingly self-referential.

It is quite possible for someone filled with self-doubt to recognize this trait in himself or herself, and to begin to doubt his or her self-doubt itself. Such psychological dilemmas are at the heart of some current theories of therapy. Gregory Bateson's "double bind," Victor Frankl's "logotherapy" and Paul Watzlawick's therapeutic ideas are all based on level-crossing paradoxes that crop up in real life. Indeed, psychotherapy is itself based completely on the idea of a "twisted system of self"—a self that wants to reach inward and change some presumably wrong part of itself.

We human beings are the only species to have evolved humor, art, language, tangled psychological problems, even an awareness of our own mortality. Self-reference—even of the sharp Epimenides type—is connected to profound aspects of life. Would Professor Brabner argue that suicide is not conceivable human behavior?

Finally, just suppose Professors Skinner and Brabner are right and no one ever says exactly, "This sentence is false." Would this mean that study of such sentences is a waste of time? Still not. Physicists study ideal gases because they represent a distillation of the most significant principles of the behavior of real gases. Similarly, the Epimenides paradox is an "ideal paradox"—one that cuts crisply to the heart of the matter. It has opened up vast domains in logic, pure science, philosophy and other disciplines and will continue to do so in spite of the skepticism of behaviorists.

DOUGLAS R. HOFSTADTER

Palo Alto, Calif.



We believe that when you travel on business, there are certain things you deserve.

You deserve service that's not only fast, but friendly. From people who understand that warm attitudes are just as helpful to travelers as cool efficiencies.

You deserve a personal welcome. A lot of respect. Extra help whenever you need it. And a commitment on our part to give you a level of attention that may surprise you, and that will certainly

make renting a car from National one of the better experiences of your business trip.

Try us. Contact your travel consultant. Or call toll-free 800-328-4567. In Minnesota call 800-862-6064. In Alaska and Hawaii call 800-328-6321. In Canada call collect 612-830-2345. We're going

to prove it: There is a difference between car rental companies, and it's called "National attention.SM"



We offer S&H Green Stamp Certificates in all 50 U.S. states.



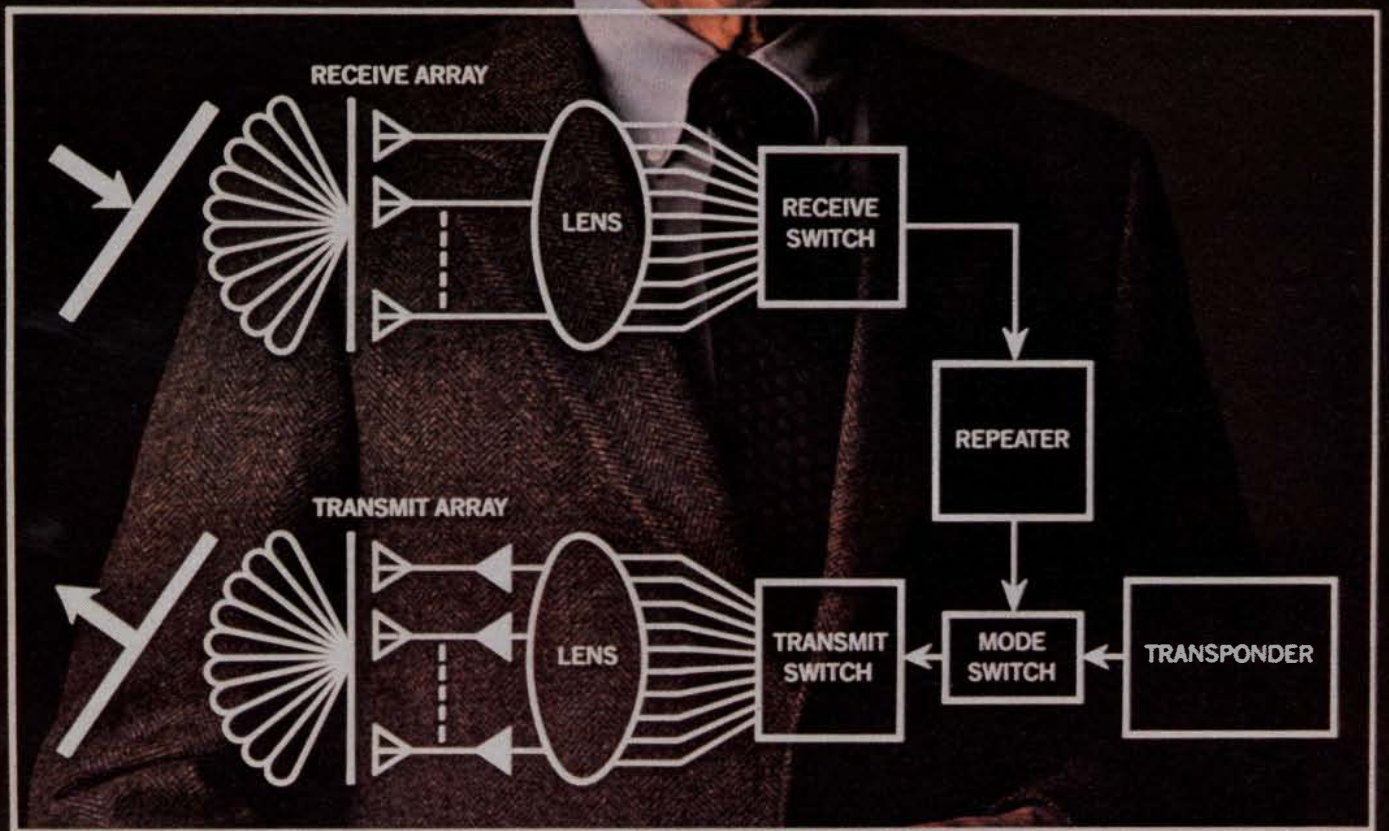
We feature GM cars like this Buick Skylark.

You deserve National attention.

National Car Rental

©1981, National Car Rental System, Inc. In Canada it's Tilden. In Europe, Africa, and the Middle East it's Europcar.

© 1981 SCIENTIFIC AMERICAN, INC



“How new multibeam technology puts an electronic shield over Navy ships.”

“When the Navy needed a new shipboard ECM system, the specs were formidable,” says Don Archer, Consulting Scientist at Raytheon.

“Ships at sea are exposed to attack by any number of electronically-guided missiles, coming in from many directions at high speeds, and operating at different frequencies.

“The problem? Devise a system that could detect, track, classify, and—as required—electronically jam or confuse many missiles simultaneously.

“Conventional systems, using a single antenna, would have restricted us in several ways: they’re limited in effective radiated power and can only receive or transmit on a very narrow bandwidth.

“Our solution was to develop a multibeam antenna made up of an array of elements fed by dielectric lenses. In effect, it’s a series of small, individual antennas, each scanning a portion of the sky.

“The resulting multibeam system can listen to a broad range of discrete frequencies at the same time. It monitors large numbers of incoming signals, on many different frequencies, and determines course, speed, and type of target.

“When a jamming response is called for, the system can send out multiple signals simultaneously, on the same frequencies as the incoming signals—and each outgoing signal can have the full power of the transmitter.”

Putting an electronic shield over the fleet is typical of how Raytheon is using advanced technology in new ways to solve difficult problems.

Raytheon... a \$5 billion company in electronics, aviation, appliances, energy, construction, and publishing. For further information, write Raytheon Company, Public Relations, Section 2-3, 141 Spring Street, Lexington, Massachusetts 02173.



Candidates interested in technical employment opportunities with Raytheon Company, An Equal Opportunity Employer, should forward a resume to Manager, Executive Placement, at the above address.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

APRIL, 1931: "A new type of stroboscope, a device for 'stopping' motion to study the behavior of machines operating at high speed, has been developed at the Massachusetts Institute of Technology by Harold E. Edgerton. The unique feature of the new instrument is the electrical circuit that causes a condenser to discharge periodically through a thyratron mercury-arc tube. An intense blue light of extremely short duration is produced by a large current through the tube and makes it possible to adapt the stroboscope for photographic as well as visual observation. So powerful is the new instrument that still and motion pictures of a 160-horsepower electric motor have been made while the machine was running at full speed. Although the rotor was turning over at a rate corresponding to a ground speed of 95 miles per hour, the moving parts were shown as clearly as if the machine were standing still."

"At Daytona Beach, Fla., Captain Malcolm Campbell on February 4 raced his specially built *Bluebird* along the sands at the tremendous speed of 245.733 miles per hour. Before passing the measured mile of the course he had made a flying start of five and a half miles. His achievement puts him far ahead of the world record for land vehicles, which was held by the late Sir Henry Seagrave, whose speed was 231.3 miles per hour. The *Bluebird* is driven by a 12-cylinder Napier-Lion aviation engine, the rating of which is said to be 1,450 horsepower. The body and wheels are carefully streamlined, and a large tail fin, similar to a rudder except that it is rigid, acts as a stabilizer."

"A French company backed by the French government is preparing to rush to completion its section of pipe-line from the Iraq oil fields to Tripoli in Lebanon. The proposed British terminal of this same pipe-line will be in Haifa and will serve both British and American oil companies interested in the Iraq fields. The British and American interests do not wish to hasten construction of their end of the pipe-line and at present contemplate leisurely construction with completion about 1938. France, eager to obtain a supply of oil wholly under her own control, desires to finish her section in 1933. As both French and British titles to the oil fields come from the

World War, there has already been much negotiation on the division of the oil. It is easy to understand both the French desire to control their own supply of oil and the opposition of the British and American companies already struggling with a worldwide overproduction to an additional supply thrown on the world market. We have watched the desperate and to date futile efforts of domestic oil companies to control the output of oil in the United States. The future developments will be interesting to observe."



APRIL, 1881: "Can we hope that the microscope will show us an atom or a molecule? Huxley said three years ago that if the makers of microscope objectives could not produce lenses that would show spaces between lines one 100,000th of an inch apart, a molecule would never be seen. But last year Helmholtz saw Nobert's lines ruled 112,000 to the inch, and yet the molecule is still unseen. The striae on *Amphipleura pellucida*, numbering 132,000 to the inch, are now plainly resolved by the best objectives of wide angular aperture. Helmholtz has determined the approximate diameter of a single molecule of water, which he estimates as one 250,000,000th of an inch. At a recent meeting of the Elmira (N.Y.) Microscopical Society, Dr. Ford described Helmholtz's method of measurement by means of a soap bubble and the wave lengths of light. Professor Ford left his attentive audience to infer that an atom and a molecule would probably never be revealed, notwithstanding the great improvements likely to be developed in microscope objectives."

"Among the various modes of transmitting power to a distance are (I) transmission by wire ropes, (II) transmission by compressed air, (III) transmission by pressure water and (IV) transmission by electricity. On the last head, however high the efficiency of an electric motor may be in relation to the chemical work of the electric battery that feeds it, the force generated by an electric battery is too expensive, on account of the nature of the materials consumed, for a machine of this kind ever to be employed for industrial purposes. If, however, the electric current, instead of being developed by chemical work in a battery, is produced by ordinary mechanical power in a magneto-electric or dynamo-electric machine, the case is different; the double transformation, first of the mechanical force into an electric current and then of that current into a mechanical force, furnishes a means for effecting the conveyance of the power to a distance. In order to achieve this desir-

able result the three following points should form the aim of experimentalists: first, the determination of the efficiency of the principal kinds of magneto-electric or dynamo-electric machines working as generators; second, the determination of the efficiency of the same machines working as motors; third, the determination of the law according to which the magnetism of the cores of these machines varies with the intensity of the current."

"H. Reinke has examined protoplasm obtained from *Aethalium septicum* and has discovered in it the following approximate constituents: plastin (an insoluble albuminoid resembling the fibrins), vitellin, myosin, pepton, peptonoid, pepsin, nuclein, lecithin, guanin, sarcin, xanthin, ammonium carbonate, parcholesterin, traces of cholesterin, *Aethalium* resin, a yellow pigment, glycogen, sugar (non-reductive), oleic, stearic, palmitic and butyric acids, carbonic acid, fatty glycerides and parcholesterides, calcium stearate, palmitate, oleate, lactate, oxalate, acetate, formiate, phosphate, carbonate, sulphate (traces), magnesium (probably phosphate), potassium phosphate, sodium chloride, iron (compound not determined) and water. Plastin can be separated by pressure from the liquid portions of the protoplasm. The albuminoids collectively scarcely amount to 30 per cent of the dry substance. Hence the supposition that protoplasm consists of albumin must be abandoned, and we must cease to compare a cell to a particle of white of egg."

"Many attempts have been made to produce ammonia from the nitrogen of the air, until now without success. Some four years ago, however, Julien patented a process for forming ammonia by the action of the electric spark on a mixture of hydrogen and nitrogen. In a new patent Muller and Geisenberger draw the combustion gases of a furnace through caustic lime, where they are freed from carbonic acid, so that nearly pure nitrogen remains. In another apparatus hydrogen gas is produced by bringing water in contact with ignited coke, and the two gases, hydrogen and nitrogen, pass into a receptacle, where they are thoroughly mixed and subjected to the action of electric sparks. The ammonia is removed as soon as it is formed."

"*La Nature* brings news of a new system of scales in which sets of separate and detached weights are dispensed with, along with the various annoyances connected with their use. In these new scales, which are the invention of M. Coulon of Paris, the weighing is effected instantaneously by means of metal slides moving on graduated beams, which form a part of the apparatus."

HERE'S TO GUT FEELINGS AND THOSE WHO STILL FOLLOW THEM.

Ted Turner does lots of things people advise him not to do. And he succeeds at them. He turned Atlanta's WTBS-TV into a "Superstation" using a communications satellite and recently founded Cable News Network, the world's first 24-hour TV news network. He bought the Atlanta Braves and moved them out of last place; won the 1977 America's Cup after being fired in the '74 races; and was named "Yachtsman of the Year" four times.

Ted Turner puts his feelings where his mouth is. He also puts a great scotch there: Cutty Sark. And while he's been called Captain Outrageous by some, one thing's sure: Ted Turner's enjoying himself.



CNN
Ted "Captain Outrageous" Turner

The Scotch with a

following of leaders.

CUTTY SARK



One Stamford Forum, Stamford, Conn. 06904

© 1981 SCIENTIFIC AMERICAN, INC

“Grow trees underground?
GEE!”

(No, GTE!)

Normally, when trees are grown and replanted there's only a 50% survival rate. But with these trees, there's a 95% survival rate.

That's because these trees are grown inside a mountain in Idaho, in an environment that's ideal for plant growth. The air temperature, carbon dioxide content and humidity are all constant, and there are no major disease or insect problems.

Light was the only growth element missing. Which was solved by installing special Sylvania Super-Metalarc lamps. These lights help the young trees not only thrive but develop a more fibrous root system which aids in the transplantation.

Who said you can't improve on nature?

GTE

AT LAST, AN IRRATIONAL REASON TO BUY A VOLVO.



Over the years, Volvo has come to be known as one of the most sensible cars you can own.

So, sadly, people who pick their cars for more emotional reasons may have denied themselves the satisfaction of driving one.

But now there's a Volvo even for those to whom practicality takes second place.

The new Volvo GLT. Its striking appearance is the first hint that the GLT is quite unlike its predecessors. But the proof's behind the small-diameter GT wheel.

The GLT is available with a fuel-injected, overhead-cam, turbo-charged engine that can propel the car from 0-60 in 9.6 seconds.

And its handling offers a feature even harder to find: predictability. The GLT is designed to react the same in panic situations as it does in normal driving. So it not only presents the driver with no surprises, it can actually improve his performance.

Every GLT also comes fully-equipped with Volvo's more traditional virtues. Like an interior capable of comfortably accommodating five full-size adults. Orthopedically contoured seats. And the superb engineering that's responsible for Volvo's famed life-expectancy: in Sweden, by latest count, an average 18.7 years.

Before you purchase a performance car, test-drive the new Volvo GLT.

The pure exhilaration you'll feel may have little to do with logic. But if your heart insists on a GLT, relax.

Your mind will be getting a Volvo.

VOLVO

A car you can believe in.

© 1981 VOLVO OF AMERICA CORPORATION

THE AUTHORS

STEVEN A. FETTER and KOSTA TSIPIIS ("Catastrophic Releases of Radioactivity") are at the Massachusetts Institute of Technology, where they are associated with the Program in Science and Technology for International Security of the department of physics. Fetter, who is an undergraduate at M.I.T., expects to receive his S.B. in physics in June. In addition to the work reported here, he writes, "I have done research on the history of the industrialization of American agriculture and on energy use in agriculture. I am also interested in energy policy, nuclear disarmament and education." Tsiapis is associate director of the Program in Science and Technology. A native of Greece, he came to the U.S. to study electrical engineering and physics in 1954. He has a B.Sc. and an M.Sc. from Rutgers University and a Ph.D. from Columbia University. He joined the physics department at M.I.T. in 1966. Since 1973 his research has been devoted primarily to technical questions involved in efforts to limit nuclear weapons and to analyses of new weapons systems.

HOWARD GEORGI ("A Unified Theory of Elementary Particles and Forces") is professor of physics at Harvard University. A graduate of Harvard College, he got his Ph.D. from Yale University in 1971. He then returned to Harvard, where he held a succession of postdoctoral fellowships before joining the faculty in 1976. In recent years he has published numerous papers, many in collaboration with other leading physicists, on aspects of the unification of the laws of physics.

STEPHEN E. LEVINSON and MARK Y. LIBERMAN ("Speech Recognition by Computer") are on the technical staff of Bell Laboratories. Levinson has a B.A. in engineering sciences from Harvard College and an M.S. and a Ph.D. in electrical engineering from the University of Rhode Island. Before going to work for Bell Laboratories in 1976 he taught computer science for several years at Yale University. Liberman's Ph.D., in linguistics, is from the Massachusetts Institute of Technology. He has been at Bell Laboratories since 1975, except for a semester of teaching at M.I.T. in 1978.

MANFRED EIGEN, WILLIAM GARDINER, PETER SCHUSTER and RUTHILD WINKLER-OSWATITSCH ("The Origin of Genetic Information") prepared their article at the Max Planck Institute for Biophysical Chemistry in Göttingen, where Eigen is director of the department of biochemical kinetics. In addition to his theoretical

work on molecular evolution Eigen continues his experimental investigation of a wide range of problems in reaction kinetics, employing the chemical-relaxation methods for which he was awarded a Nobel prize in 1967. He was educated at Göttingen, receiving his doctoral degree in 1951. Gardiner, who is professor of chemistry at the University of Texas at Austin, came to the field of molecular evolution from experimental and computer-simulation studies of gas kinetics and explosions. His degrees are from Princeton University (A.B., 1954) and Harvard University (Ph.D., 1960). Schuster, who got his doctorate from the University of Vienna in 1967, is professor and director of that university's Institute for Theoretical Chemistry and Radiochemistry. He combines his interest in the dynamics of reacting systems with studies of hydrogen bonding. Winkler-Oswatitsch acquired her doctorate from the Technical University of Vienna in 1969. Her research interests range from the kinetics of reactions of ions with antibiotics to evolutionary problems and the phylogenetic analysis of RNA. She and Eigen coauthored the book *Das Spiel (The Game)* in 1975.

ROBERT E. WILLIAMS ("The Shells of Novas") is professor of astronomy at the University of Arizona and an astronomer at the university's Steward Observatory. He did his undergraduate work at the University of California at Berkeley and his graduate work at the University of Wisconsin at Madison, obtaining his Ph.D. from the latter in 1965. He has been a member of the faculty at Arizona ever since. In 1972 he spent a sabbatical as a Fulbright lecturer at University College London. Williams' main research interests, apart from nova shells, include the analysis of the radiation from quasars and the structure of accretion disks in close binary-star systems.

RICHARD W. MERRITT and J. BRUCE WALLACE ("Filter-feeding Insects") are entomologists with a common interest in the ecology of insects that live in streams and rivers. Merritt is associate professor of entomology at Michigan State University; he also has a joint appointment with the Kellogg Biological Station in Hickory Corners, Mich. His degrees are from California State University at San Jose (B.A., 1968), Washington State University (M.S., 1970) and the University of California at Berkeley (Ph.D., 1974). His doctoral dissertation at Berkeley, he writes, "involved an ecological study of the insect complex associated with cattle droppings. Although my major research emphasis has changed to stream

SAVE BIG on CALCULATORS

TEXAS INSTRUMENTS

TI 59.....\$200	TI MBA.....53
TI 58C.....89	TI BA2.....39
TI PC-100C...157	TI Bus Anal I...15
TI 58/59 Lib...32	TI Bus Card...39
TI PROG.....47	TI Inv. Anal...42
TI 57.....29	TI 5040.....60



TI 59
\$200.00

HEWLETT PACKARD CALCULATORS

HP-41C.....\$189
HP-41C V.....239

Optical Wand.....95
Card Reader.....169
Printer.....289
Quad R.A.M. (for HP41C)....75



HP-41C
\$189

System I.....\$380
(HP 41C V + Card Reader)

System II.....640
(HP-41C V + Card Reader, Printer)

HP-97.....\$570
HP-67.....289
HP-38C.....115
HP-37E.....59
HP-34C.....115
HP-33C.....69
HP-32E.....43

HP-85 COMPUTING SYSTEM



The HP-85 Personal Computing System for Professionals helps solve difficult engineering problems and simplifies financial analysis. The HP-85 is a powerful BASIC language computer with keyboard, CRT display, printer and tape drive complete in one compact unit. (HP-83 is the same as above but without the built-in printer and tape drive.)

	Mfr. Sugg. Ret.	Elek-Tek Price
HP-85 COMPUTER	\$3200	\$2600
HP-83 COMPUTER	2250	1800
as HP-85 but no Printer or Tape Drive)		
26318 PRINTER	3650	3000
7225A PLOTTER (incl. Personality Module)	2800	2450
VisiCalc PLUS	200	180
APPLICATION PACKS	95	81
HP-85 SOL BOOKS	10	9
and all related accessories and software at SIZEABLE DISCOUNTS		

CALL TOLL FREE 800-621-1269
EXCEPT Illinois, Alaska, Hawaii

Accessories discounted too. Corporate Accounts Invited. Mastercharge or Visa by mail or phone. Mail Cash. Ck. Mon. Ord Pers. Ck. [2 wks to clear]. Add \$4.00 1st. item, \$1.00 ea. add'l shpg. & handl. Shipments to IL address add 6% tax. Prices subject to change. WRITE for free catalog. ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE

ELEK-TEK, inc.
5344 W. Devon Ave., Chicago IL 60648
(800) 621-1269 (312) 631-7800

Celestron[®]

C90 TELEPHOTO



C90 PHOTO OF TIGER AT 30 FEET

C14 SATURN PHOTO

C90 TELESCOPE

For the photographer, astronomer, naturalist, educator, or casual observer — Celestron has established a new standard in superb telescopes and telephoto lenses at affordable prices (from \$295). All feature large observatory mirror/lens type optics folded into a compact, lightweight, portable telescope or telephoto lens. Close-up detail borders on the fabulous from celestial objects light years away, to ships, planes, flowers, wildlife or people. Here is truly an instrument that allows you to observe or photograph the world around you. Celestron also has a full line of binoculars of astronomical quality.

Send \$2.00 for 32-page full color catalog on how to select and use a Celestron telescope or telephoto lens.

Celestron International
2835 Columbia St., Box 3578-SA
Torrance, Calif. 90503, U.S.A.
Telephone: (213) 328-9560



C8 Telescope

(Dealer Inquiries invited.)

ecology at Michigan State, there are many similarities between the two habitats with respect to the insect fauna inhabiting them." Wallace is professor of entomology at the Institute of Ecology of the University of Georgia. A graduate of Clemson University, he got his M.S. and his Ph.D. from the Virginia Polytechnic Institute and State University in 1963 and 1967 respectively. His current work is on the response of aquatic ecosystems to alterations in the watershed.

VERNARD FOLEY and **WERNER SOEDEL** ("Ancient Oared Warships") have collaborated on a variety of topics in the history of science and technology, among them the engineering of the ancient catapult; it was this project that led to their article "Ancient Catapults" in the March 1979 issue of *SCIENTIFIC AMERICAN*. Both are at Purdue University. Foley, who is associate professor of history, has degrees from McPherson College, the University of Kansas and the University of California at Berkeley. He joined the Purdue faculty in 1970, soon after getting his Ph.D. in history from Berkeley. He is currently investigating Leonardo da Vinci's role in the development of the milling machine. Soedel, a professor of mechanical engineering, was born in Prague and educated mainly in West Germany, where he received his first engineering degree. He also has an M.S. and a Ph.D. from Purdue. In addition to his interest in the history of engineering he teaches and does research on the mechanics of vibration-related phenomena.

DANIEL C. TOSTESON ("Lithium and Mania") is dean of the Harvard Medical School, where he is also Caroline Shields Walker Professor of Physiology. He was graduated from Harvard College in 1944 and from the Harvard Medical School in 1949. After completing his internship and residency at Presbyterian Hospital in New York, he worked as a research fellow at the Brookhaven National Laboratory, the National Heart Institute, the Biological Isotope Research Laboratory in Copenhagen and the Physiological Laboratory in Cambridge, England. He began his academic career in 1958 as associate professor of physiology at the Washington University School of Medicine and moved on in 1961 to become professor and chairman of the department of physiology at the Duke University Medical Center. Before returning to Harvard in 1977 he was Lowell T. Coggeshall Professor of Medical Sciences and dean of the Pritzker School of Medicine and vice-president of the Medical Center of the University of Chicago. His research, Tosteson writes, "is directed toward understanding the cellular functions and molecular mechanisms of ion transport across membranes."

Choose one of these five sets and save up to \$216¹⁵

You simply agree to buy 4 books within the next two years
as a member of the Book-of-the-Month Club.[®]



The Compact Edition of The Oxford English Dictionary

The "most complete, most scholarly dictionary of the English language"—*The Christian Science Monitor*. Through photo-reduction, the original 13-volume set has been reproduced in this two-volume Compact Edition. Magnifying glass included.

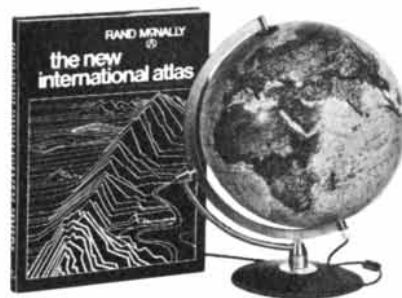
~~\$125.00~~ Publisher's price
~~-\$105.05~~ Your savings
\$ 19.95 Book-of-the-Month Club price



The Complete Beethoven Nine Symphonies

Sir Georg Solti conducts the Chicago Symphony in this nine-record set, hailed by the *San Francisco Sunday Examiner & Chronicle* as "one of the two or three great recording ventures of the century," and by *Time* as "a glorious musical combustion."

~~\$71.82~~ List price
~~-\$61.82~~ Your savings
\$ 10.00 Book-of-the-Month Club price



The New International Rand McNally Atlas & the Hammond Geo-Scan Globe

This new, enlarged (11 $\frac{3}{8}$ " x 15 $\frac{1}{2}$ ") atlas features 323 pages of distinctive, full-color maps and charts, 160,000-place index. The Hammond Geo-Scan Globe is 16" high, 12 $\frac{3}{4}$ " in diameter. Interior light illuminates world's political boundaries.

~~\$130.00~~ Total price for both
~~-\$102.50~~ Your savings
\$ 27.50 Book-of-the-Month Club price



The Story of Civilization by Will and Ariel Durant

For more than a generation, Will and Ariel Durant have been tracing the continuity of our history—economic and political organization, science and art, religions, philosophies, customs and manners—to show the foundations of society today. Now complete, the Durants' 11-volume masterwork is history come alive. The enormous work covers ancient and modern civilization, including Oriental as well as Western history. With clarity and authority the Durants illuminate "the steps by which man passed from barbarism to civilization." This set is a magnificent centerpiece for any home library, and one of the most useful sources of knowledge that the modern reader can possess.

~~\$241.10~~ Publisher's prices total
~~-\$216.15~~ Your savings
\$ 24.95 Book-of-the-Month Club price



The Encyclopedia of Philosophy

The most comprehensive encyclopedia of philosophy ever published, this set traces ancient, medieval, modern, Eastern and Western thought. An essential and rewarding reference source for home libraries.

~~\$150.00~~ *Publisher's price
~~-\$130.05~~ Your savings
\$ 19.95 Book-of-the-Month Club price

*While the publisher's price has risen from \$125 to \$150, our special price remains the same.

Prices shown are publishers' U.S. prices. Outside the U.S., prices are generally somewhat higher.

Facts About Membership. You receive the *Book-of-the-Month Club News*[®] 15 times a year (about every 3 $\frac{1}{2}$ weeks). Each issue reviews a *Main Selection* plus scores of Alternates. If you want the *Main Selection* do nothing. It will be shipped to you automatically. If you want one or more Alternate books—or no book at all—indicate your decision on the reply form always enclosed and return it by the date specified. **Return Privilege:** If the *News* is delayed and you receive the *Main Selection* without having had 10 days to notify us, you may return it for credit at our expense. **Cancellations:** Membership may be discontinued, by either you or the Club, at any time after you have purchased 4 additional books.

Book-of-the-Month Club, Inc., Camp Hill, Pennsylvania 17012

A170-4

Please enroll me as a member of Book-of-the-Month Club and send me the work I have checked below, billing me for the appropriate amount, plus shipping and handling charges. I agree to buy 4 books during the next two years. A shipping and handling charge is added to each shipment.

CHECK ONE BOX ONLY

904. Beethoven Symphonies **\$10**
912. Compact OED **\$19.95**
913. The Story of Civ. **\$24.95**
917. Ency. of Philosophy **\$19.95**
926. McNally/Hammond Set **\$27.50**

Mr. _____ I-61
Mrs. _____
Miss _____ (Please print plainly)

Address _____ Apt. _____

City _____

State _____ Zip _____

BOOK-OF-THE-MONTH CLUB[®]

MATHEMATICAL GAMES

How Lavinia finds a room on University Avenue, and other geometric problems

by Martin Gardner

Most of the short problems that make up this month's column involve geometry, taking the word in a broad sense to include such areas as combinatorial geometry, topology and graph theory. All the problems will be answered in my next column. Comments on letters from readers cannot appear until at least four months after that.

1. Lavinia Seeks a Room. The line in the illustration below represents University Avenue in a small college town where Lavinia is a student. The spots labeled *A* through *K* are buildings along the avenue in which Lavinia's 11 best friends are living.

Lavinia has been living with her parents in a nearby town, but now she wants to move to University Avenue. She would like a room or an apartment at a location *L* on the street that minimizes the sum of all its distances from her 11 friends. Assuming that a place is available at the right location, specify where Lavinia should live and prove it does make the sum of all its distances to the other locations as small as possible.

2. Mirror-symmetric Solids. On plane figures an axis of symmetry is a straight line that divides the figure into congruent halves that are mirror images of each other. The hearts on playing cards, for example, have one axis of symmetry. So do spades and clubs, but the diamond has two such axes. A square has four axes of symmetry. A regular five-pointed star has five, and a circle has an infinite number. A swastika or a yin-yang symbol has no axis of symmetry.

If a plane figure has at least one axis of symmetry, it is said to be superposable on its mirror image in the following sense. If you view the figure in a vertical mirror, with the edge of the mirror resting on a horizontal plane, you can imagine sliding the figure into the mirror and,

if necessary, rotating it on the plane so that it coincides with its mirror reflection. You are not allowed to turn the figure over on the plane because that would require rotating it through a third dimension.

A plane of symmetry is a plane that slices a solid figure into congruent halves, one half a mirror reflection of the other. A coffee cup has a single plane of symmetry. The Great Pyramid of Egypt has four such planes. A cube has nine: three are parallel to a pair of opposite faces and six pass through corresponding diagonals of opposite faces. A cylinder and a sphere each have an infinite number of planes of symmetry.

Think of a solid object bisected by a plane of symmetry. If you place either half against a mirror, with the sliced cross section pressing against the glass, the mirror reflection, together with the bisected half, will restore the shape of the original solid. Any solid with at least one plane of symmetry can be superposed on its mirror image by making, if necessary, a suitable rotation in space.

Discussing this in my book *The Ambidextrous Universe* (Charles Scribner's Sons, 1979), I stated on page 19 that if a three-dimensional object had no plane of symmetry (such as a helix, a Möbius strip or an overhand knot in a closed loop of rope), it could not be superposed on its mirror image without imagining its making an impossible rotation that "turned it over" through a fourth dimension.

The statement is false! As many readers of the book pointed out, there are solid figures totally lacking a plane of symmetry that nonetheless can, by a suitable rotation in ordinary space, be superposed on their mirror images. In fact, one is so simple that you can fold it in a trice from a square sheet of paper. How is it done?

3. The Damaged Patchwork Quilt. The patchwork quilt in the top illustration on page 23, of size 9 by 12, was originally made up of 108 unit squares. Parts of the quilt's center became worn, making it necessary to remove eight squares as indicated.

The problem is as follows. Cut the quilt along the lattice lines into just two parts that can be sewn together to make a 10-by-10 square quilt. The new quilt cannot, of course, have any holes. Either part can be rotated, but neither one can be turned over because the quilt's underside does not match its upper side.

The puzzle is an old one, but the solution is so beautiful and the problem is so little known that I constantly get letters about it from readers who are not aware of its origin. The solution is unique even if it is not required that the cutting be along lattice lines.

4. Acute and Isosceles Triangles. An acute triangle is one with each interior angle less than 90 degrees. What is the smallest number of nonoverlapping acute triangles into which a square can be divided?

I asked myself that question some 20 years ago, and I solved it by showing how to cut a square into eight acute triangles as is indicated in the top figure in the bottom illustration on page 23. Reporting this in a column, reprinted as Chapter 3 of my *New Mathematical Diversions from Scientific American* (Simon and Schuster, 1966), I said: "For days I was convinced that nine was the answer; then suddenly I saw how to reduce it to eight."

Since then I have received many letters from readers who were unable to find a solution with nine acute triangles but who pointed out that solutions are possible for 10 or any higher number. The middle figure in the bottom illustration on page 23 shows how it is done with 10. Note that obtuse triangle *ABC* is cut into seven acute triangles by a pentagon of five acute triangles. If *ABC* is now divided into an acute and an obtuse triangle by *BD*, as is indicated in the bottom figure, we can use the same pentagonal method for cutting the obtuse triangle *BCD* into seven acute triangles, thereby producing 11 acute triangles for the entire square. A repetition of the procedure will produce 12, 13, 14, ... acute triangles.

Apparently the hardest dissection to find is the one with nine acute triangles. Nevertheless, it can be done, as I shall disclose in my next column.

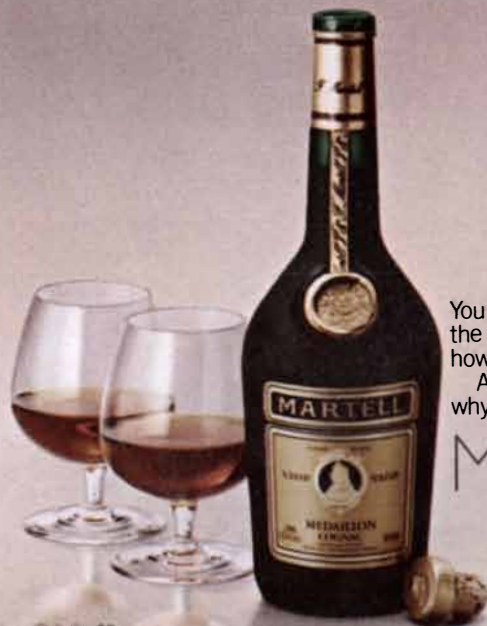
There are many comparable problems about cutting figures into nonoverlapping triangles, of which I shall mention



University Avenue, where Lavinia must find a room



THE MOMENT WAS MARTELL.



You remember how very special
the moment was, how very relaxed,
how totally perfect.
And, of course, you remember
why. Martell Cognac.

MARTELL
COGNAC
SINCE 1715

V.S.P., V.S.O.P. and Cordon Bleu.
Imported by the Jos. Garneau Co., New York, N.Y. © 1980. 80 Proof

© 1981 SCIENTIFIC AMERICAN, INC

New drives



(16) EPA EST. MPG. 24 HWY. EST. Compare to estimated MPG of other cars. Your mileage

Introducing the next logical Lincoln.

New dimensions in size and technology were achieved in 1980. Now, these achievements continue in the new 1981 Lincoln Town Car.

This is a truly remarkable automobile.

Among its technological features:

- Electronic fuel injection and electronic engine controls for

fuel efficiency and performance.

- Four-speed Automatic Overdrive Transmission for excellent highway mileage.

- Trim, contemporary size with room for six adults.

- Optional electronic instrument panel with message center that answers twenty different driver inquiries about time/distance/fuel usage.

- Optional keyless entry system that unlocks doors and/or decklid electronically simply by depressing buttons in numerical sequence.

Lincoln Town Car for 1981. A fine sense of tradition...remarkable achievements in technology.

You are cordially invited to experience it at your Lincoln-Mercury Dealer. There, you may exercise your option to lease or buy.

nsions

1981 LINCOLN TOWN CAR

LINCOLN-MERCURY DIVISION



may differ depending on speed, weather and trip length. Actual highway mileage may be less.

Once every generation it seems, something comes along that changes the entire nature of home entertainment. That time has come again.

Pioneer introduces LaserDisc™ a whole new technology that makes your television and stereo do things they've never done before. Now you can watch a movie or hear and see a concert as easily as playing a record.

And when you hook up the LaserDisc player to your TV and hi-fi, for the first time, along with a television picture you get true stereo.

And a television picture that's actually a 40% sharper picture than home videotape.

And because the disc is read by a light beam rather than a needle or video head, nothing touches the

disc. So, with normal use, it doesn't wear out from play. You can enjoy the disc forever.

But Pioneer LaserDisc offers you more than superb fidelity, it's truly the first form of personal entertainment. Your television viewing is no longer limited to what's on television. Because you can see what you want to see when you want to see it. A whole library of Pioneer Artists and MCA/Discovision discs with movies like *Jaws*, *Animal House*, *The Blues Brothers*; concerts like Paul Simon, Liza Minelli, Loretta Lynn, and Abba; sports like NFL football, and title fights.

You can study the standard-play discs in fast motion, slow motion, stop motion. Even create your own instant replay.

But there's something even more remarkable. A micro-computer built

into the Pioneer LaserDisc player lets you access any individual frame on the disc at the touch of a button. You can go right to your favorite scene in a movie, or a song in concert in seconds.

And because you can view the up to 108,000 frames on the disc one frame at a time, it just may be the future of education.

The cost of Pioneer LaserDisc is surprising as well. Just \$749* (with \$50* more for optional remote control). And a full-length movie on LaserDisc can cost less than taking your family out to the movies.

The only way to even begin to experience the magic of this remarkable LaserDisc system is to see it. And we've arranged for that. Just call at 800-621-5199** for the store nearest you. ** (In Illinois 800-972-5822).

*Suggested retail price. Actual price set by dealers.

Finally, high-fidelity television with stereo sound.

Simulated TV picture from Paul Simon.



© 1981 U.S. Pioneer Electronics Corp.

LaserDisc
PIONEER
We bring it back alive.

© 1981 SCIENTIFIC AMERICAN, INC

only two. It is easy to divide a square into any even number of triangles of equal area, but can a square be cut into an odd number of such triangles? The surprising answer is no. As far as I know, this was first proved by Paul Monsky in *American Mathematical Monthly* (Vol. 77, No. 2, pages 161-164; February, 1970).

Another curious theorem is that any triangle can be cut into n isosceles triangles provided n is greater than 3. A proof by Gali Salvatore appeared in *Crux Mathematicorum* (Vol. 3, No. 5, pages 134-135; May, 1977). The case of the equilateral triangle is of particular interest. It is easy to cut it into four isosceles triangles (all equilateral) or into three isosceles triangles. (Some triangles cannot be cut into three or two isosceles triangles, which is why the theorem requires that n be 4 or more.) Can you cut an equilateral triangle into five isosceles triangles? In my next column I shall show how it can be done with none of the five triangles equilateral, with just one equilateral and with just two equilateral. It is not possible for more than two of the isosceles triangles to be equilateral.

5. Measuring with Yen. This problem was originated by Mitsunobu Matsuyama, a reader in Tokyo. He sent me a supply of Japanese one-yen coins and told me of the following remarkable facts about them that are not well known even in Japan. The one-yen coin is made of pure aluminum, has a radius of exactly one centimeter and weighs just one gram. Thus a supply of yens can be used with a balance scale for determining the weight of small objects in grams. It also can be used on a plane surface for measuring distances in centimeters.

It is easy to see how one-yen coins can be placed on a line to measure distances in even centimeters (two centimeters, four, six and so on), but can they also be used to measure odd distances (one, three, five and so on)? Show how a supply of one-yen coins can be used for measuring all integral distances in centimeters along a line.

6. A New Map-coloring Game. This problem comes to me from its originator, Steven J. Brams, a political scientist at New York University. He is the author of *Game Theory and Politics* (1975) and *Paradoxes in Politics* (1976), both published by the Free Press, and *The Presidential Election Game* (Yale University Press, 1978). His latest book, *Biblical Games* (The MIT Press, 1980), is a surprising application of game theory to Old Testament episodes of a game-like nature in which one of the players is assumed to be an omniscient deity.

Suppose we have a finite, connected map on a plane and a supply of n crayons of different colors. The first player, the minimizer, selects any crayon and colors any region of the map. The sec-

ond player, the maximizer, then colors any other region, using any of the n colors. Players continue in this way, alternately coloring a region with any of the n colors but always obeying the rule that no two regions of the same color can share any portion of a common border. Like colors may, of course, touch at points.

The minimizer tries to obviate the need for an $n + 1$ color to complete the map. The maximizer tries to force the use of it. The maximizer wins if either player is unable to play, using one of the n crayons, before the map is fully colored. If the entire map is colored with n colors, the minimizer wins.

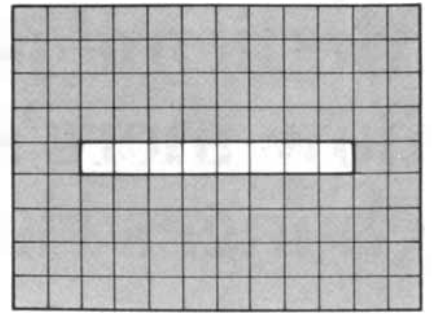
The subtle and difficult problem is: What is the smallest value of n such that when the game is played on any map, the minimizer can, if both players play optimally, always win?

To make the problem clearer, consider the simple map in the top illustration on page 26. It proves that n is at least 5. Of course, if no game is played, the map can be easily colored with four colors, as indeed any map on the plane can be. (This is the famous four-color theorem, now known to be true.) But if the map is used for playing Brams's game, and only four colors are available, the maximizer can always force the minimizer to use a fifth color. If five colors are available, the minimizer can always win.

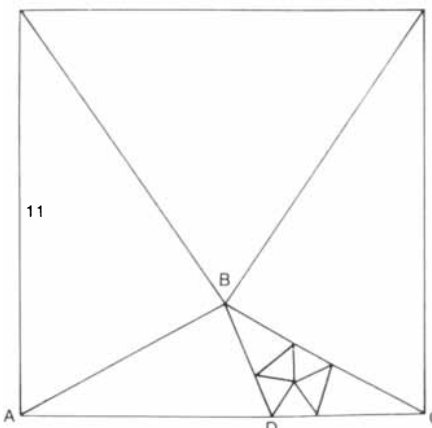
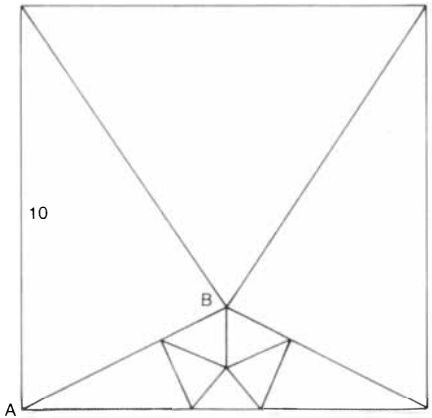
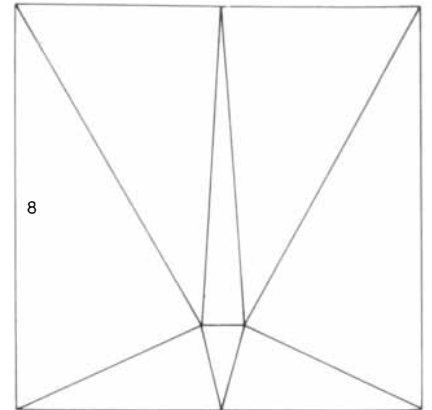
Brams conjectures that the minimum value of n is 6. A map has been found on which five colors allow the maximizer always to win. Can you construct such a map and give the maximizer's winning strategy? Remember, the minimizer goes first, and neither player is under any compulsion to introduce a new color at any turn if he can legally play a color that has already been used.

7. Whim. In his Pulitzer-prize book *Gödel, Escher, Bach* (now a Vintage paperback) Douglas R. Hofstadter introduces the concept of self-modifying games. These are games in which on his turn a player is allowed, instead of making a legal move, to announce a new rule that modifies the game. The new rule is called a metarule. A rule that modifies a metarule is a metametarule, and so on. Hofstadter gives some chess examples. Instead of moving, a player might announce that henceforth a certain square may never be occupied, or that all knights must move in a slightly different manner, or any other metarule that is on a list of allowed alterations of the game.

The basic idea is not entirely new; before 1970 John Horton Conway proposed a whimsical self-modifying variation of nim that he called whim. Nim is a two-person game played with counters that are arranged in an arbitrary number of piles, with an arbitrary number of counters in each pile. Players take turns removing one or more counters from any one pile. In normal nim the person taking the last counter wins; in *misère*,



A damaged patchwork quilt



Cutting a square into acute triangles

Medicine is always changing, now there's a text that changes right along with it.

SCIENTIFIC AMERICAN *Medicine*.

The answers that you get from SCIENTIFIC AMERICAN *Medicine* are current, because each month new chapters replace the old. You will know where to look for the answers, because this advanced text follows the familiar structure of medical knowledge, changing as medicine changes.

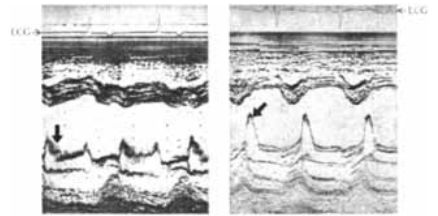
Comprehensive. Leading authorities from Harvard and Stanford cover the full range of medicine for you the seasoned clinician.

Complimentary CME. At no extra cost, we offer a CME program through which you can earn 32 top credits* per year.

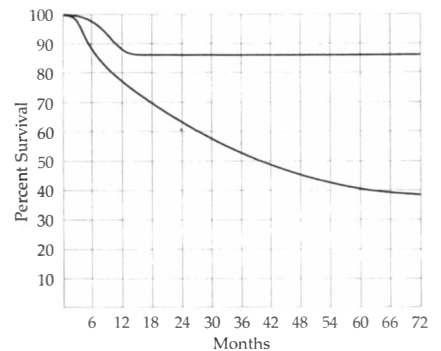
Trial Offer. We invite you to try SCIENTIFIC AMERICAN *Medicine* for two months with our compliments. Send us the coupon and you will receive the

two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, you will owe us \$185 for the first 12 months (renewal is \$160); otherwise, return the volumes; there will be no charge.

Please mail the coupon today so that we can put medical progress to work for you.

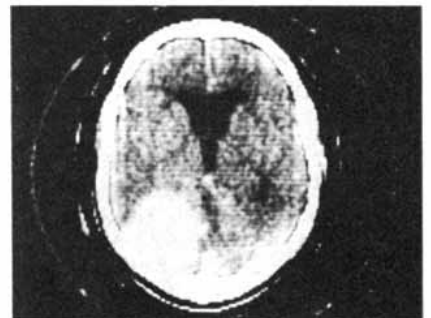


Echocardiograms from patients with aortic regurgitation.



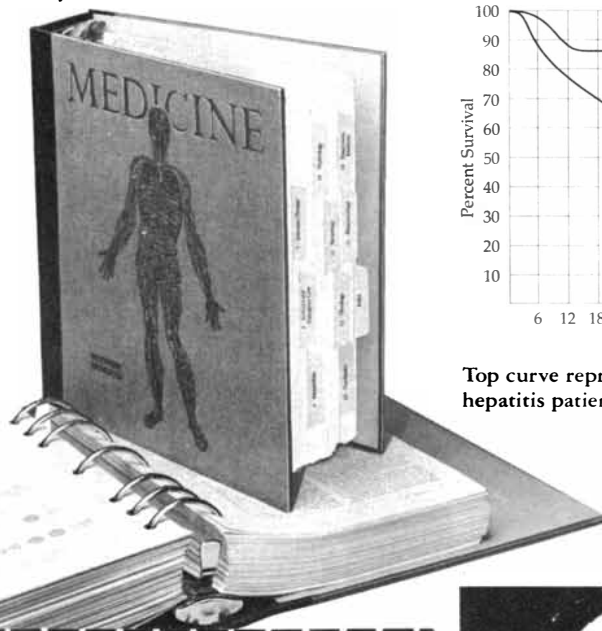
Top curve represents survival in chronic hepatitis patients treated with steroids.

CT scan reveals large meningioma and mild hydrocephalus.



*As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions.

This program is approved by the American Academy of Family Physicians for 32 elective hours of CME credit.



SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, New York 10017

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. I understand that the annual subscription of \$185 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.

- 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 \$185*
- Please bill me

*Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio, and New York

Name _____ MD Specialty _____
 Address _____
 City _____ State _____ Zip _____
 Signature _____

Subscribers outside the U.S. and possessions will be charged extra for shipping and handling; updates will be sent by surface routes unless airmail delivery is requested. Please allow 6-8 weeks for delivery.

1W

Who will be first with the electronic eyes of the free world?

It could be you and Hughes Electro-Optical and Data Systems.

We built the first working laser and the first ground laser sensor ever used by the Army. Today more than 30 free-world nations use our sensing devices, and our systems are applied in defense applications and in space.

The Space Shuttle will launch Hughes sensors that can gaze at the sun or probe the depths of space, with the "intelligence" to detect, track, identify and classify almost anything in sight.

In fact, today's Hughes is one of the nation's largest employers in virtually every engineering, scientific, computer

and technical discipline — with 1,500 projects and a \$5 billion backlog. Yet we're decentralized to give you the kind of environments that stimulate innovation and individual recognition.

And our new facility in El Segundo will be a 140-acre city of research and development, manufacturing and testing excellence, to help thrust your vision into the future.

Who will be first with the optical and data systems of the future? It could be you and Hughes.

At Hughes Electro-Optical and Data Systems, we'll introduce you to people, ideas and jobs that could change your world. And maybe ours.

Call or send resume to:

Hughes Electro-Optical
& Data Systems
P.O. Box 913, Dept. AS-4
El Segundo, CA 90245
(213) 616-3374

Current openings:

Electro-Optical/Optical Design
Infrared Systems
Materials & Processes
Metallurgical
Computer Science/Software
CAD/CAM/CAT
Analog/Digital Design
Components & Materials
Control Systems Design/Analysis
Cryogenic
Electromechanical Packaging
Servo
Space Sensors
Facilities
Test Equipment Design

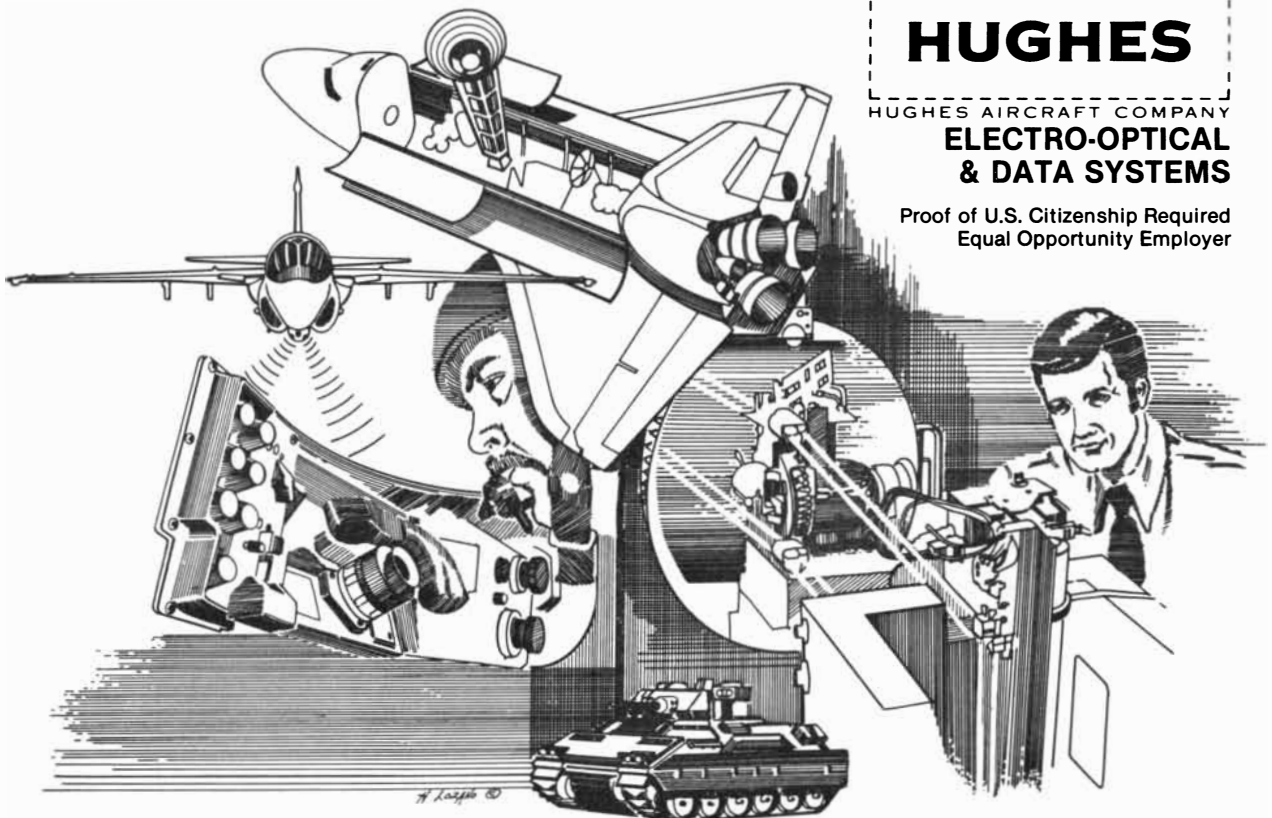
It could be you and Hughes **Electro-Optical and Data Systems.**

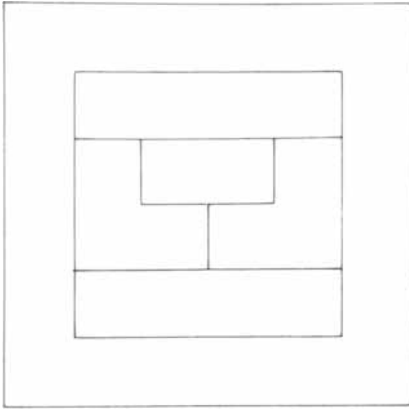
HUGHES

HUGHES AIRCRAFT COMPANY

ELECTRO-OPTICAL & DATA SYSTEMS

Proof of U.S. Citizenship Required
Equal Opportunity Employer





Brams's map-coloring game

or reverse, nim the person taking the last counter loses. The strategy for perfect play has long been known. You will find it in the chapter on nim in my *Scientific American Book of Mathematical Puzzles & Diversions* (Simon and Schuster, 1959).

Whim begins without any decision on whether the game is normal or misère. At any time in the game, however, either player may, instead of making a move, announce whether the game is normal or misère. This "whim move" is made only once; from then on the game's form is frozen. It is well known that in nim the strategy is the same for both forms of the game until near the end, so that you may be tempted to suppose whim strategy is easily analyzed. Try playing a few games and you will find it is not as simple as it seems!

Suppose you are the first to play in a game with many piles and many counters in each pile, and the position is a nim loss for you. You should at once make the whim move because it leaves the position unchanged and you become the winning player. Suppose, however, you are first to play and the position is a win for you. You dare not make a winning move because this allows your opponent to invoke the whim and leave you with a losing position. Hence you must make a move that would lose in ordi-

nary nim. For the same reason your opponent must follow with a losing move. Of course, if one player fails to make a losing move, the other wins by invoking the whim.

As the game nears the end, reaching the point where the winning strategy diverges for normal or for reverse nim, it may be necessary to invoke the whim in order to win. How is this determined? And how can one decide at the start of a game who has the win when both sides play as well as possible? Conway's strategy is easy to remember but, as he once remarked, hard to guess even by someone well versed in nim theory.

Kenneth W. Abbott, a New York computer consultant, sent an amusing generalization of the discrete "taxicab" geometry discussed in last November's column. As in taxicab geometry, the points of the non-Euclidean space are the intersections of lattice lines on a square grid. In Abbott's generalization the "distance" between any two points is an integer defined as being equal to $n\sqrt{x^n} + y^n$, where x is measured horizontally, y is measured vertically and n is any positive integer.

When n is 1, we have the simple taxicab geometry explained in November. All "circles" in this geometry are sets of points equidistant from the circle's center. They have the forms shown at the left in the illustration below, where the radii are 1, 2, 3, 4 and 5.

When n is 2, circles of the same radii take the forms shown in the middle of the illustration. Note that the first four circles consist of just four points lying on the two axes that pass through the common center of the five circles. We shall call such circles "trivial." When n is 1, only the circle of radius 1 is trivial. All other circles are not. When n is 2, the fifth circle is nontrivial. In this geometry there are an infinite number of both kinds of circles. Pi is $2\sqrt{2}$ for all trivial circles, but it has different values for the nontrivial ones. For the fifth circle, which has a radius of 5, pi is $(4\sqrt{10} + 2\sqrt{2})/5$.

When n is 3, the first five circles (at right in illustration below) are all trivial. In this geometry pi is $2^{(n+1)/n}$ for all trivial circles.

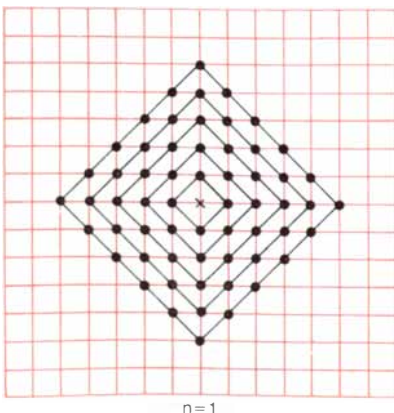
We now state a remarkable conjecture. Any generalized taxicab geometry with n greater than 2 can contain only trivial circles. This unsolved conjecture, as Abbott has pointed out, is easily seen to be equivalent to Fermat's last theorem!

My December column on prime numbers mentioned a sequence of eight primes having the form $k, k + 2, k + 6, k + 8, k + 12, k + 18, k + 20, k + 26$, and it stated that the only known instance is 11, 13, 17, 19, 23, 29, 31, 37. John C. Hallyburton, Jr., who works for the Digital Equipment Corporation, found seven other such sequences. The starting numbers of each are

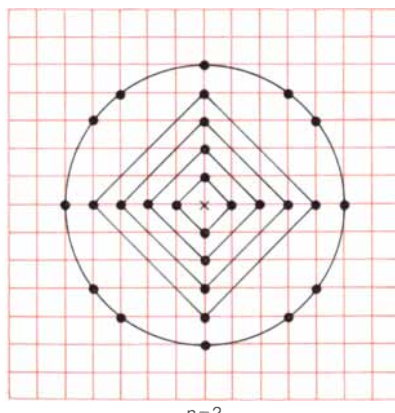
- 15,760,091
- 25,658,441
- 93,625,991
- 182,403,491
- 226,449,521
- 661,972,301
- 910,935,911

Mark Templer reported an error in my listing of four primes from his 1980 paper; the first prime should be 31, not 37. Many readers called attention to another mistake. The patterned sequences of primes, 7, 37, 337, 3337, ..., should have been 31, 331, 3331, The pattern first fails for 333333331, the product of 17 and 19,607,843.

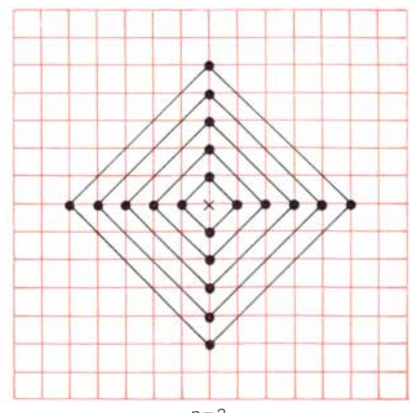
Last September Dr. Matrix mentioned the unsolved problem of determining the last prime on an alphabetical listing of the English names of primes. Donald E. Knuth, the noted computer scientist at Stanford University, gave the task to his students in a programming seminar. More than half a dozen students independently found the answer: two vigintillion two undecillion two trillion two thousand two hundred ninety-three.



n=1



n=2



n=3

"Circles" of radii 1, 2, 3, 4 and 5 in a generalized "taxicab" geometry



“The Best In The House” for over a century.

Hiram Walker's unique, light-tasting whisky earned its reputation as the finest whisky in exclusive clubs from the time it first appeared. That's how it became “Club Whisky.”

In 1891, a new U.S. law required the country of origin to appear prominently on a product's label. Hiram Walker proudly added “*Canadian*” to Club Whisky and people just as proudly began ordering “*Canadian Club*.”

Today people enjoy Canadian Club so many ways: on the rocks, smoothing out sours or Manhattans and with favorite mixers. Because it's lighter than Scotch, smoother than bourbon. And it's still “The Best In The House”[®] in 87 lands.

Canadian Club

BOOKS

How engineers overcome stress and gravity, and everything about music and mathematics

by Philip Morrison

STRUCTURES, OR WHY THINGS DON'T FALL DOWN, by J. E. Gordon. Plenum Press (\$17.95). WHY BUILDINGS STAND UP: THE STRENGTH OF ARCHITECTURE, by Mario Salvadori. W. W. Norton & Company, Inc. (\$14.95). Fry a sausage and nearly always the steamy, swelling interior cracks open the skin along the length, almost never across it. That result follows directly from one of the simplest practical formulas of the mechanics of stress in a material. The cheap but heavy gas cylinder in the garage and the elegantly optimized spherical oxygen bottles of spacecraft bear out the same mechanical truism. The taut rigging and flattened Dacron sails of the tensely constructed racing yacht contrast acutely with the partitioned sails of a Chinese junk, which bulge between supports to adjust pressure curvature, the stress on the sail remaining roughly constant as the wind freshens. Bats are rigged on Chinese-junk lines, and they pay a small price in aerodynamic costs; so probably did the big pelagic pterodactyls.

Still more striking energy containment lies deep in living forms. A constant-stress biological membrane, with its stress-strain curve running horizontally across the graph, must resemble a soap film. Such membranes can support stress only if they are formed as spheres or part spheres: bubble trains, like a segmented worm. But a tube or a pipe, the cylinders that transport essential stuffs throughout the kingdom of larger life forms, like our own vital arterial plumbing, cannot function with constant stress. "The circumferential stress is ineluctably twice the axial stress," and so the stress-strain curve of the walls of tubing must slope upward to accommodate the doubled stress. Rubbery elastomers indeed can work, but their stress-strain curves are S-shaped; they are therefore unstable at high strains, blowing out into local bubbles like a snake that has eaten a football. A kitchen observation has led to the logic of the deadly aneurysms.

A fuller analysis demonstrates that a material that is stable under fluid pressure even at high strains must show a stress-strain curve that is rather flat at first but then rises steeply. Very often

found in living organisms is a composite tissue whose soft matrix (the protein elastin), almost as yielding as a surface-tension film, is reinforced by a felted mat of the stiffer protein collagen. The collagen fibers come taut only after much extension, to provide the increased resistance needed at high strain. It is the tough collagen in muscle that cooking breaks down to tenderer gelatin at a temperature below that which destroys elastin (enough to restore "one's faith in the beneficence of Providence").

Leave soft materials; consider stone. Men have built well in heavy brick and stone from earliest times. The crushing of well-baked bricks by the sheer dead weight of the walls towering above them sets a very grand limit indeed: the Tower of Babel could have been raised two kilometers high with parallel walls, even higher if it had tapered like a mountain. (The tallest actual masonry structures indeed rise half as high as modern steel towers.) The real limit is evident; we all learned it playing with blocks. Block walls do not crush but fall down from lack of stability. They fall sideways. Masonry cannot withstand any tension; tensioned joints crack and gape. Then, if the wall thickness still bearing weight dwindles any more, the uptilted block will eventually rock around an outside edge; the wall tips upward and out and collapses utterly. Masonry design is a matter of arranging things so that the lines of thrust from the weight above remain well within the lower walls. Mere weight is not the issue; cathedral walls actually gained stability from the massive pinnacles and sculptured saints added to their top. Fortunately structures that fail not by breakage but from such geometrical instability can be scaled up proportionately both from models and from previous experience.

The cathedral builders had plenty of experience. The audacity and beauty of those transparent, pierced, lofty walls, their necessary supports artfully flying outside the structure, is superb. The style ended at Beauvais; there the stone tower reached an incredible 500 feet above the still incomplete nave. It failed in 1573 and was never rebuilt. The Gothic impetus was spent; the schools for the skilled tradesmen had died out,

with the lore perhaps not fully mastered. After all, 350 years had passed since the beginning of construction at Beauvais. The epoch had changed; as Beauvais was dying, the masters of the Renaissance were proudly building in a new way. The death of St. Peter's at Beauvais came as St. Peter's in Rome grew in heavy-domed triumph.

Today we still gather ceremonially, although often for a secular sacrament around the playing field, under marvelous composite inverted domes. The current Madison Square Garden (there have been three) is one of those, a dish roof made of a thin concrete slab held on tensed cables. The web of radial cables (modern steel cables easily support 10 times the stress found safe for the wrought-iron chains of 19th-century bridges, some of which are still in faithful service) acts to compress their concrete ring support on the outer building wall and also to extend the inner steel anchor ring at the center of the roof. The slabs for the tight roof can be thin (usually only a couple of inches) because the concrete merely spans the small distance between the cables. The entire structure can be erected without any expensive scaffolding.

The dish roof was made practical within the past 25 years by the pioneer design of the Uruguayan engineer Leonel Viera. He had the happy idea of loading the roof with brick ballast before joining the slabs with a good mortar. Once the slab had set, the ballast was removed, having tensed the cables within the solid concrete. Thus stiffened, the roof no longer responded much to wind gusts—as had earlier, lightly loaded roofs held by flexible cables. The prestressed, stiff, cable-held light concrete dish (there are now other ways to achieve it) can span 400 or even 500 feet easily. We can expect better materials still, and can "dream of dish roofs... under which hundreds of thousands of people gather for the enhancement of their deepest experiences."

The paragraphs above conflate two rich and readable books. J. E. Gordon, the author of *Structures*, is professor of materials science at the University of Reading, with a background in Glasgow naval architecture, in yachting and in aircraft design. His book is much the more wide-ranging. Its tables of strengths and moduli include not only brick and steel and wood but also a few dozen other substances, from eggshell membrane to diamond. Gordon presents his topic in four parts. He begins with a careful account of the meaning and history of the ideas of stress and strain, goes on to cover modern fracture mechanics (that is, the energy considerations of crack propagation) and then structures of all kinds in tension and in compression, and finally pays due attention to more complicated stress patterns such as bending (and, unexpectedly,

E L D O R A D O

More going for it than the luxury imports...

...more advanced technology like V8-6-4.

No luxury import offers an engine that goes from 8 to 6 to 4 cylinders, but Eldorado does and it's standard. Also standard is another Cadillac exclusive, Digital Fuel Injection—a generation ahead of conventional electronic fuel injection.

...more meaningful features like front-wheel drive.

Surprisingly, most luxury imports lack front-wheel drive while Eldorado has had it since 1967. Other Eldorado standard features often lacking on luxury imports include electronic level control, cornering lights and MPG Sentinel—a miles-per-gallon monitor which can help you become a more efficient driver.

...more ways to go.

No single luxury import model offers you Eldorado's unprecedented engine choice... standard V8-6-4... available V6... or available Diesel.

Eldorados are equipped with GM-built engines produced by various divisions. See your Cadillac dealer for details.

...more interior roominess.

In addition to impressive traction, front-wheel drive helps give Eldorado more interior roominess than most luxury import models. Source: EPA Interior Volume Index.

...more convenient service.

Eldorado has a larger dealer network than any luxury import... for readily available service and parts across America.

...more in demand.

Eldorado outsells any single luxury import model. Any single luxury import model. Source: R. L. Polk Registrations, Jan.-Oct., 1980.

...and more... distinctive styling.

...more going for it... and you.

When you compare everything you get with an Eldorado that you don't get with a luxury import, we think you'll agree that there is no comparison. Eldorado has more going for it... and you. Whether you buy or lease, see a Cadillac dealer, and see for yourself.



CADILLAC MOTOR CAR DIVISION, U.S.A.



The Polo: a water-resistant quartz Piaget. 136 grams of 18K gold carved link-by-link into the most dynamic sports watch ever made.

Designed to accommodate the most diverse lifestyle.

Shock-resistant. Water-resistant to three atmospheres. You can go through the day and night never having to take it off your wrist.

Styling and ruggedness only Piaget could combine.

Rugged yet distinctly elegant is the uniqueness of the Polo. On your wrist, it is one continuous flow—bracelet to case to bracelet.

18 karat gold bracelet down to the smallest screw.

Each link, carved from a solid



Each link of the bracelet is handcarved from a solid block of 18K gold. And individually attached by a tiny gold screw. The result is a new dimension in flexibility.

block of 18K gold, is individually attached by tiny gold screws. Piaget has achieved a new dimension in beauty and flexibility.

Impeccable accuracy is inherent in a Piaget.

The Polo's electronic quartz movement is completely hand-crafted in Switzerland, accurate to within 30 seconds a year. And never needs winding. You simply replace the battery every two years.

The Polo. For the man who deserves the most expensive watch in the world: a Piaget.



For more information write Piaget, Dept. SA, 650 Fifth Avenue, New York, New York 10019.

bias-cut dressmaking, à la Mlle Vionnet of Paris in the 1920's, complete with period fashion illustrations). His style is personal, witty and ironic, with a good deal of wordplay and much attention to apt epigraphs from Kipling.

Mario Salvadori is a widely known architectural engineer at Columbia University; his writing is lucid, hopeful and direct. His gaze is tightly fixed on buildings. *Why Buildings Stand Up* discusses the principles of stress analysis without formulas, although at a less physical level than Gordon's book. It includes chapter after chapter sketching the design and history of famous structures from the Pyramids to the Brooklyn Bridge, the Eiffel Tower and the wonderful works of Pier Luigi Nervi in reinforced concrete (even a beautiful sailing ketch). Salvadori's title is accurate, but Gordon treats many topics entirely apart from gravity.

Both men offer strongly stated views on the relations between aesthetics and technology. Salvadori argues for the naturalness of design that reveals structure, although with due concern for novel experience. Gordon disdains austerity and hopes for lots of ornament: "Let there be figureheads... crinolines... flags." Both books are lasting additions to the slim stock of readable works on structure and form; the Gordon book is a find for the scientific reader at large, the Salvadori for the reader who is caught up in the roofs and bridges of all the world.

THE NEW GROVE DICTIONARY OF MUSIC AND MUSICIANS, edited by Stanley Sadie. Macmillan Publishers Limited (\$1,900). ENCYCLOPEDIA OF MATHEMATICS, by the Mathematical Society of Japan, edited by Shōkichi Iyanaga and Yukiyoji Kawada. Translation reviewed by Kenneth O. May. The MIT Press (\$40). These two works of reference share much. Each of them offers comprehensive matter (with an order of magnitude or so between them in cost and bulk), here and there fully to be grasped by any general reader but most frequently so embedded in its discipline as to require specialized preparation for understanding. Again in both cases the biographies of the men and women who built the two structures are evidently useful references for anyone interested.

The Grove music dictionary is presented in 20 thick volumes, with a resounding breadth of coverage. Only a small sample (Volume I, *A* through *Bacilly*) was examined for this review. (M. de Bacilly, by the way, was a composer and singing teacher in 17th-century Paris.) Nearly 900 pages, the volume is well illustrated with music scores, photographs and drawings. The claim that this new work in an old and respected tradition represents the active scholarship of music in all its aspects is not overdrawn.

a small hotel




PARK HYATT

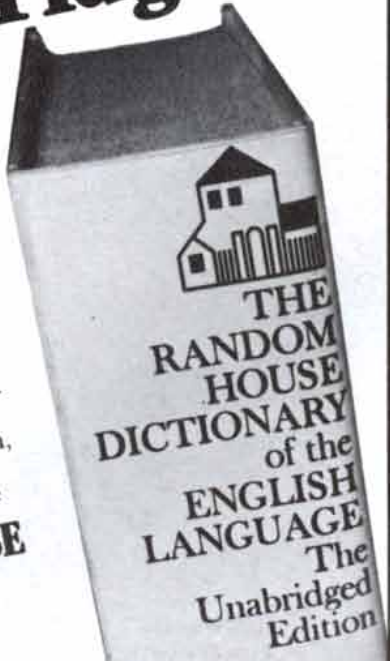
ON WATER TOWER SQUARE, CHICAGO

312 280-2222

THE ANSWER BOOK The Unabridged!

- THE MOST AUTHORITATIVE DICTIONARY OF ITS KIND
- Comprehensive—more than 260,000 entries; 2,091 pages; large format 9" x 12" page size; 9 lbs 14 oz.
- Up-to-Date—with new words and terms
- Easy to Use—more than 50,000 example phrases and sentences; 2,000 illustrations; 10,000 synonym lists and studies; thumb-indexed
- Full-color ATLAS; much, much more

\$49.95, now at your bookstore



RANDOM HOUSE
THE
RANDOM
HOUSE
DICTIONARY
of the
ENGLISH
LANGUAGE
The
Unabridged
Edition

(Most of the 2,500 contributors overall are American; U.S. musicology has plainly ripened.) Its catholicity is noteworthy: in the first volume an expert overview of African music, north and south of the Sahara, serves to introduce a large number of other articles, country by country; Angola, for instance, gets four pages illustrated with an old map, a period drawing, photographs and a transcribed part-song. The editors hold that no other work has ever so extensively treated the music of non-Western and folk tradition. The evidence is here.

A few articles catch the eye of a scientific peruser. Acoustics? A long article opens with a remarkable survey that begins with architectural acoustics, from antiquity. The geometrical Greek theater was designed with voice projection a major aim; the Romans disposed resonant bronze vases around the seats to reinforce the song. The account goes on over time all the way to Tanglewood and the Royal Albert Hall. The rest of the article treats somewhat arbitrarily the sound-source physics of strings, woodwinds, keyboard strings and the voice, in much detail (including the Tibetan drone chant) but with little mathematics. (To be sure, major articles on hearing, sound and the physics of music appear in other volumes.) The articles clustered around the name Bach occupy

some 100 pages. A naive reader was astonished to learn that the mighty river of music that flowed through Johann Sebastian (1,087 listed works) and on to his musical kin for two more generations had risen in J. S. Bach's great-great-grandfather's day. That family "produced musicians of every kind in number beyond parallel: from fiddlers and town musicians to organists, Kantors, court musicians and Kapellmeisters."

This admirable reference work, a library's pricey treasure, is being published in the U.S. under rather special circumstances, with no little public-relations fanfare. It is a pleasure to report that the quality of the work fulfills the promises, as far as one volume of 20 can be held fair sample for a nonexpert reviewer.

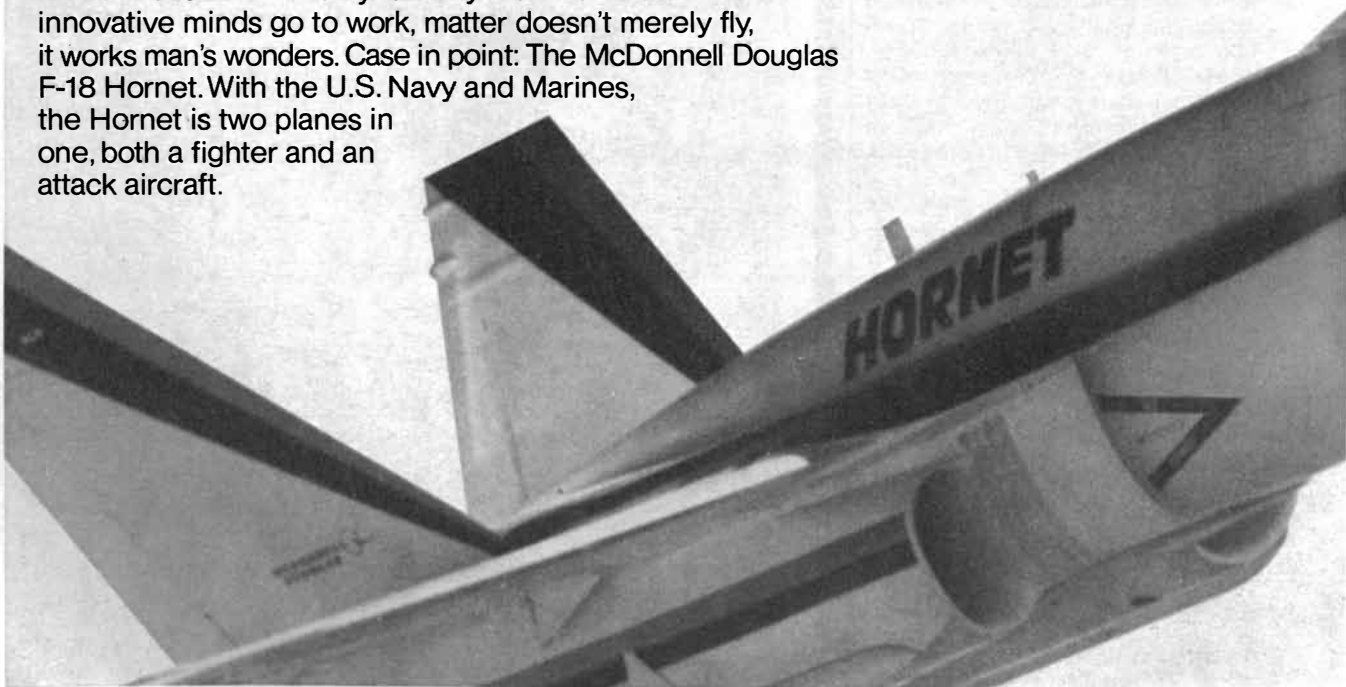
The second of these reference works looks at an even purer order long sought by the human mind. It arose as a work of energy and devotion by the Mathematical Society of Japan, to which about 200 experts of the group contributed. The edition at hand, translated by the same society, has been brought up to date as of about 1976. It has now reached this relatively inexpensive and very beautiful realization, the crisp, tall pages of the two paperbound volumes having the air of an old Japanese book. (There was an earlier hardcover edition at \$150.) The

text presents 436 articles, from Abel to zeta functions, spanning with authority the whole of modern mathematics. Each article is of modest length and includes both concise definitions and some sense of the relations among concepts, but it will not serve an outsider, say a journalist who wants a simple account of the nature of some field of mathematics. Look up homotopy, for example, and learn that if a family of continuous mappings from one topological space into another is continuous also with respect to a certain index, it is called a homotopy. (Here the symbols have been paraphrased.) With cross-references explicit, such matter can inform serious students at the research or advanced-undergraduate level, but hardly tyros. (The entry under number theory, elementary, is not so daunting.)

Some history and biography are included, and applications (for example quantum mechanics) are here along with pure mathematics, although they are seen from a strictly mathematical point of view. There is a persuasive tone of understanding. This is a genuine work of union, although perhaps not one of unification. The indexes are rich, and the 100-odd pages of formulas and constants are surely the most catholic to be found, from the quadratic equation through special functions, differential

MINDS OVER MATTER: One plane for two missions.

Matter doesn't want to fly. Gravity won't let it. But when innovative minds go to work, matter doesn't merely fly, it works man's wonders. Case in point: The McDonnell Douglas F-18 Hornet. With the U.S. Navy and Marines, the Hornet is two planes in one, both a fighter and an attack aircraft.



equations and differential geometry, on to such arcana as the diagrams of knot theory and the homotopy groups of spheres. These days even the last hints at application.

A CONCISE WORLD ATLAS OF GEOLOGY AND MINERAL DEPOSITS, by Duncan R. Derry. Mining Journal Books Ltd. Distributed by Halsted Press, a division of John Wiley & Sons, Inc. (\$61.95). An expert compilation by a distinguished Canadian exploration and economic geologist, this attractive book offers an up-to-date geological map of the world in 10 sheets, intelligently adjusted for scale and projection to suit the regions shown. The maps are full of information legibly displayed, although (because of the necessities of scale) never very detailed. For example, the two Americas lie as a whole on one two-page spread, sensibly from left to right, about two feet across from the North Pole to Cape Horn.

This is no landscape atlas; relief and soil cover are almost unseen. Its cut is to the rocky bone. The Americas show five ancient continental nuclei, deep in the Precambrian, heavily imprinted with symbols marking the main deposits of gold, copper-zinc, iron and nickel. The Precambrian is divided. The earliest bedrock, more than 2,600 million years

old and stretching wide from Lake Superior to the Greenland coast, is the true ancient platform; it is marked by intense volcanism and by sedimentary rocks that are reworked volcanic outflows rather than the long-weathered layered structures of the familiar sandstones, shales and limestones that dominate the later sediments. Most iron ore worked today is the legacy of Precambrian microbes. Along with the map and its many colored legends there goes a detailed text of some 10 pages, chronologically discussing the chief events of hard-rock geology in this region and their eventual economic consequences.

So the volume continues, with one or two sheets each for the other continents. More room is given to northern and Mediterranean Europe; the U.S.S.R., Africa, southern Asia and the rest appear at intermediate scale; the two polar regions share with the Americas the largest compression of scale. The mineral deposits noted include the chief metals, some 17 elements from antimony to zinc; such nonmetals as diamond, asbestos, potash and phosphate, and the energy minerals, including uranium. Volcanics are marked and some famous individual mines as well, such as Homestake, Tsumeb and Broken Hill. The area from Kimberley across the Witwatersrand earns a special insert map, but

the Pacific islands, from beyond Samoa east to the coast of the Americas, are not shown at all; Hawaii is absent. The sea bottom is well treated, mapped with broad depth contours; we see the great ridges and major faults and the deep-sea nodule deposits.

The introduction by a senior British geologist, Sir Kingsley Dunham, makes clear the novelty of the entire plan. This is the first geological atlas squarely based on the "two great advances in thought about world geology" of the past generation. First of all, of course, there is the marvelous unifying stimulus of plate tectonics, whose concepts dominate every page and almost every line. Second, on the practical side, the great growth in the discovery and consumption of world petroleum resources has both enriched the information brought together here and provided the point of view (toward the finiteness of resources) that underlies the economic discussion. Some 10 pages summarize world production by mineral and country, mainly as of 1979, and also present guarded estimates of known reserves.

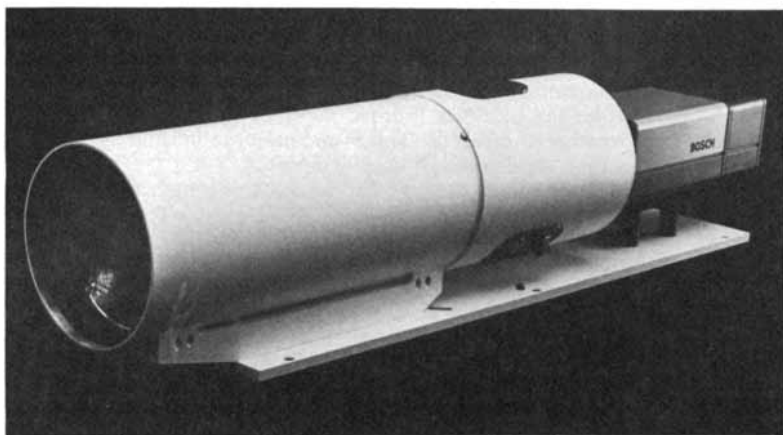
The author-editor set himself three objectives. First, the book should be written so that it is "comprehensible to those who have a minimal knowledge of geological science." And so the maps are marked with the age of formations

Working magic with computers.
Design an oil rig to stand up to hurricanes. Schedule the working hours for thousands of telephone operators. McDonnell Douglas Automation Company solves tough problems with innovative, hard-working computer programs for clients in dozens of industries.

A plane that flies in 69 languages. The challenge: Move more than 300 people, on short trips or long, across land and sea at 600 miles an hour—in all the comfort of home. The solution: The DC-10. It now serves 90 nations, flying a *million miles a day*.

MCDONNELL DOUGLAS

For more information, write: McDonnell Douglas, Box 14526, St. Louis, MO 63178



The Questar 40-120

A unique instrument for which a patent has now been granted, the Questar 40-120 establishes prime focus at both 40 and 120 inches (1000 and 3000 mm.). Other lenses—particularly the zooms—can vary their focal lengths but with a compromise of quality and with relatively low magnification. The 40-120 resolves 100 lines per millimeter at the lower focal length and at least 55 lines per millimeter at the upper; one can move in a few seconds between the two, and since the shift is managed by internal optical change the barrel length remains at a constant 30 inches and weighs only 40 pounds.

These boundaries of size and weight allow unequalled versatility in the use of the instrument for security and tracking purposes. Typically it is teamed with a video monitor, which permits constant observation of a border area, a missile site, an airfield or a harbor.

In many ways this is the most sophisticated of the Questar Security Systems, ideal for a variety of uses where the observer must be at a great distance from the area or activity under scrutiny. The dual focal length is particularly important for objects in motion, and its usefulness is enhanced by tower installations, where "seeing" limitations imposed by heat and humidity can be more readily overcome.

This optical system is as unique in quality as in concept. Both are necessary to give the user such remarkable access to his subject, and enable him to make his judgments with assurance; he will see exactly what is there.

QUESTAR

Box 4020, New Hope, Pa. 18938 (215) 862-5277

literature on request.

expressed in years; the jargon of geological-epoch names is quite secondary. A fine set of little maps shows the continents adrift back over 350 million years, and other maps show the plates, their boundaries and their active zones. Second, says Dr. Derry, the book was sized to fit into a briefcase; it is meant for those who fly and read, one imagines. The final aim, to set a cost low enough to be acceptable to a wide readership, is perhaps the least successfully attained. On the other hand, cash-flow problems do not much affect the well-backed community of the "traveling amateur and the professional earth scientist." Students will need to persuade friends or librarians to get them the atlas.

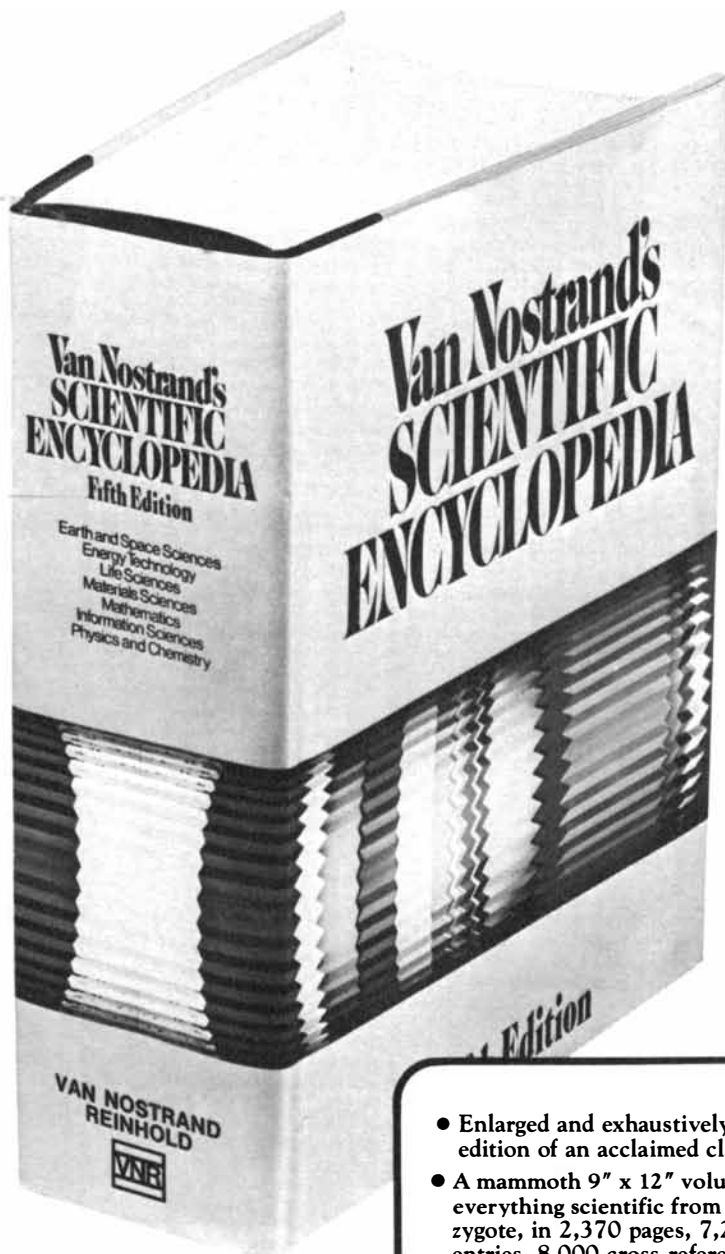
This is a true resource, the first of its kind in any language for more than a dozen years. There is a glossary, and a long list of sources for more maps and reports is arranged by map sheet. (To learn more about the Tibesti "hot spot" write to the listed agency in N'Djamena, Chad, box number included.)

THE ART OF ELECTRONICS, by Paul Horowitz and Winfield Hill. Cambridge University Press (\$24.95). The computer is so salient now that we associate the very idea with those word-and-symbol processors that take in our names and put out our paychecks, or those that respond to our mathematical queries. The same technology is also commonplace in the laboratory, where it skeptically queries and senses the world, more of Hume than of Plato. We know about the pulsars as we know about feature recognition in neurons of the cerebral cortex: because some well-formed configuration of transistor circuitry has told us.

This volume, avoiding both the lee shore of excessive handbook detail and the windy tack of cookbook oversimplification, is a basic navigator's guide to laboratory electronics as it is today. It is meant for serious study, hands on the oscilloscope probe, for those who would go on to carry out experimental research in just about any branch of science or technology. It omits all theory that is less useful than it is conventional for students, to aim straight at the art of circuit design; it is roughly quantitative, studiously analytical, full of clever circuits and sharp insights, but with a surprising minimum of mathematics. The style is wry and casual, almost deprecatory, the very speech of the laboratory. The depth is genuine, as is the richness of examples, data and apt tricks. There is an indispensable introduction to the Anglo-American electronics culture: data sheets, magazines, how to recognize a mystery chip, which firms supply a "nice selection of transformers and quick delivery," even how to avoid mislabeled counterfeit integrated circuits.

It takes until the third chapter to reach the sovereign idea of feedback, past a

The most comprehensive
and authoritative scientific reference
in the world.



Van Nostrand's Scientific Encyclopedia

a \$69.50 value

Yours **\$6⁹⁷**
for only

When you join the **Natural Science Book Club**.
You simply agree to buy 3 more books—at handsome discounts
—within the next 12 months.

Find the answers to virtually *all* your scientific questions, in the thoroughly updated and expanded fifth edition of Van Nostrand's SCIENTIFIC ENCYCLOPEDIA. It has been hailed by *The New York Times* as "an amazing book . . . For both the general and scientific reader."

Thousands of photos, diagrams, graphs and charts completely illuminate the ENCYCLOPEDIA's clear and thorough coverage of every scientific field—including archaeology, astronomy, general science, biology, medicine, physics, ethology and meteorology. Whether you're looking for a basic explanation of earthquakes or a comprehensive analysis of nuclear reactors—you'll find it all in Van Nostrand's SCIENTIFIC ENCYCLOPEDIA.

What is the Natural Science Book Club?

A unique book club devoted to bringing you exciting, refreshingly different books on all areas of the natural sciences. From the life story of early man to the death of a solar system . . . from the largest mammals to the invisible kingdom of microscopic organisms . . . all the wonders of our natural world are revealed.

Begin enjoying the Club's benefits by accepting Van Nostrand's SCIENTIFIC ENCYCLOPEDIA. It's the perfect guide to the marvels of nature . . . and it's a great bargain, too!

- Enlarged and exhaustively updated edition of an acclaimed classic.
- A mammoth 9" x 12" volume covering everything scientific from abacus to zygote, in 2,370 pages, 7,200 editorial entries, 8,000 cross-references and 2.2 million words . . . plus 2,450 photos, diagrams, and charts as well as 550 tables.
- Authoritatively compiled by over 180 internationally respected scientific experts.

If reply card has been removed, please write:
Natural Science Book Club—Dept. 4-AL6, Riverside, N.J. 08075
to obtain membership information and application.



FROM 0-115 MPH, IT'S ONE OF THE

Beautiful?

That's usually the last thing people have to say about Saabs.

But we've found that the people who think Saabs are anything but beautiful are the people who've never driven one.

Those who have are of a different opinion.

Performance, not appearance, is what people remember after they've driven the Saab 900 Turbo. Here are a few reasons why:

THE POWER PLANT, BEAUTIFUL.

The Saab 900 Turbo power plant is a beautiful solution to a seemingly ugly problem. Namely, how to achieve

increased power and fuel economy from the same engine.

The answer: turbocharging.

But not just any turbocharger. Saab engineers use a smaller, lighter turbocharger that spins faster, and which is, therefore, more efficient. And unlike other turbocharged cars, the Saab doesn't have to be taken to Watkins Glen to be tested out. The Saab 900 Turbo-charger kicks in at about 1,500 rpm, which is low enough for you to feel a few "G's" on your way to the supermarket.

At Saab, *that's* what we call beautiful.

BEAUTIFUL HANDLING.

1980 will be remembered as the year

the whole automotive world decided that front-wheel drive gives a car better handling.

Saab came to that conclusion 31 years ago.

So it can certainly be said that Saab is experienced in designing and producing front-wheel drive cars.

This may account for our incredible string of victories in the world's toughest winter rallies.

And together with its rack-and-pinion steering, independent double-wishbone front suspension and unsplit rear axle, it certainly accounts for the fact that the Saab 900 Turbo handles like no other car



MOST BEAUTIFUL CARS IN THE WORLD.

on the road.

Effortlessly, flawlessly, superbly.

At Saab, *that's* what we call beautiful.

AERODYNAMICALLY BEAUTIFUL.

But if you really want to see how beautiful the Saab 900 Turbo is, you have to look at it in a wind tunnel.

You have to see the way its wrap-around windshield, its bumpers, and even its side mirrors slice through the wind.

You have to notice that unlike some other luxury imports, there are no sharp, angular corners on a Saab that can cause air turbulence. Even its rear end is

shaped to give it efficient air breakaway.

At Saab, form allows function. And we've managed to design the Saab 900 Turbo 5-speed to function at ⁽¹⁹⁾EPA estimated mpg, and 30 estimated highway mpg. (Remember, use estimated mpg for comparison only. Mileage varies with speed, trip length and weather. Actual highway mileage will probably be less.)

That's what we call beautiful.

BEAUTIFUL TEST DRIVE.

We could go on and on about what a beautiful car the 900 Turbo is but the only way you're going to agree is to take a test drive.

Here's what Motor Trend magazine had to say after they took a spin: "*For about the same price as you'd pay for a Type S Turbo Riviera, you can have this Saab for a playmate. The 900 Turbo has it all: precise steering, lithe suspension, ultra-responsive engine, boat-anchor brakes and, above all, a wonderfully balanced personality.*"

We think what they are trying to say is, "It's a beautiful car."

SAAB
The most intelligent car
ever built.



These cassette deck makers have very high standards.



And this is how they set them.

With TDK SA. It's the high bias reference standard for these major cassette deck manufacturers. Which means that when they set the high bias position on their decks, there's a TDK SA inside. Since all these decks are adjusted to sound their best with TDK SA, it stands to reason yours will, too.

A lot of deck owners agree. When they listen to the more than 35 manufacturers who recommend TDK SA, they try it. When they listen to their music on TDK SA, they're convinced. They feel TDK SA's Super Avilyn formulation handles their music better than any chrome tape. And TDK SA's Super Precision Mechanism assures years of reliable performance. Each TDK SA cassette is backed by a lifetime warranty.*

You can listen to all the cassettes that might meet the high bias standard. Or you can buy the one that sets it. TDK SA.

TDK
The Amazing Music Machine

*In the unlikely event that any TDK cassette ever fails to perform due to a defect in materials or workmanship, simply return it to your local dealer or to TDK for a free replacement. © 1980 TDK Electronics Corp. Garden City, N.Y. 11530

chapter on the current-voltage foundations of circuit theory and one on discrete transistor circuits (made workably simple, with solid-state physics just about ignored). Feedback is introduced by way of operational amplifiers, those clever packages of a score of transistors and resistors that act as very-high-gain differential direct-current amplifiers and are never used without external feedback. The initial treatment is based on an idealized set of rules, in fact the limiting case of feedback with infinite open-loop gain. The fundamental gain equation itself enters only after 35 pages of powerful and quite understandable examples.

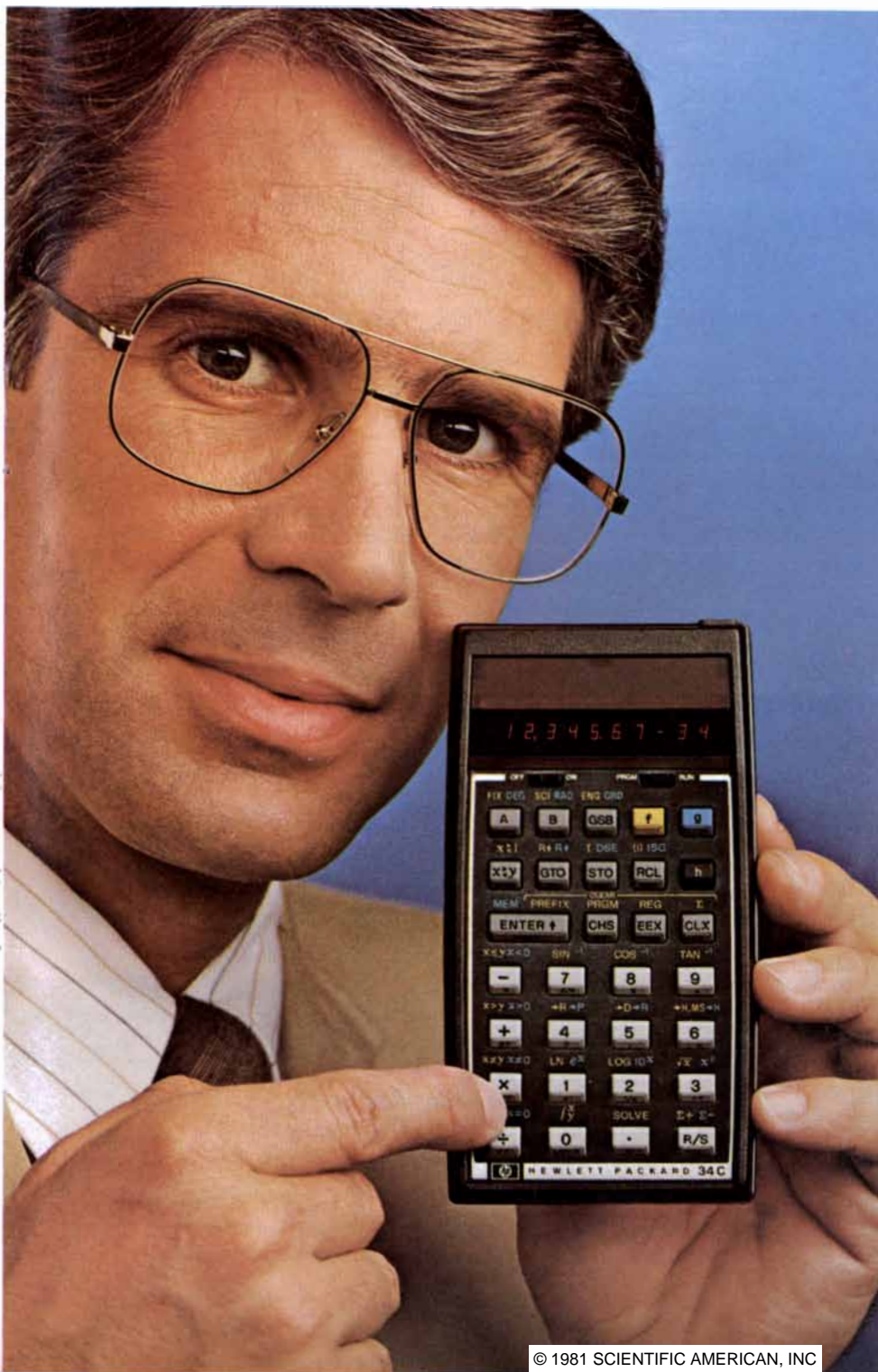
This scheme of making maximum use of the insights of simple approximations described in words, and introducing more formal design only after understanding has gained real headway, is the pedagogical core of this original treatment. Filter design, for example, is treated in the same nonmathematical, estimational fashion, with tables and charts but with barely a whisper of the intricacies of poles in the complex plane and the Chebyshev polynomials. Yet broad understanding is never sacrificed along with the formulas.

Half of the text treats analogue and half treats digital. We approach the digital magic in the same sensible way, first with arrays of gates alone and their logic. Only then do we step into the realm of sequential logic, where output depends not only on input but also on history: the same basis as mind. Modest entry here is made by way of the basic flip-flop, two simple cross-fed NAND gates. After some additional work on analogue-digital interfaces and on noise (including the intriguing digital pseudo-random noise schemes) there is a long look at minicomputers and their software and then at a microprocessor and how to feed, teach and control it. The problems of power supplies, of precision and even of construction are not forgotten, with mechanical and thermal attention from the breadboard up.

The last chapter, perhaps the most readable one for someone not engaged in laboratory work itself, is a fine overview of measurement. It explains a variety of transducers for sensing the world, the wonderful precision with which we can handle frequency and time, and the power of some ingenious schemes, such as lock-in detectors and signal averaging, that can all be seen as ways to narrow the detection bandwidth.

A dozen appendixes give instructions for the use of the oscilloscope, for drawing diagrams neatly and for finding data sheets, a math review and the necessary color codes. The volume is a nice match of the conceptual and the practical, produced by a pragmatic Harvard physicist, widely known for his electronics flair, and an engineer from the demanding redoubts of Route 128.

HP-33C, 34C, & 38C with Continuous Memory. They're ready when you are.



These advanced programmable calculators from Hewlett-Packard retain your data and programs even after they've been turned off. Just think how much more quickly and easily you'll be able to solve your business or scientific problems.

But these remarkable calculators do more than just remember. Consider the HP-34C, for instance. It provides up to 210 program lines. Up to 3 keystroke instructions are automatically merged into one program line, expanding memory to the equivalent of 370 program lines. And editing is a snap. The HP-34C's editing features let you review your programs and insert or delete instructions as needed. The HP-34C also has two special function keys, Solve and Integrate, to determine the roots of an equation or compute the definite integrals of a function.

Hewlett-Packard supports the HP-33C, 34C and 38C with the most comprehensive instruction manuals and software solutions to problems for science, engineering and business applications.

Visit a Hewlett-Packard dealer now and experience the advantages of Continuous Memory in one of these outstanding calculators.

- HP-33C Programmable Scientific with Continuous Memory. *New low price* \$90*
- HP-34C Advanced Programmable Scientific with Continuous Memory \$150*
- HP-38C Advanced Financial Programmable with Continuous Memory \$150*

For details and the address of your nearest HP dealer, CALL TOLL-FREE 800-547-3400, Department 236Y, except from Hawaii or Alaska. In Oregon, call 758-1010. Or write to Hewlett-Packard, 1000 N.E. Circle Blvd., Corvallis, OR 97330, Dept. 236Y.

*Suggested retail price excluding applicable state and local taxes—Continental U.S.A., Alaska and Hawaii. Display photographed separately to simulate typical appearance. 610/13





Jameson Irish
"Scotch on the rocks."

If you like fine Scotch, you'll love light, imported Jameson Irish.

Try a glass of Jameson Irish the way you would your favorite Scotch. With water. Soda. On the rocks.

You'll notice how much it tastes like fine Scotch — only lighter and more

delicate.

The dedicated Scotch drinker will instantly appreciate this flavor difference.

Though it may take a little time getting used to saying, "Jameson Irish on the rocks, please."

Jameson. World's largest-selling Irish Whiskey.

80 PROOF • CALVERT DIST. CO., N.Y.C.

Catastrophic Releases of Radioactivity

The gravest conceivable accident to a nuclear reactor is far less destructive than the detonation of a nuclear weapon, even if it is imagined that the weapon causes harm only by radiation

by Steven A. Fetter and Kosta Tsipis

A large population could be exposed to a dangerous amount of radioactivity in a number of ways. Such an event would be the inevitable consequence of even the most limited nuclear war. It could happen if an accident to a nuclear reactor caused its containing vessel to burst and allowed material from the core of the reactor to escape into the atmosphere. The inadvertent release from a reactor of water or gases bearing radioactive nuclei would create the danger of an exposure of lesser magnitude. Still another possibility is an accident during the manufacture, transportation, reprocessing or storage of radioactive material for nuclear reactors or nuclear weapons.

There are large differences in the amounts of radioactivity that could be released in such events, and so each possibility must be considered separately. We shall therefore describe the radioactivity that would probably be released in each of three events. The first is the detonation of a thermonuclear weapon at ground level. The second is the melting of the core of a nuclear reactor and the bursting of its containment vessel, with a resulting escape of radioactivity. The third is the explosion of a thermonuclear warhead on a nuclear reactor.

We want to emphasize that we shall not include in these comparisons the blast and the heat that constitute the prompt, explosive effect of a thermonuclear weapon. We shall examine and compare only the delayed effects brought on by the release of radioactivity. It emerges nonetheless that the detonation of a nuclear weapon is far more to be feared than any accident to a reactor. Moreover, the detonation of a

weapon on a reactor is many times more damaging than the detonation of a weapon on the ground. The nuclear attack turns the reactor into a devastating radiological weapon.

A thermonuclear weapon is usually made up of three parts. The first part is in essence a trigger whose most important component is a few kilograms of plutonium. The fissioning of the plutonium nuclei gives rise to the heat that is needed to set off a thermonuclear explosion.

The second part is the thermonuclear explosive: a mixture of deuterium and tritium, the heavy isotopes of hydrogen. The thermonuclear fusion of a deuterium nucleus (which has one proton and one neutron) with a tritium nucleus (one proton and two neutrons) yields a helium nucleus (two protons and two neutrons). The extra neutron is emitted at high velocity and a quantity of energy is released as heat. The products of this reaction do not have long-lasting radioactivity.

The third part of the weapon is a mantle of uranium that surrounds the layer of deuterium and tritium. The nuclei of uranium atoms fission when they are bombarded by the neutrons emitted in the course of thermonuclear fusion. The fragments of the fissioned nuclei are a source of copious radioactivity. In a three-part thermonuclear weapon about half of the energy released comes from thermonuclear fusion and half from the fissioning of the uranium.

The heat generated by the detonation of a thermonuclear weapon vaporizes the weapon itself almost instantly, and so the nuclear reactions stop. Most of

the nuclei created by the fissioning of the uranium are now in an abnormally energetic state. Their return to a state of lower energy is accompanied by the emission of radiation in the high-energy part of the electromagnetic spectrum, the part made up of X rays and gamma rays. This radiation heats the surrounding air, forming a shock wave that heats additional layers of air. The result is a luminous fireball. For a weapon that has an explosive yield of one megaton (the equivalent in energy of a million tons of chemical explosive) the fireball rises at a rate approaching 400 feet per second to an altitude of approximately 60,000 feet. A one-megaton yield is typical of a warhead on an intercontinental ballistic missile in the arsenal of the U.S.S.R.

The updraft generated by the rising fireball lifts large quantities of soil and debris. A one-megaton explosion at ground level can excavate a crater as much as 400 feet deep and 1,200 feet in diameter. As the fireball cools, the radioactive nuclei created by the explosion condense onto particles of dirt, which over a period of time return to the earth as radioactive fallout.

The fallout is radioactive in part because some of the newly created nuclei are unstable: in general they have an excess of neutrons. The instability is relieved when a neutron is transformed into a proton by the process called beta decay; in the course of the decay the nucleus expels an electron, which in this context is called a beta ray. Such transformations can again leave the nucleus in an excited state, from which it returns to a lower energy level by emitting electromagnetic radiation, mainly gamma rays. The fallout particles continue to

emit beta and gamma rays for many decades after the explosion. These emissions come at random times. Given a quantity of radioactive nuclei, one can predict only the average number of emissions in any given interval. With the passing of time the number of nuclei in an excited or an unstable state decreases, and so the intensity of the radioactivity diminishes.

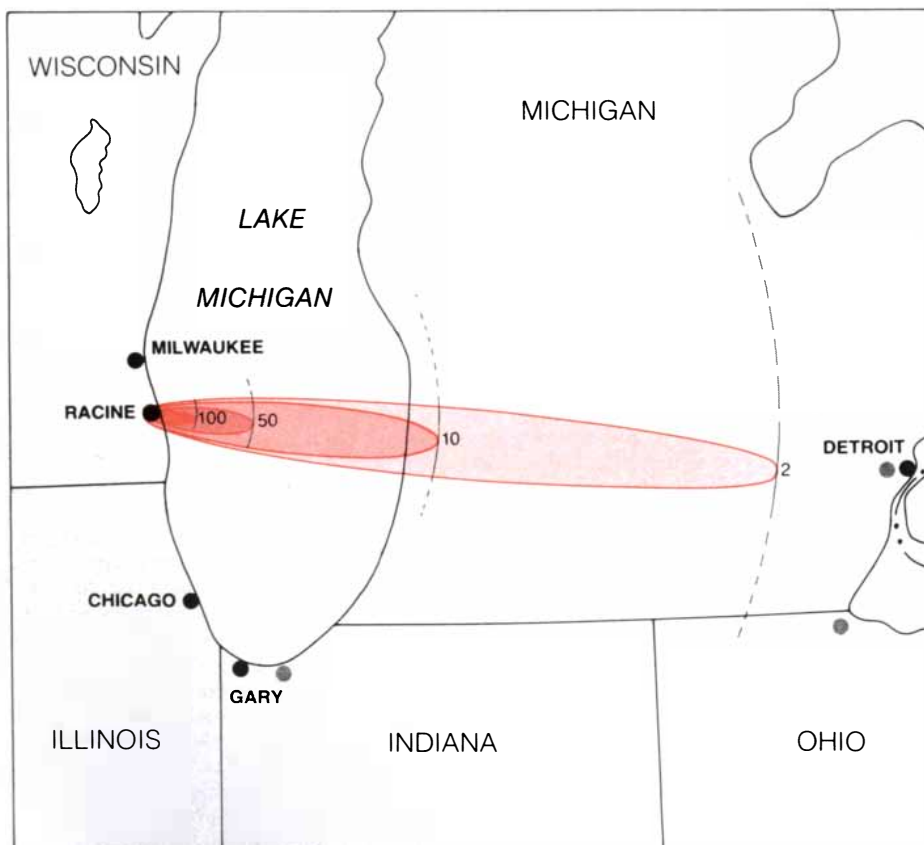
Several units of measure are employed to describe an amount of radioactivity or the amount of energy the radioactivity can deposit in living tissue. The standard unit for radioactivity itself is the curie, which is defined as 3.7×10^{10} emissions per second. The definition makes no reference to the kind of radiation or to the energy it carries. The unit for the deposition of energy by radioactivity is the rad, which is defined as the absorption of 100 ergs by a gram of matter such as living tissue. Another unit is the roentgen, which pertains exclusively to gamma rays and X rays. Exposure to one roentgen of gamma rays is equivalent to the absorption of about 94 ergs by a gram of tissue. Hence the rad and the roentgen are almost equal.

Because none of these units describes the amount of biological damage caused by the radiation, still another unit of measure is needed. The unit is the rem, an abbreviation of roentgen equivalent man. A dose of radiation measured in rem takes into account the fact that different kinds of radiation may have quite different effects on a living organism, even if they deposit the same amount of energy and cause damage by the same general mechanism, namely the ionization of atoms in intracellular molecules. The differences in damage reflect such characteristics of the radiation as the distance to which it penetrates in a given tissue. A dose in rem is equal to a dose in rads multiplied by a factor called the relative biological effectiveness (RBE) of the particular form of the radiation. For beta and gamma rays the RBE is approximately 1. In the discussion that follows, therefore, a dose in rads will be considered equal to a dose in rem. A sense of the size of the dosages we shall consider can be gained by comparing them with the following pair of examples: a chest X ray entails a dose of some .01 rem, absorbed in a fraction of a second; the natural background radia-

tion at sea level amounts to about .075 rem per year.

The biological effects of radiation vary considerably from one person to another. They depend, for example, on age and health. For this reason it is not possible to cite precise radiation levels at which one could expect to see particular symptoms of radiation sickness, such as loss of hair, vomiting, diarrhea, internal bleeding or lesions in the mouth and throat. Still, it has been established that if the human body is exposed to more than 500 or 600 rem in an interval not much longer than a day or two, survival is almost impossible. If the dose is between 200 and 450 rem, survival is possible but by no means assured, even if medical care is available. All things considered, it seems reasonable to assume that a dose of 400 rem in a day implies a mortality rate of 50 percent or more. Exposure of a population to 100 rem in the same period would cause sickness and some deaths. At this dosage, however, most people could be expected to recover even without medical attention.

In calculating the land area made uninhabitable by a given release of radioactivity we shall take the maximum ac-



DOSE CONTOURS resulting from three hypothetical releases of radioactivity are compared. The map at the left shows the pattern that develops a week after an accident in which the core of a one-gigawatt nuclear reactor releases a third of its radioactivity. The site of the hypothetical accident is Racine, Wis. The amount of radioactivity released is 100 million times the amount released during the accident at

the Three Mile Island Nuclear Generating Station near Harrisburg, Pa., in March, 1979. The map in the middle shows the pattern that develops a week after a one-megaton thermonuclear weapon explodes at ground level in Racine, sparing the reactor but initially releasing a far greater quantity of radioactivity. The map at the right shows the pattern that develops a week after a one-megaton weapon vapor-

ceptable dose to the general public to be 2 rem per year. This dose is more than 10 times the maximum recommended by the U.S. Environmental Protection Agency, and it is more than 20 times the natural background radiation. On the other hand, it is less than the 5 rem per year now considered an upper limit for workers exposed to radiation over a period of years. A standard of 2 rem per year might well be adopted in the aftermath of a peacetime reactor accident. In a nuclear war, however, the public could hardly be excluded from all areas in which the radiation level is 2 rem per year. Indeed, people driven by hunger or other imperatives might be willing (or compelled) to occupy areas in which they would absorb more than 50 rem per year, a dose that causes radiation sickness in more than half of the people exposed. A dose of 50 rem per year also causes occasional fatalities and may cause cancer in some people years after the exposure.

We now turn to the consequences of the release of radioactivity by a one-megaton thermonuclear weapon detonated at ground level. Most of the

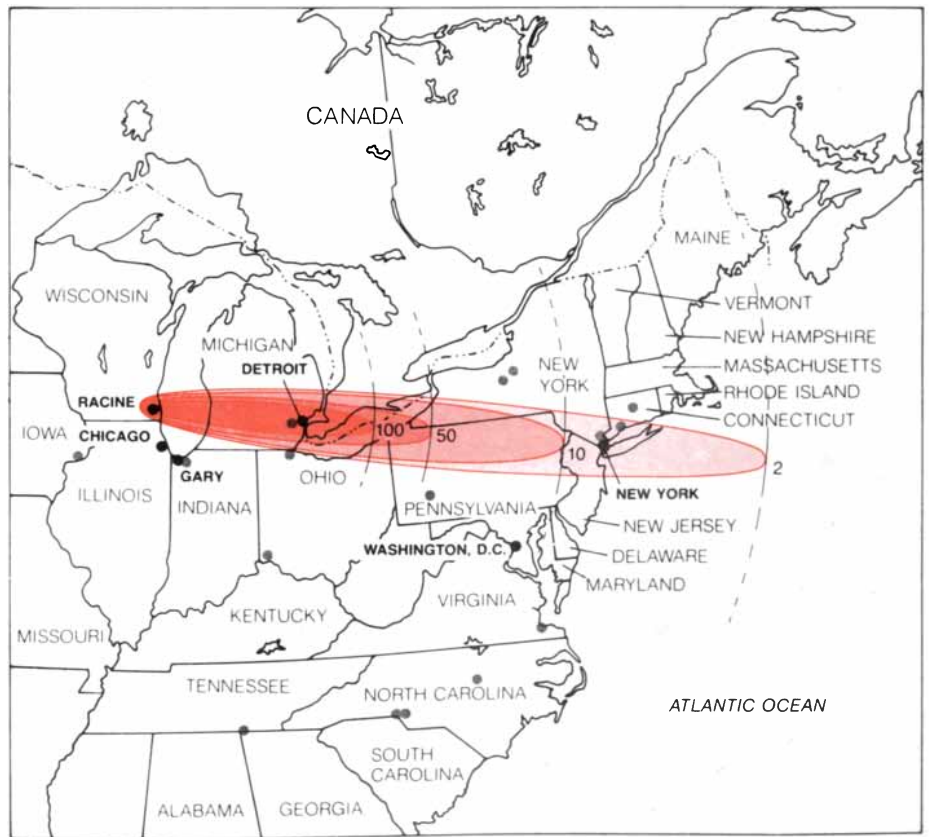
resulting radioactive fallout returns to the earth downwind from the explosion, and as much as 70 percent of the fallout takes the form of relatively large particles that return to the earth within a day. The intensity of the radioactivity decreases with distance from the site of the explosion. For one thing, the cloud of debris loses dust particles as it drifts with the wind. In addition the total radioactivity diminishes as the radioactive nuclei decay.

If the wind is steady, the pattern of accumulated dose of radioactivity (in rem) is a set of nested cigar-shaped contours. Each contour denotes a particular dose and all points within the contour are points where the dose is greater. We shall assume a wind speed of 15 miles per hour. In that case the lethal zone—the area circumscribed by the contour line denoting an exposure to 400 rem in 24 hours—amounts to roughly 400 square miles. The number of fatalities in the lethal zone would depend largely on the population density. In the U.S., for example, the population density ranges from 100,000 per square mile in metropolitan areas during business hours to fewer than five per square mile. The ra-

diation from the detonation of a single warhead could therefore kill from a few hundred to several million people. The total would depend not only on the site of the explosion but also on the time of day, the weather conditions, the efficacy of any warning issued and the available protection from radiation.

Those who escape death in the lethal zone would be unable to return to the zone for a long time because the ground would be contaminated by radioactive particles. The survivors would have to wait until the effects of radioactive decay and the settling of the contaminants into the ground by way of rain and snow reduced the radioactivity to an acceptable level. At a maximum acceptable dose of 2 rem per year some 1,200 square miles of land would remain unfit for human use for a year. Larger areas would be affected for a shorter time. The disruption of society would be immense. For example, more than 20,000 square miles would be uninhabitable for a month. Plainly this could dislocate many hundreds of thousands of people.

In an attack involving several weapons the cumulative radioactivity would almost certainly deny to the surviv-



izes the core of a one-gigawatt reactor. Radioactivity from both the reactor and the weapon is spread over the landscape. In each case the prevailing wind is from the west at 15 miles per hour. The plume of debris could drift, however, in a number of directions (gray circles). Doses are given in rem per year. One rem represents the amount of radiation that deposits 100 ergs in a gram of tissue. The natural back-

ground radiation at sea level is approximately .075 rem per year. The exposure of a population to 2 rem in the course of a year might increase the long-term incidence of cancer. An exposure to 50 rem in a year would cause instances of radiation sickness. Gray dots mark places where a commercial nuclear reactor has been built or is being built within 25 miles of a city with population of more than 100,000.

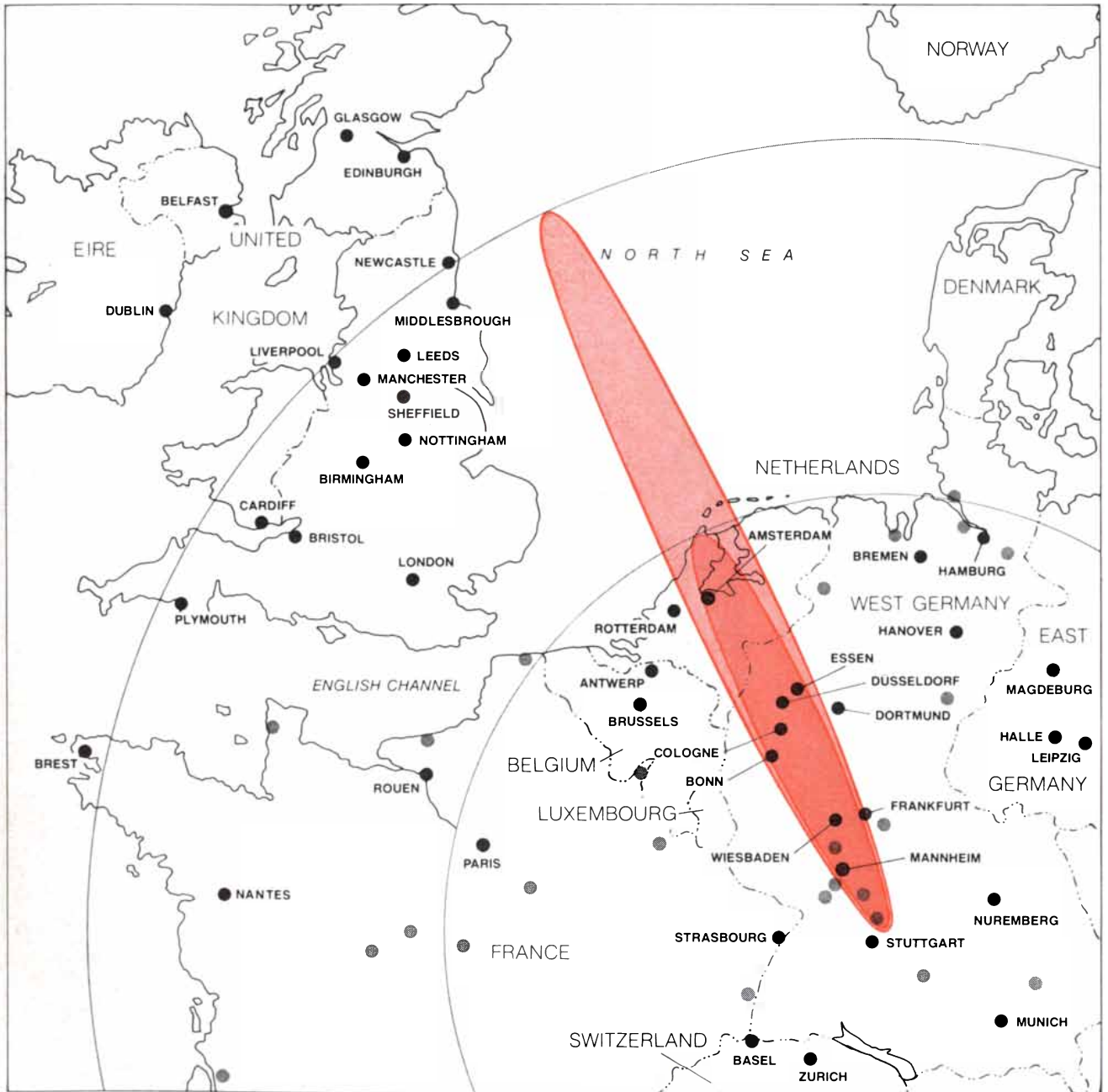
ing population the use of contaminated workplaces and farms that had escaped destruction in the explosions. Even if the survivors were willing to occupy areas in which they were exposed to doses far greater than 2 rem per year, they would still be unable to use large tracts of land. Each one-megaton weapon creates a zone of 1,500 square miles in which the dose of radiation remains at least 50 rem per year for a month.

In contrast to a nuclear weapon, a nuclear reactor cannot explode. The re-

actor liberates energy by means of nuclear fission, but even in a reactor that is entirely out of control the rate at which energy is released is more than 10^{12} times slower than it is in a nuclear weapon. Moreover, the energy released in the reactor is absorbed initially by the mass of the core, which is hundreds of times greater than the mass of a nuclear weapon. As a result the temperature of the core even in a runaway reactor rises only slowly.

If the temperature of the core became

too great, the fuel elements would melt and the core would fall apart before a chain reaction could generate an explosive amount of energy. A breach of the reactor's containment vessel could then result in a release of radioactivity. In one conceivable accident the total loss of coolant to the fuel rods in the core allows the rods to overheat and melt. The molten material then makes contact with water and a steam explosion tears open the containment vessel, with an ensuing release of radioactive material. In



ATTACK ON A SINGLE REACTOR with a single nuclear weapon could devastate a substantial part of Europe. Here a hypothetical attack has been made on a one-gigawatt nuclear reactor at Neckarwestheim in West Germany. The weapon has an explosive yield of one megaton. The prevailing wind is southeasterly with a speed of 15

miles per hour. A month after the attack the zone in which the dose is 10 rem per year (*light color*) might extend well into the U.K. A year after the attack the 10-rem zone (*dark color*) still includes much of the industrial capacity of West Germany. Gray dots mark the locations of commercial nuclear reactors in West Germany and France.

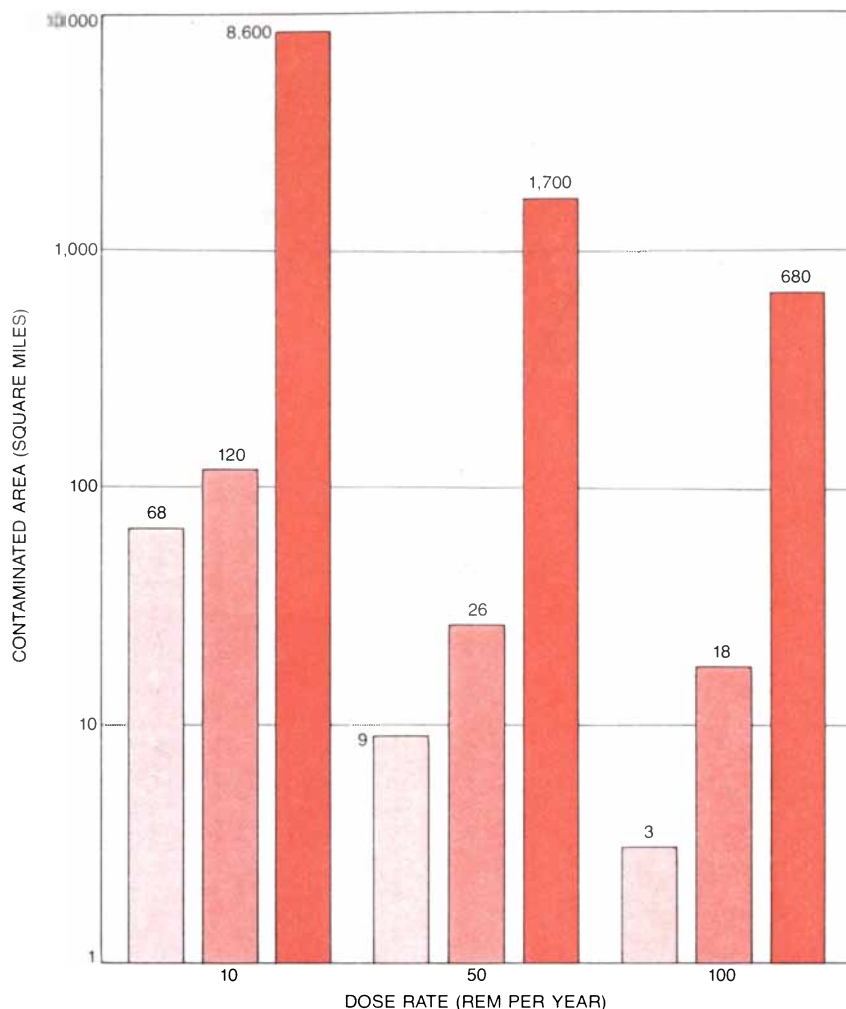
another conceivable accident the overheating of the core generates hydrogen or other flammable gases, which mix with atmospheric oxygen and then ignite and explode. Again the containment vessel is breached and radioactivity escapes.

In order to compare the hazard posed by a release of radioactivity from a reactor with the release caused by a nuclear weapon, we shall examine the consequences of such a worst-case accident, in which the containment vessel is ruptured. It should be pointed out that the probability of such an event is calculated to be many orders of magnitude smaller than the probability of a lesser accident, such as the one at the Three Mile Island Nuclear Generating Station near Harrisburg, Pa., in March, 1979.

The amount of radioactive material that would escape from a reactor and the composition of the material would depend on the exact nature of the accident and on the time that had passed since the reactor was last refueled. The dispersal of the radioactivity would depend on the shape of the plume of debris released by the accident and on the local weather conditions. Two general conclusions emerge. First, the rate at which radioactivity is released by a thermonuclear weapon is initially far greater than the rate of release by a reactor accident. The radioactivity from the weapon, however, includes a much larger proportion of isotopes whose radioactivity is short-lived. Second, comparatively little heat is released by the reactor accident. As a result the plume of debris remains at low altitude, and it deposits its radioactivity rather promptly. This tends to limit the size of the contaminated area. In sum, the area of land contaminated by the reactor accident is far smaller, but the land stays contaminated longer.

We shall consider a one-gigawatt (1,000 megawatt) nuclear reactor in which a third of the fuel is replaced each year. We shall assume that an explosion breaches the containment vessel and releases into the atmosphere a third of the reactor's content of radioactive nuclei. An hour after the release the radioactivity of the escaped material would be roughly 1.5 billion curies. The detonation of a one-megaton thermonuclear weapon would result in radioactivity 1,000 times greater. The accident at Three Mile Island released 100 million times less. (It released 17 curies of radioactive iodine.)

Let the wind speed again be 15 miles per hour. The crucial fact about the outcome of such an accident is that the contaminated area is indeed relatively small. Moreover, the exposure to radiation remains close to a level of 2 rem per year for people in all but a small part of the contaminated area. Specifically, the



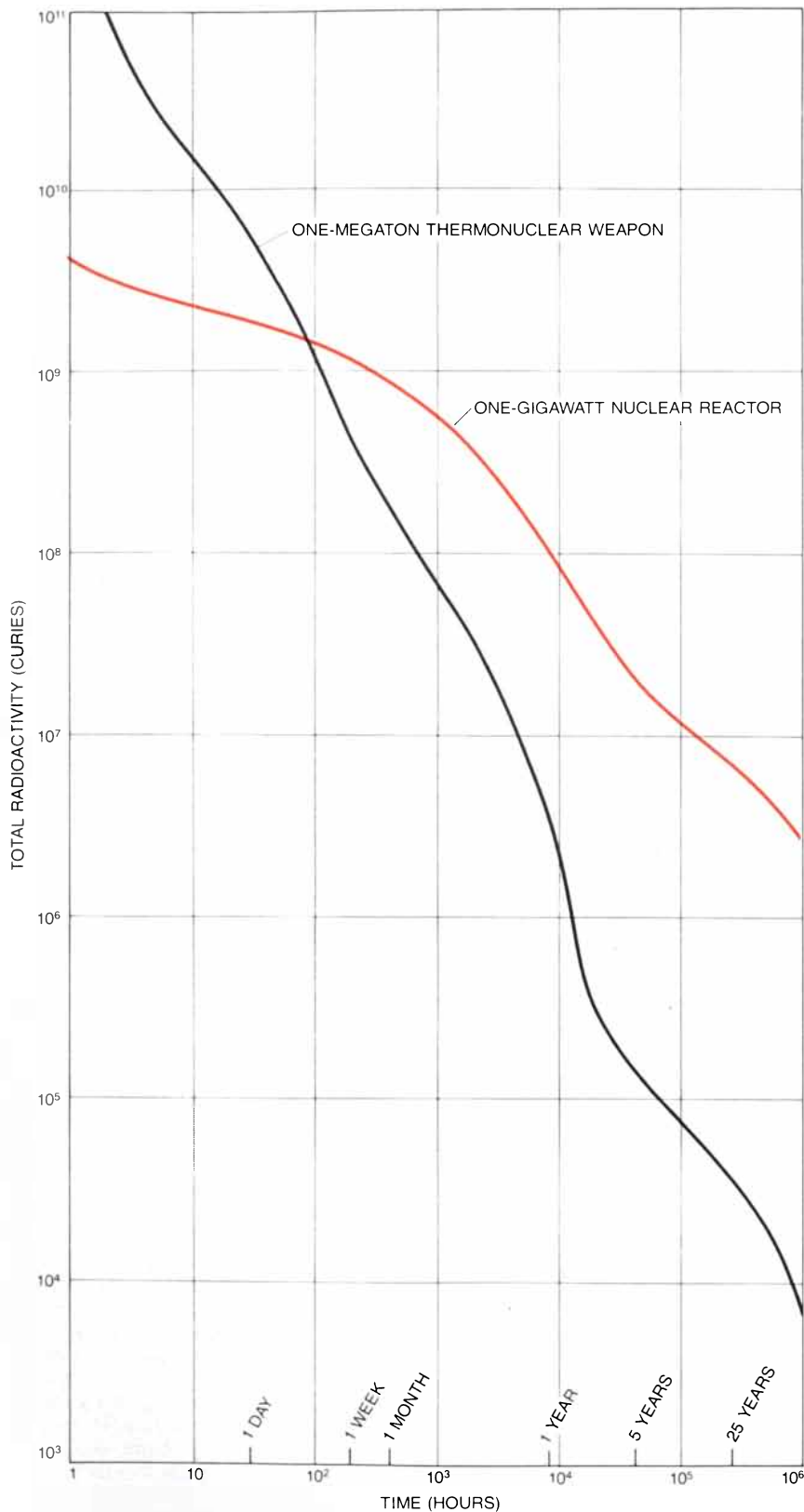
DENIAL OF LAND to the survivors of a release of radioactivity depends on the dose of radiation the survivors would be willing (or compelled) to absorb. Presumably a dose rate of even a few rem per year would be intolerable after a peacetime accident, whereas the survivors of a nuclear attack might attempt to endure far more. The bars show the amount of land that must remain uninhabited for a year if the maximum acceptable dose rate is 10 rem per year (left), 50 rem per year (middle) or 100 rem per year (right). Again three possible sources of radioactive contamination are considered: a grave reactor accident (light color), the ground-level detonation of a thermonuclear weapon (medium color) and the detonation of a thermonuclear weapon on a reactor (dark color). If more than 10 rem per year is unacceptable, the amount of land that must remain uninhabited for a year after the attack on the reactor is 8,600 square miles.

dose remains 2 rem per year for a month in an area of about 1,800 square miles. (The comparable figure for the detonation of a one-megaton weapon is more than 20,000 square miles.) The lethal zone, where the dose attains a level of 400 rem per day, is less than one square mile. (The lethal zone for the detonation of a one-megaton weapon is 400 square miles.)

The lower dose rate and the smaller size of the contaminated area in the case of a reactor accident suggest that people might be evacuated before they inhaled substantial amounts of radioactive dust. (This is the principal danger after the reactor accident.) It also seems possible that land could be decontaminated. In the case of the nuclear weapon the decontamination of land would be pre-

cluded by the far greater deposition of radioactivity.

In comparing the destructive outcome of a reactor accident with that caused by the detonation of a nuclear weapon it is useful to consider briefly the prompt destruction resulting from each event. A reactor meltdown causes no significant damage from blast or heat. A nuclear weapon, on the other hand, causes immediate devastation for five to 10 miles around the site of the explosion. In all likelihood, therefore, the detonation of the weapon would destroy or disrupt emergency and medical facilities. It is reasonable to conclude that if one population were exposed to a dangerous dose of radiation from a nuclear weapon and another population were exposed to the same dose released in a reactor acci-



DECAY OF RADIOACTIVITY released by the detonation of a nuclear weapon differs from the decay of the radioactivity released by a reactor accident because the two inventories of radioactive nuclei have different proportions of various isotopes. After an hour the radioactivity released by the detonation of a one-megaton thermonuclear weapon is 1,000 times greater than the radioactivity that would escape in the worst conceivable peacetime reactor accident. On the other hand, the radioactivity from the reactor takes longer to decay. The unit of radioactivity is the curie. A curie is 3.7×10^{10} emissions per second of various forms of radiation.

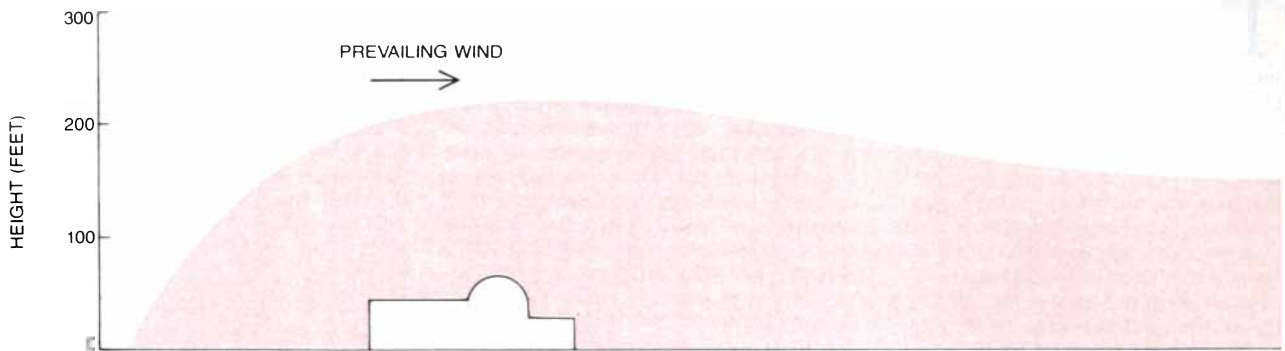
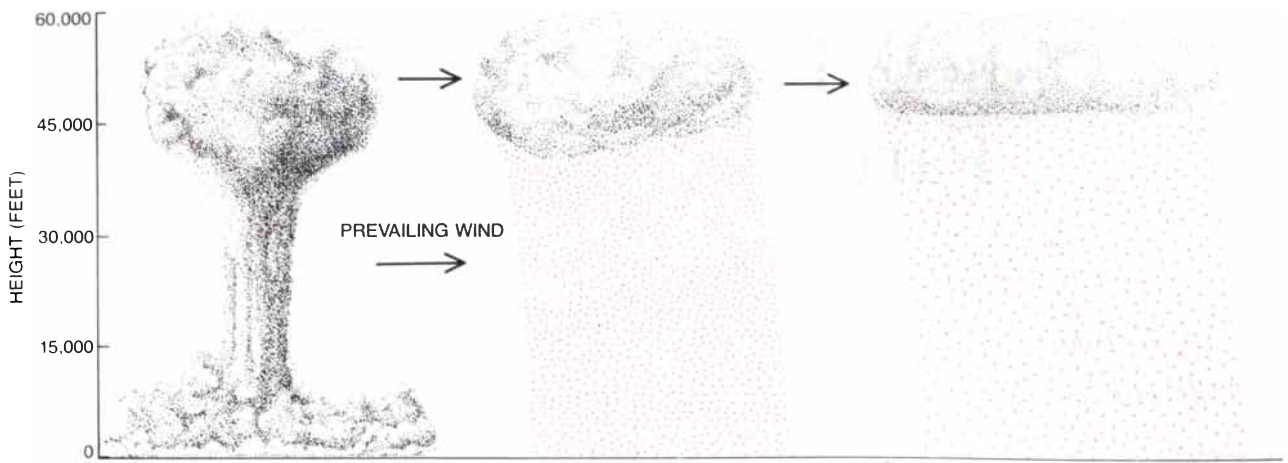
dent, the former population would have many fewer survivors than the latter. The reason is the disruption of the services needed by the victims of radiation exposure.

The probability of the two events we have considered is difficult to assess. Still, there seems to be a belief among both defense analysts and specialists in nuclear power that the probability of the detonation of a nuclear weapon somewhere in the world in the next 10 years is far greater than the probability of the catastrophic meltdown of a nuclear reactor. Among the reasons that might be cited in support of this view are the ever increasing number of nuclear weapons in the arsenals of several nations, the growing truculence that characterizes relations between the U.S. and the U.S.S.R. and the efforts of military planners to shift their nation's strategic policy from that of deterring a nuclear war to that of preparing to fight one.

We now consider the radioactivity that would be released if a one-megaton thermonuclear weapon were detonated on a one-gigawatt nuclear reactor. We shall assume that the radioactive material in the core of the reactor is completely vaporized by the explosion. The radioactivity of the core would then combine with the radioactivity derived from the weapon itself; both would rise with the fireball and return to the earth in the manner characteristic of the fallout from the explosion of a weapon alone.

Since the rate of radioactivity in the reactor is initially much less than the rate of radioactivity given off by the detonation of the weapon, the pattern of contamination in the first week would not differ greatly from the pattern caused by the weapon alone. Because the radioactivity from the reactor is relatively long-lived, however, the time a given area would remain contaminated is significantly greater. In essence, the debris from the weapon would contribute to the contamination a high level of early radioactivity and the debris from the reactor would contribute long-lasting radioactivity. The lethal zone for the detonation of the weapon on the reactor would be more than 500 square miles, an area a third larger than the lethal zone created by the detonation of the weapon alone. The area in which the cumulative dose remained 2 rem per year for a month would be 64,000 square miles, or three times larger. The area in which the dose would remain 2 rem per year for a year is 25,000 square miles, or 20 times larger. An area of 180 square miles would continue for more than a century to expose any occupant to a dose of at least 2 rem per year. Such an area would be a permanent monument to the catastrophe.

Vaporizing the cores of nuclear reac-



PLUME OF DEBRIS caused by the detonation of a nuclear weapon is also different from the plume caused by a reactor accident. The detonation of a one-megaton thermonuclear warhead at ground level (*upper drawing*) creates updrafts that lift the debris of the explosion to an altitude of some 60,000 feet. The debris, which is carried down-

wind, rains back onto the earth as radioactive fallout. In contrast, the non-nuclear explosion that breaches the containing vessel of the reactor (*lower drawing*) has relatively little energy, and so the released radioactivity is not carried to high altitude. The near absence of a plume reduces the spread of the radioactivity by the prevailing wind.

tors with nuclear weapons is clearly an efficient way to desolate large parts of a nation. Indeed, by waiting for suitable weather conditions, a determined or desperate combatant could devastate a substantial fraction of an opponent's industrial capacity with a single thermonuclear weapon. For example, an attack on a reactor in the Rhine-Neckar River valley could render uninhabitable about a third of West Germany's 96,000-square-mile area for a period of a month or more, even if cumulative doses of radioactivity much higher than 2 rem per year were acceptable to the survivors. The only condition on the attack is that the prevailing wind come from the southeast.

In thinking about such devastation it is well to remember that in central Europe, where the population is dense and the land is intensively exploited, power reactors may lie not far from military installations. The likelihood that a nuclear weapon aimed at a military target would unintentionally destroy a nearby reactor is therefore not negligible. It is

also well to remember that storage pools for radioactive wastes are on the same site as the reactor that produces the wastes. The wastes in a typical pool may soon represent an inventory of radioactivity two times greater than that of the reactor core itself. In addition reactors are often constructed in pairs a few hundred feet apart. All things considered, the dose rate following the detonation of a nuclear weapon on a reactor complex could easily be from two to six times higher than what we have calculated.

We can find no public evidence that military planners have carefully considered in any of their scenarios for nuclear war the deliberate or accidental vaporization of the core of a nuclear reactor in a nuclear attack. The best way to minimize the probability of such an event is to avoid all nuclear war. Some helpful steps would be the negotiation of a multinational agreement not to designate nuclear facilities as targets and efforts to ensure that military installations are not situated near civilian reactors.

If a single conclusion is to be drawn from the analysis we have offered, it must be that even a single nuclear weapon would contaminate a much greater area with radioactive fallout than the worst conceivable accident to a nuclear reactor. In view of this the preoccupation of the public with the risks of the generation of electricity by nuclear reactors appears to be misplaced. A catastrophic reactor accident would doubtless cause considerable disruption in its immediate vicinity. It would probably cause long-term medical problems and even some loss of life. Still, the impact of the accident could be moderated, because social, governmental and medical services would be intact and functioning, even in the contaminated area. Moreover, the risks posed by reactors can be minimized by the thoughtful application of technology. A nuclear attack is fundamentally different. The point cannot be overstated: nuclear war poses the ever present danger of suffering and death on a scale unparalleled in human history.

A Unified Theory of Elementary Particles and Forces

At a range of 10^{-29} centimeter the world may be a simple place, with just one kind of elementary particle and one important force. If the proposed unified theory is correct, all matter is unstable

by Howard Georgi

There can be nothing simpler than an elementary particle: it is an indivisible shard of matter, without internal structure and without detectable shape or size. One might expect commensurate simplicity in the theories that describe such particles and the forces through which they interact; at the least one might expect the structure of the world to be explained with a minimum number of particles and forces. Judged by this criterion of parsimony, a description of nature that has evolved in the past several years can be accounted a reasonable success. Matter is built out of just two classes of elementary particles: the leptons, such as the electron, and the quarks, which are constituents of the proton, the neutron and many related particles. Four basic forces act between the elementary particles. Gravitation and electromagnetism have long been familiar in the macroscopic world; the weak force and the strong force are observed only in subnuclear events. In principle this complement of particles and forces could account for the entire observed hierarchy of material structures, from the nuclei of atoms to stars and galaxies.

An understanding of nature at this level of detail is a remarkable achievement; nevertheless, it is possible to imagine what a still simpler theory might be like. The existence of two disparate classes of elementary particles is not fully satisfying; ideally one class would suffice. Similarly, the existence of four forces seems a needless complication; one force might explain all the interactions of elementary particles. An ambitious new theory now promises at least a partial unification along these lines. The theory does not embrace gravitation, which is by far the feeblest of the forces and may be fundamentally different from the others. If gravitation is excluded, however, the theory unifies all elementary particles and forces.

The first step in the construction of the unified theory was the demonstration that the weak, the strong and the electromagnetic forces could all be described by theories of the same general kind. The three forces remained distinct, but they could be seen to operate through the same mechanism. In the course of this development a deep connection was discovered between the weak force and electromagnetism, a connection that hinted at a still grander synthesis. The new theory is the leading candidate for accomplishing the synthesis. It incorporates the leptons and the quarks into a single family and provides a means of transforming one kind of particle into the other. At the same time the weak, the strong and the electromagnetic forces are understood as aspects of a single underlying force. With only one class of particles and one force (plus gravitation), the unified theory is a model of frugality.

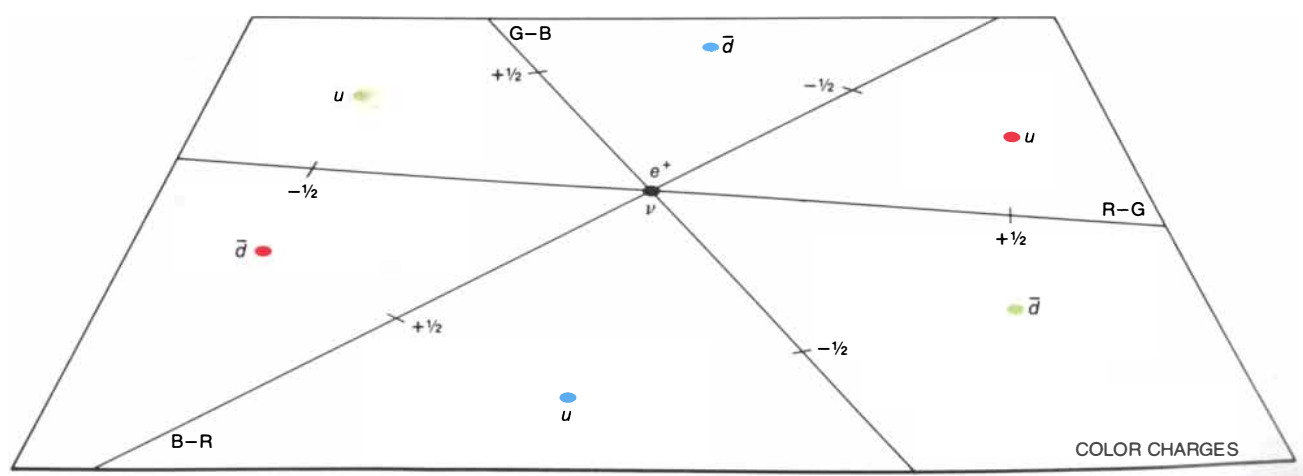
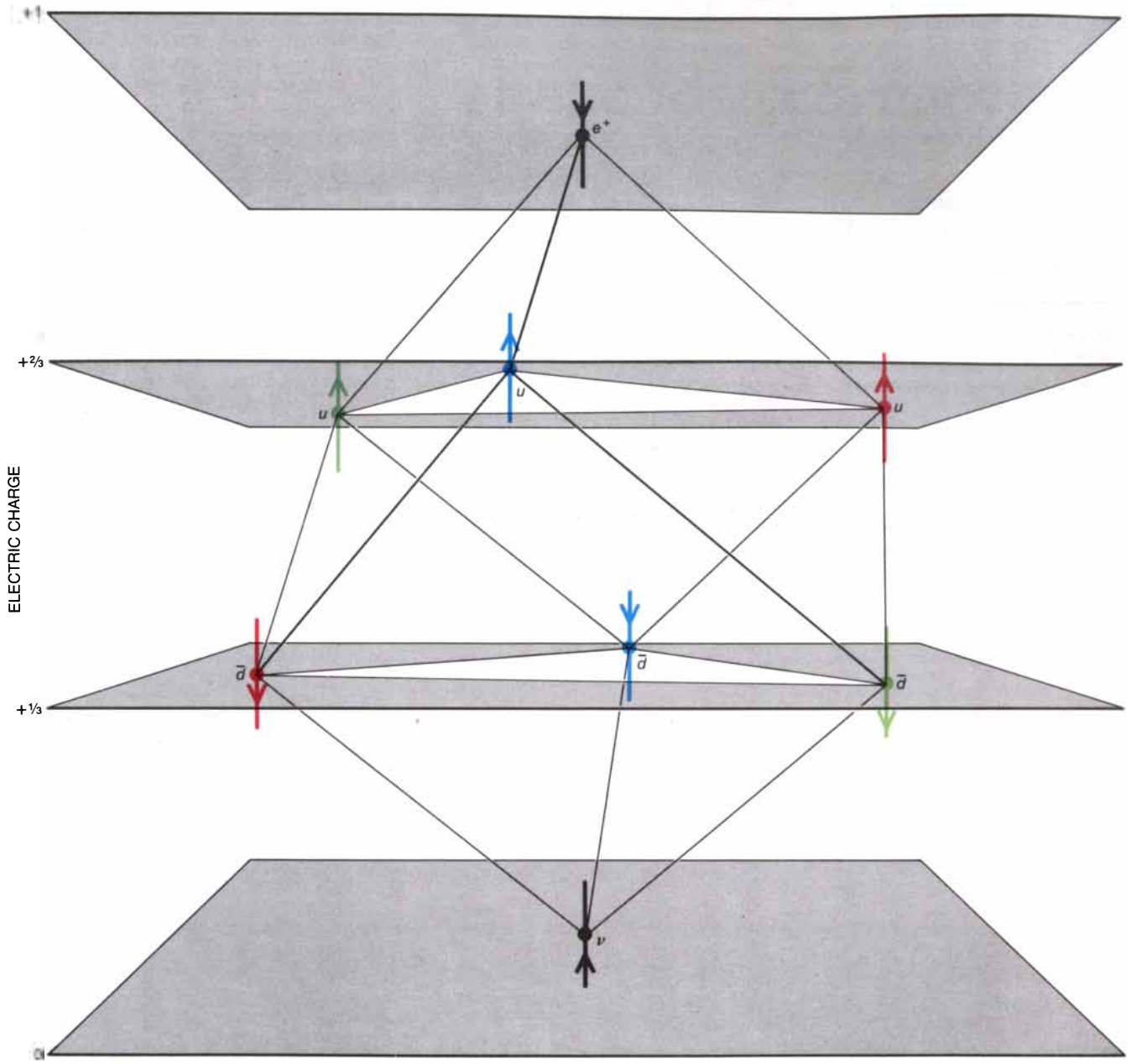
Leptons and quarks are known to have quite different properties; how can they be consolidated into a single family? The weak, the strong and the electromagnetic forces differ in strength, range and other characteristics; how can they be derived from a single force? The unified theory does not attempt to conceal the differences, but it asserts they are not fundamental. The differences are conspicuous mainly because the universe is now quite cold, so that particles generally have low energy. If experiments could

be done at extremely high energy, the unification would become apparent in all its simplicity. Leptons and quarks would be freely interconverted and the three forces would all have the same strength.

The energy needed in order to see the unification of particles and forces in this dramatic form is estimated to be about 10^{15} gigaelectron-volts, or GeV. (One GeV is the energy imparted to an electron when it is accelerated through a potential difference of a billion volts.) This energy exceeds the capabilities of even the largest planned particle accelerators by a factor of 10 trillion, and it is most unlikely such an energy will ever be attained in the laboratory. It may therefore seem that the theory can never be tested, but such is not the case; the theory has definite consequences at readily accessible energies.

First, the theory provides a rationale for several established features of the physical world that have long seemed mysteriously arbitrary. It accounts for the quantization of electric charge: the observation that charge always comes in discrete multiples of a fundamental smallest charge. It gives a value for the relative strengths of the three forces (measured at ordinary laboratory energy) that is in reasonably good agreement with experimental results. It might explain why there is more matter than antimatter in the universe. Equally important, the unified theory predicts new

CUBIC SYMMETRY emerges when certain properties of elementary particles are graphed in three dimensions. The particles are members of the families called leptons and quarks. The position of each particle on the horizontal plane is determined by three kinds of "color charge." The quarks designated u come in three colors and lie at the vertexes of an equilateral triangle; the antiquarks labeled \bar{d} have the corresponding three anticolors and form an oppositely oriented triangle. The leptons, which are represented here by the positron (e^+) and the neutrino (ν), have no color charge and lie at the center of the plane. When each particle is displaced vertically by a distance proportional to its electric charge, the cube takes shape. The fact that this arrangement of the particles yields a simple and symmetrical solid suggests some connection between the leptons and the quarks, a connection that can be explained by a unified theory.



phenomena that cannot be deduced from earlier theories. The most noteworthy of these predictions is the decay of the proton, a particle that had been considered absolutely stable. If the proton can decay, the atom itself is unstable, and all matter is impermanent.

The unified theory is not intended to supplant the established theories of the weak, the strong and the electromagnetic forces. On the contrary, the individual theories are embedded in the structure of the larger one. In order to explain the nature and the origin of the unified theo-

ry it is therefore best to begin with the individual component theories, with the forces the theories describe and with the elementary particles on which the forces act.

The apparent differences between the leptons and the quarks are substantial. Six leptons are known, and among them the electron can be considered prototypical. It has a small mass, equivalent in energy units to about 500,000 electron volts, and it has one unit of electric charge; by convention the charge of the electron is negative. Two other leptons, namely the muon and the particle designated tau, have the same charge and indeed seem to be identical with the electron in all properties except mass. The muon is more than 200 times as massive as the electron; the tau lepton, which was discovered only five years ago, has a mass almost 3,500 times that of the electron.

The remaining leptons are three kinds of neutrino, which are electrically neutral and have a very small mass (if they have any mass at all). One neutrino is associated with each charged lepton. In addition, for each of the six leptons there is an antilepton with the same mass but with the opposite electric charge. Thus the antielectron (or positron), the antimuon and the antitau particle all have a charge of +1. The antineutrinos, like the neutrinos themselves, have no electric charge.

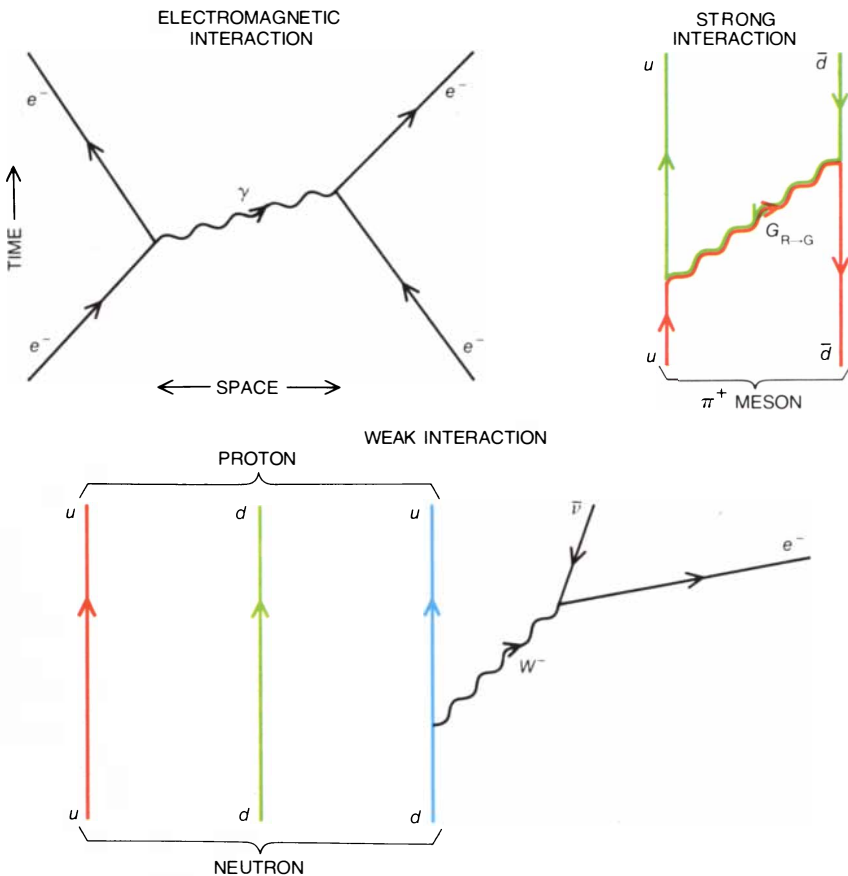
Whereas the leptons appear as free particles, no one has yet been able to examine a quark in isolation. The quarks are observed only as constituents of the particles called hadrons, a large and diverse class that includes the proton, the neutron, the pi meson and more than 100 other known particles.

There is compelling evidence for the existence of five kinds of quark: they are designated down (*d*), up (*u*), strange (*s*), charm (*c*) and bottom (*b*). A sixth kind labeled top (*t*) has been predicted, but so far it has not been seen. The kinds of quark are generally called flavors; in addition the quarks have another property called color. (Flavor and color are arbitrary labels, which have nothing to do with the sensations of taste or sight.) Each flavor of quark comes in three colors: red, green and blue. The property of color marks a major difference between the leptons and the quarks. The five or six flavors of quark correspond roughly to the six varieties of lepton, but there is no counterpart among the leptons of the quark colors. The distinction has observable consequences. The strong force is an interaction between colors, and since the leptons have no color, they are not susceptible to the strong force.

Another distinctive property of the quarks is their electric charge. The *d*, *s* and *b* quarks have a charge of $-1/3$, whereas the *u*, *c* and *t* quarks have a

		LEPTONS		QUARKS					
THIRD GENERATION	ν_τ	0		<i>t</i>	$+2/3$	<i>t</i>	$+2/3$	<i>t</i>	$+2/3$
	τ^-	-1		<i>b</i>	$-1/3$	<i>b</i>	$-1/3$	<i>b</i>	$-1/3$
SECOND GENERATION	ν_μ	0		<i>c</i>	$+2/3$	<i>c</i>	$+2/3$	<i>c</i>	$+2/3$
	μ^-	-1		<i>s</i>	$-1/3$	<i>s</i>	$-1/3$	<i>s</i>	$-1/3$
FIRST GENERATION	ν_e	0		<i>u</i>	$+2/3$	<i>u</i>	$+2/3$	<i>u</i>	$+2/3$
	e^-	-1		<i>d</i>	$-1/3$	<i>d</i>	$-1/3$	<i>d</i>	$-1/3$

LEPTONS AND QUARKS differ in a number of important properties, and so they have generally been classified in separate families. One of the most conspicuous differences is in electric charge, which is shown here for each particle: the lepton charges are integers, whereas the quark charges are fractions. Furthermore, the leptons exist as free particles, whereas the quarks are found only as constituents of the composite particles called hadrons. It is customary to divide the leptons and the quarks into three generations; only the particles of the first generation have a place in the structure of ordinary matter. The *t* quark has not been observed experimentally.



THREE FORCES OF NATURE bring about interactions between the elementary particles. Each such interaction can be described as the exchange of a "virtual" particle, which is the carrier of the force. In an electromagnetic interaction particles with electric charge exchange a photon (γ). The strong force is conveyed by gluons (G), which are exchanged by particles with color charge. Particles with weak charge can exchange a W^- (shown here) or a W^+ or a Z^0 . In the diagrams the charge of an antiparticle is given by an arrow pointing backward in time.

charge of $+2/3$. The antiquarks, which are labeled \bar{d} , \bar{u} and so on, have opposite values of electric charge; hence the \bar{d} antiquark has a charge of $+1/3$ and the \bar{u} antiquark has a charge of $-2/3$. The antiquarks also have opposite colors, namely antired, antigreen and antiblue.

Quarks can be combined in two ways to form a hadron. Either three quarks are bound together, with one quark of each color, or one quark of a given color is joined to an antiquark of the corresponding anticolor. These combinations are said to be colorless, and it turns out they have another significant trait as well. In all the allowed combinations the fractional electric charges of the quarks add to yield an integral total charge; no other combinations (except multiples of the allowed ones) have this property. The proton has the quark composition uud , giving it a total electric charge of $2/3 + 2/3 - 1/3$, or $+1$. The neutron consists of the quarks udd , with charges of $2/3 - 1/3 - 1/3$, for a net charge of zero. The positive pi meson is made up of a u quark and a \bar{d} antiquark; the component charges of $+2/3$ and $+1/3$ give a total of $+1$.

The fact that all atoms are electrically neutral implies that the charge of the proton has exactly the same magnitude as that of the electron, although of course the signs are opposite. For the same reason the charge of the neutron must be exactly zero. It follows from these observations that the charges of the quarks must be exactly commensurable with those of the leptons. For example, the charge of the d quark must be exactly one-third that of the electron and not merely approximately so. This precise relation among particles that appear to be independent is another seemingly arbitrary property one would like to see explained in a unified theory.

It has become customary to classify the leptons and the quarks in three generations. Each generation is made up of a charged lepton, its associated neutrino and two quarks, one quark with charge $-1/3$ and one with charge $+2/3$. The first generation consists of the electron, the electron-type neutrino, the d quark and the u quark. Because the quarks exist in three colors there are eight particles in the generation. All atoms and all ordinary matter can be assembled from these eight particles; the higher generations are observed almost exclusively in laboratory experiments with accelerated particles. In the unified theory the three generations are described independently but in essentially the same way. I shall therefore discuss only the first generation.

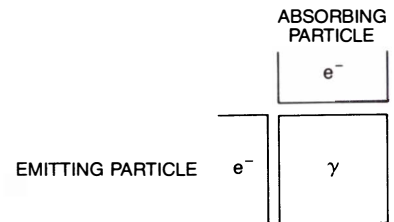
Of the three forces that I shall consider here electromagnetism was the first to be described by an accurate theory; indeed, the accuracy of the theory has never been surpassed. The the-

ory¹ is quantum electrodynamics, or QED, and it was developed over a period of about 25 years culminating in the early 1950's. It has served as a model for theories of the other forces.

The concept of a force is closely connected with that of a charge. Electric charge is the property attributed to a particle that responds to the electromagnetic force, and the amount of the charge determines the response. When two charged particles approach each other, an attraction or a repulsion is set up whose magnitude is directly proportional to the product of the charges. The force is also inversely proportional to the square of the distance between the charges. These two rules constitute Coulomb's law of the electric force. It is important to note that if one of the particles has zero charge, there is no attraction or repulsion; such neutral particles are not directly susceptible to the electromagnetic force.

How strong is the electromagnetic interaction between charged particles? For any given particles the answer depends on the charges and on their distance of separation, but Coulomb's law can provide an absolute answer. Suppose the force between two particles is multiplied by the square of the distance between them; the product is a measure of the strength of the electromagnetic interaction that is independent of the particles' separation, although it does depend on the system of units in which the separation is expressed. Dividing by the speed of light and by Planck's constant (two quantities that are built into the structure of the relativistic quantum-mechanical world) yields a result that is independent of units. The result is a pure or dimensionless number; it has the same value whether measurements are made in grams, centimeters and seconds or in tons, feet and years, provided only that the speed of light and Planck's constant are expressed in the same units as the measurements.

The strength of any given electromagnetic interaction evidently still depends on the size of the participating charges. Thus the interaction would be four times as strong if each charge were doubled. Since electric charge is quantized, however, the interaction of two protons or two electrons has a special role. All particles observed in isolation (that is, all particles except the quarks) have charges that are integral multiples of the proton's charge, so that the proton-proton interaction is a measure of the minimum strength of the electromagnetic interaction. This quantity is called the electromagnetic coupling constant, and it constitutes an absolute measure of the strength of the interaction. Experimental determinations of the coupling constant yield a value of about $1/137$. Because the value is less than 1 the electromagnetic interaction is fairly weak.



ELECTROMAGNETISM is described by a symmetry designated $U(1)$, a term borrowed from the mathematical theory of groups. $U(1)$ is the group of transformations that can be carried out on a single object or on a one-by-one matrix. In its application to electromagnetism the $U(1)$ symmetry signifies that the electromagnetic force cannot change the identity of a particle. The one-by-one matrix is occupied by the photon, which can only transform an electron into another electron.

It should be emphasized that the quantization of charge is not required by or predicted by quantum electrodynamics; rather it is an experimental fact. The theory would work as well if there were observable particles with fractional charges or even with irrational quantities of charge, such as pi or the square root of 2.

In quantum electrodynamics the interaction of two charged particles, such as two electrons, is related to the exchange of a third particle. The intermediary particle is a photon: a quantum of electromagnetic radiation. The photon is a massless particle that has no electric charge of its own and that moves (by definition) with the speed of light. Describing the electromagnetic force as an exchange of photons avoids the trou-

	RED	GREEN	BLUE
RED	$G_1 + G_2$	$G_{R \rightarrow G}$	$G_{R \rightarrow B}$
GREEN	$G_{G \rightarrow R}$	$G_1 + G_2$	$G_{G \rightarrow B}$
BLUE	$G_{B \rightarrow R}$	$G_{B \rightarrow G}$	$G_1 + G_2$

STRONG FORCE is described by a theory with an $SU(3)$ symmetry, in which the couplings of gluons to quarks can be represented by a three-by-three matrix. Any quark color in the column at the left edge of the matrix can be transformed into any of the colors in the row at the top of the matrix; the transition is mediated by the gluon specified at the intersection of the column and the row. A red quark can emit a $G_{R \rightarrow B}$ gluon to become a blue quark. Two gluons do not alter color but mediate transformations such as red to red.

COLOR CHARGES

		R-G	G-B	B-R	
QUARKS	RED	+1/2	0	-1/2	= 0
	GREEN	-1/2	+1/2	0	= 0
	BLUE	0	-1/2	+1/2	= 0
		= 0	= 0	= 0	
GLUONS	G ₁	0	0	0	= 0
	G ₂	0	0	0	= 0
	G _{R-G}	+1	-1/2	-1/2	= 0
	G _{G-R}	-1	+1/2	+1/2	= 0
	G _{G-B}	-1/2	+1	-1/2	= 0
	G _{B-G}	+1/2	-1	+1/2	= 0
	G _{R-B}	+1/2	+1/2	-1	= 0
	G _{B-R}	-1/2	-1/2	+1	= 0
	= 0	= 0	= 0		

COLOR CHARGES OF QUARKS AND GLUONS can be identified as red minus green (R - G), green minus blue (G - B) and blue minus red (B - R). Each of the quark colors red, green and blue is defined by a combination of the three charges. It is significant that the charges contributing to each color add up to zero. This pattern implies that the three charges are not fully independent, and indeed any two suffice to identify the color of a particle. (Here all three charges are retained for clarity.) For a triplet of quarks made up of one quark in each color the sums of the values of each charge are also zero. Six of the gluons have color charges, with just the values needed to convert a quark from one color into another. The pattern of charges in the quark triplet and the presence of charges in the gluons together demand the quantization of color charge: the only possible values of color charge are discrete multiples of 1/2 unit.

		R-G	G-B	B-R			R-G	G-B	B-R
ANTIQUARK ASSEMBLY	GREEN	-1/2	+1/2	0	GLUON ASSEMBLY	RED	+1/2	0	-1/2
	+ BLUE	0	-1/2	+1/2		+ ANTIGREEN	+1/2	-1/2	0
	ANTIRED	-1/2	0	+1/2		G _{R-G}	+1	-1/2	-1/2
	+ RED	+1/2	0	-1/2		RED	+1/2	0	-1/2
	+ BLUE	0	-1/2	+1/2		RED	+1/2	0	-1/2
	ANTIGREEN	+1/2	-1/2	0		+ BLUE	0	-1/2	+1/2
RED	+1/2	0	-1/2	G _{R-G}	+1	-1/2	-1/2		
+ GREEN	-1/2	+1/2	0						
ANTIBLUE	0	+1/2	-1/2						

ASSEMBLY PROCEDURE for antiquarks and gluons correctly predicts their color properties by supposing they are combinations of quarks drawn from the fundamental color triplet. Any antiquark (which must have color charges opposite to those of the corresponding quark) can be "built" by adding the color charges of two quarks. Any gluon can be assembled from the colors of a quark and an antiquark or, since the antiquark itself can be decomposed, from the colors of three quarks. All possible combinations are accounted for. The procedure is a formal one only; antiquarks and gluons should not be considered physical composites of quarks.

blesome notion of action at a distance. The interaction is confined to two point-events: the emission and the absorption of the photon. At the same time, however, the description introduces another equally serious problem: the exchange of the photon seems to violate the laws of nature that require energy and momentum to be conserved.

The apparent violation can be illustrated by imagining two electrons held stationary a short distance apart. Since a force could be measured between the electrons, it must be assumed that photons are being exchanged. Ordinarily when a photon is emitted, it carries away part of the energy and momentum of the emitting particle; similarly, when a photon is absorbed, it adds to the energy and momentum of the absorbing particle. In this way the total quantity of energy and momentum in the system is conserved. In the situation being considered here, however, the emitting particle is held stationary, so that its energy and momentum cannot change, and the same is true of the absorbing particle. Evidently the exchanged photon has rather special properties, different from those of the photons that make up sunlight or radio waves. In recognition of the distinction the exchanged photon is called a virtual photon.

The explanation of the peculiar properties of the virtual photon lies in the uncertainty principle introduced into quantum mechanics by Werner Heisenberg. The uncertainty principle does not invalidate the conservation laws of energy and of momentum, but it does allow a violation of the laws to go unnoticed if it is rectified quickly enough. The stationary electrons have the same energy and momentum before the virtual photon is emitted and after it is absorbed; the conservation laws seem to be violated only during the brief passage of the photon. The uncertainty principle states that such an apparent violation can be tolerated if it does not last too long or extend too far.

How long is too long and how far is too far? The answers depend on the magnitude of the apparent violation: the greater the imbalance in energy and momentum caused by the emission of the virtual photon is, the sooner the photon must be reabsorbed. A high-energy virtual photon can survive only briefly, whereas a low-energy one has a long grace period before the books must be balanced. To be explicit, the product of the energy imbalance and the lifetime of the photon cannot exceed Planck's constant. The minimum energy any particle can have is the energy equivalent of the particle's rest mass, and so the maximum range of a virtual particle depends inversely on its mass. The range of the electromagnetic force seems to be infinite, and so the rest mass of the photon is thought to be exactly zero.

The presence of virtual particles great-

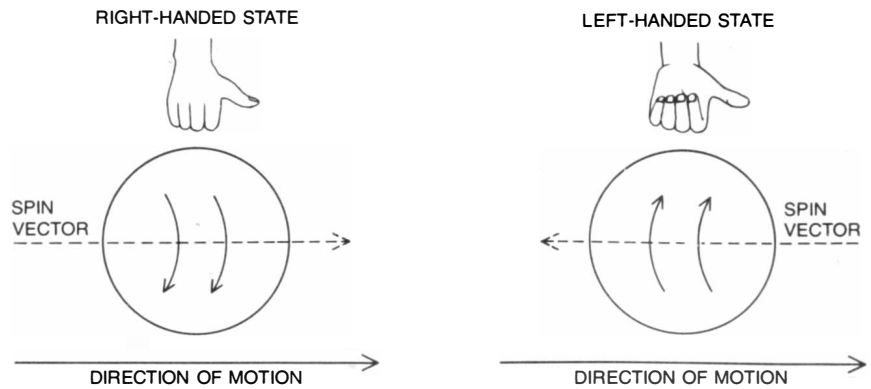
ly complicates the structure of the universe. Because of the virtual particles the vacuum is not mere empty space. A virtual photon can appear spontaneously at any instant and disappear again within the time allotted by the uncertainty principle. Other virtual particles can be created in the same way, including electrically charged particles; the only constraint is that particles with an electric charge must appear and disappear in matched pairs of particle and antiparticle. This process has profound consequences for the theory of electromagnetism.

Consider what happens when a real electron is embedded in a cloud of virtual photons and virtual electron-positron pairs. The photons have little effect, but the charged virtual particles become polarized: negative virtual charges are repelled by the real negative charge, whereas positive virtual charges are attracted to the real electron. As a result the electron becomes surrounded at close range by a cloud of positive charges, which screen or shield part of the electron's charge.

It follows from this analysis that the "bare" charge of the electron is much greater than the measured charge. Indeed, in quantum electrodynamics the bare charge is assumed to be infinite. The measured charge is merely the finite residue that remains when the shielding charge is subtracted from the bare charge. If the electron's charge could be measured at extremely close range, it would be found to increase as the screening layer was penetrated. A further consequence is that the coupling constant of electromagnetism is not a constant at all but varies with the distance at which charged particles interact with one another. The coupling constant increases (signifying that the electromagnetic interaction becomes stronger) as the range is reduced. The measured coupling constant of about $1/137$ is observed at atomic distances of roughly 10^{-8} centimeter.

Even in the ephemeral realm of the virtual particles there is one conservation law that is never violated: the conservation of electric charge. Because the photon itself is neutral, charge is automatically conserved in the exchange of a virtual photon; no charges are altered. Moreover, when charged matter is created or annihilated, it is always in pairs of particles and antiparticles, so that the sum of the charges after the event remains the same as it was before it.

The conservation of electric charge and the masslessness of the photon are related to a group of symmetries in the mathematical system that describes quantum electrodynamics. The group of symmetries is designated $U(1)$, and so QED is a $U(1)$ theory. $U(1)$ is a term employed in the mathematical theory of groups. The 1 refers to the fact that



HANDEDNESS OF A PARTICLE is determined by the orientation of its intrinsic spin angular momentum. When the vector that defines the spin axis is parallel to the particle's direction of motion, the particle is said to be right-handed; the reason is that when the fingers of the right hand are curled in the way the particle rotates, the thumb indicates the direction of motion. When the spin vector is antiparallel to the trajectory, the thumb of the left hand gives the direction of motion, and so the particle is said to be left-handed. Electromagnetism and the strong force are indifferent to handedness, but handedness has an important influence on the weak interactions. A massive particle can change its handedness but a massless one cannot.

the photon interacts with only one kind of particle at a time. The photon never transforms one kind of particle into another kind. The strong force and the weak force are more complicated in this respect, and the groups that describe them are more complex.

The prevailing theory of the strong force is modeled directly on quantum electrodynamics. The theory is called quantum chromodynamics, or QCD, "chromo-" signifying that the force acts not between electric charges but between color charges. As in QED the magnitude of the force between two charges is proportional to the product of the charges. Particles that have no color charge are not subject to the force. A dimensionless coupling constant defines the intrinsic strength of the interaction. The coupling constant is larger than the constant of electromagnetism, as might be expected of a force that is named strong.

Although QCD is constructed on the same principles as QED, it is a more elaborate theory. The main source of the added complexity is the multiplicity of color charges. Whereas electromagnetism is associated with just one kind of charge, the strong force acts on three colors: red, green and blue. Each color represents a combination of underlying color charges.

There are several ways of analyzing the color charges. The way I shall adopt begins by supposing there are three kinds of color charge. I shall call the charges red minus green ($R - G$), green minus blue ($G - B$) and blue minus red ($B - R$). Each charge can have a value of $+1/2$, $-1/2$ or 0, and each color of quark is distinguished by a particular combination of values. A quark is red if it has an $R - G$ charge of $+1/2$, a $G - B$ charge of 0 and a $B - R$ charge of $-1/2$. A green quark has the color

charges $R - G = -1/2$, $G - B = +1/2$ and $B - R = 0$. In a blue quark the three charges are $R - G = 0$, $G - B = -1/2$ and $B - R = +1/2$. The anticolors associated with the antiquarks are formed simply by reversing the signs of all the charges.

Several observations can be made about this distribution of charges. First, there are 27 possible combinations of three charges when each charge can have any one of three values. Nevertheless, only quarks with the three combinations that yield the colors red, green and blue seem to exist in nature. Second, this subset of the possible color charges is a highly distinctive one. Each of the observed combinations is arranged so that the sum of the three charges is zero, and the observed combinations are the only ones having this property. (Actually there is one other combination with a total color charge of zero: the combination in which each charge is zero. A particle that has no color charge at all, however, is not a quark.)

The finding that the sum of the three color charges is always equal to zero indicates that one of the three charges is not independent of the other two. If any two of the charges are known, the third can be found by subtraction. Hence it can be concluded that there are really only two varieties of color charge, which are sufficient to specify the three colors completely. It makes no difference which two charges are considered fundamental and which one is discarded; here I shall assume that the $R - G$ and $G - B$ charges are fundamental, but I shall often retain the $B - R$ charge for clarity, even though the information it provides is redundant.

A further relation among the charges can be noted. In a state made up of one red, one green and one blue quark the total quantity of each color charge is again zero. In other words, combining

the three colors gives rise to a color-neutral state, much as combining an electron and a proton creates a state (the hydrogen atom) that is neutral with respect to electric charge. It is in just this way that color-neutral hadrons such as the proton are formed. A colorless system is also created by combining a color with the corresponding anticolor; since the color charges are then opposite, they cancel exactly. Linking a color with its anticolor is the other formula for making a hadron, on the model of the pion. Except for multiples of these combinations (such as a six-quark system that includes two quarks of each color) there is no other way of combin-

ing colored quarks so that all the color charges have a sum of zero.

The mechanism by which the strong force is transmitted is comparable to the corresponding mechanism in electromagnetism: the interaction between two charged particles is described as the exchange of a third particle. Again, however, quantum chromodynamics is a more elaborate theory. Whereas QED has a single massless photon, QCD has eight massless particles called gluons. Furthermore, whereas the photon has no electric charge, some of the gluons do carry color charge. The presence of charged carrier particles fundamentally alters the character of the force.

Since the gluons are themselves charged, they not only can transmit the strong force but also can alter the colors of quarks. The emission or absorption of photons, in contrast, can never alter the electric charge of a particle. There are nine possible transitions among the quark colors, defined by a three-by-three matrix. For example, a red quark can be transformed into a red quark (the identity transformation), into a green quark or into a blue quark. The three identity transformations (red becomes red, green becomes green and blue becomes blue) make up the diagonal elements of the matrix. It is apparent that the gluons responsible for the identity transitions cannot have color charge or they would alter the quark colors. It might seem there should be three such color-neutral gluons, one neutral gluon for each identity transformation. Since just two independent color charges are needed to specify the three quark colors, however, there are only two color-neutral gluons. I shall designate them G_1 and G_2 .

The six remaining transitions between the quark colors do entail changes of color. Each of these transitions is associated with its own gluon, and each of these six gluons does bear a color charge. I shall label the color-charged gluons descriptively. For example, a red-to-green gluon, or $G_{R \rightarrow G}$, can be emitted by a red quark, which is thereby transformed into a green quark.

The color charges carried by the gluons can be deduced from the requirement that color charge be conserved. Consider the process in which a quark changes from red to green by the emission of a $G_{R \rightarrow G}$ gluon. In the course of the transition the R - G charge of the quark changes from $+1/2$ to $-1/2$; if the total quantity of charge is to remain unchanged, the gluon must therefore have an R - G charge of $+1$. In the same way the G - B charge of the quark goes from 0 to $+1/2$, so that the gluon is required to carry off a G - B charge of $-1/2$. The B - R charge of the quark is transformed from $-1/2$ to 0, which again implies a gluon B - R charge of $-1/2$. Hence the gluon color charges are respectively $+1$, $-1/2$ and $-1/2$. The quark that mediates the reverse transformation, from green to red, must have charges of the same magnitude but opposite sign.

The presence of gluon color charges has a further implication: it automatically ensures that the color charge is quantized. In electromagnetism a photon could in principle be emitted or absorbed by a particle with any quantity of electric charge. Particles with color charge, however, can interact by exchanging gluons only if the charges are separated by intervals that are multiples

	WEAK CHARGE	U(1) CHARGE	ELECTRIC CHARGE	PARTICLES	TRANSITIONS
DOUBLETS	$+1/2$		0	$\uparrow \nu_{\text{LEFT}}$	$W^+ \rightleftarrows W^-$
	$-1/2$	$-1/2$		$\uparrow e^-_{\text{LEFT}}$	
	$+1/2$		$+2/3$	$\uparrow u_{\text{LEFT}}$	$W^+ \rightleftarrows W^-$
	$-1/2$	$+1/6$	$-1/3$	$\uparrow d_{\text{LEFT}}$	
	$+1/2$		$+1$	$\downarrow e^+_{\text{RIGHT}}$	$W^+ \rightleftarrows W^-$
	$-1/2$	$+1/2$	0	$\downarrow \bar{\nu}_{\text{RIGHT}}$	
SINGLETS	$+1/2$		$+1/3$	$\downarrow \bar{d}_{\text{RIGHT}}$	$W^+ \rightleftarrows W^-$
	$-1/2$	$-1/6$	$-2/3$	$\downarrow \bar{u}_{\text{RIGHT}}$	
	0	-1	-1	$\uparrow e^-_{\text{RIGHT}}$	
	0	$+1$	$+1$	$\downarrow e^+_{\text{LEFT}}$	
	0	$+2/3$	$+2/3$	$\uparrow u_{\text{RIGHT}}$	
	0	$-2/3$	$-2/3$	$\downarrow \bar{u}_{\text{LEFT}}$	
0	$-1/3$	$-1/3$	$\uparrow d_{\text{RIGHT}}$		
0	$+1/3$	$+1/3$	$\downarrow \bar{d}_{\text{LEFT}}$		

WEAK CHARGE depends on the handedness of a particle and also has a curious relation to electric charge. Left-handed particles and right-handed antiparticles form doublets in weak interactions and are assigned weak charges of plus or minus $1/2$. The W^+ and W^- , which mediate the weak force, transform one member of a doublet into the other member. Right-handed particles and left-handed antiparticles remain singlets and have no weak charge, so that no weak transitions among them are possible. The electric charge of each particle is the sum of the weak charge and another quantity called the U(1) charge, which is equal to the average electric charge of the particles in the singlet or the doublet. This relation of the charges indicates that there is an underlying connection between the weak force and electromagnetism.

of $1/2$. It can also be shown that the color charges of the system must be symmetrical about zero, that is, the total of all the positive charges and the total of all the negative charges must be equal in absolute value.

The quantization of the color charge can be demonstrated in another way. Any system of particles with color can be "built" out of the simplest such system: the triplet made up of a red, a green and a blue quark. The triplet of antiquarks can be formed by combining the quarks in pairs. I do not mean to suggest that a physical antiquark consists of two quarks in a bound state. Nevertheless, all the color properties of the antiquark are correctly given by such a synthesis. Note that a red quark has $R - G$, $G - B$ and $B - R$ color charges of $+1/2$, 0 and $-1/2$ respectively. An antired antiquark must have the opposite charges: $-1/2$, 0 and $+1/2$. These are exactly the values found by adding the charges of a green quark ($-1/2$, $+1/2$ and 0) and a blue quark (0 , $-1/2$ and $+1/2$). Hence antired is in some sense equivalent to the sum of green and blue. Likewise antigreen consists of red plus blue and antiblue consists of red plus green. This remarkable correspondence is a simple consequence of the way the color charges are distributed in the quark triplet. Because the total charge of the triplet is zero, the sum of any two charges must equal the negative of the remaining charge.

The gluons can be constructed in a similar way from a quark and an antiquark, although again no physical building process should be inferred. The red-to-green gluon, with charges of $+1$, $-1/2$ and $-1/2$, can be imagined as a compound of a red quark ($+1/2$, 0 and $-1/2$) and an antigreen antiquark ($+1/2$, $-1/2$ and 0). Of course, the antigreen antiquark can be further decomposed into a red quark and a blue quark, so that the red-to-green gluon has the color properties of two red quarks and a blue quark.

There is yet another consequence of giving the gluons color charges. As explained above, an electron in a vacuum becomes surrounded by a cloud of virtual photons and virtual electron-positron pairs; the charged virtual particles become polarized and shield a part of the electron's bare charge. By the same mechanism a quark in a vacuum becomes enveloped in a cloud of virtual gluons and virtual quark-antiquark pairs, but the result is quite different. The cloud of virtual quarks and antiquarks is polarized in the usual way, with antiquarks clustered near the real color charge and tending to screen it. The virtual gluons, however, have the opposite effect. The predominant color charge of the gluons near the quark is

the same as the charge of the quark itself. Moreover, the virtual gluons are more numerous than the virtual quarks, so that the influence of the gluons is the stronger one. The result is that the charge of the quark appears to be spread out in space, and the effective charge diminishes as the quark is approached more closely.

In the absence of gluon charges the strong force might be expected to vary with distance in much the way electromagnetism does. Since the gluons are massless, like the photon, the force would have an infinite range but would decrease in intensity as the square of the distance. The fact that the gluons carry color charges alters the character of the force. Because the cloud of virtual gluons spreads out the color charge, the color force between two quarks does not increase as fast as the electromagnetic force does when the distance between the particles is reduced. As a result the coupling constant of QCD decreases as the distance at which it is measured gets smaller (unlike the coupling constant of QED, which increases at close range). The quarks are said to be asymptotically free, meaning that the QCD coupling constant falls to zero as the distance approaches zero.

Asymptotic freedom was discovered by H. David Politzer, who is now working at the California Institute of Technology, and by David Gross and Frank Wilczek of Princeton University. It has been tested and confirmed in a number of experiments that probe the quark structure of hadrons at small distances. The nature of the strong interactions between quarks at longer distances is not as well established, but it seems the force may not decrease as the square of the distance but may instead remain constant regardless of distance. Unlimited energy would then be needed to separate two color charges, which would explain why quarks seem to be confined permanently in hadrons.

Quantum chromodynamics is called an $SU(3)$ theory, where $SU(3)$ is another term from the theory of groups. The 3 refers to the three colors that are transformed into one another by the gluons. The S indicates that the sum of the color charges in each $SU(3)$ family is zero. In a close analogy with the $U(1)$ of QED, the $SU(3)$ of QCD describes a group of symmetries of the theory that is associated with the conservation of color charge and the masslessness of the gluons.

The high order of symmetry in the color- $SU(3)$ theory can be suggested in a geometric presentation. The three color charges $R - G$, $G - B$ and $B - R$ can be represented by position along three axes on a plane. The axes are symmetrically arrayed at angles of 120 degrees to one another. If the three colors are placed according to their component

color charges on such a graph, they are found to lie at the vertexes of an equilateral triangle. The anticolors lie opposite the corresponding colors, so that they form another triangle turned 180 degrees from the first one. The two superposed triangles form a Star of David.

A hint of the still greater symmetry in the unified theory is provided by adding a third dimension to the graph. Suppose the color charges already plotted are assumed to be those of the u quark and the \bar{d} antiquark. Two leptons are now added: the electron-type neutrino and the positron. Since the leptons have no color charge, they lie at the origin of the three axes in the center of the plane. The third dimension is electric charge: each particle is to be displaced vertically by an amount proportional to its electric charge. The neutrino remains in its place, but the three \bar{d} antiquarks are shifted upward one-third of a unit, the u quarks move upward two-thirds of a unit and the positron moves upward one unit. If the vertical and horizontal scales are suitably chosen, the eight particles define the vertexes of a cube tipped on one corner. That quarks and leptons can be arranged in the configuration of this simple solid suggests some underlying connection between them.

In order to discuss the last of the three forces, the weak force, it is necessary to introduce another property of the elementary particles: spin angular momentum. All the leptons and all the quarks are observed to have the same fixed quantity of angular momentum, equal to $1/2$ when measured in fundamental units. The particles can be imagined as spinning on an internal axis, like the earth or like a top, but without loss of energy. The angular momentum is represented by a vector, or arrow, along the axis of spin.

A particle with one-half quantum of intrinsic spin can have only two possible orientations; in the simplest case, where the particle is in motion, the spin vector can point either in the same direction the particle is moving or in the opposite direction. The two orientations represent two distinguishable states of the particle. The state in which the vector is parallel to the direction of motion is said to be right-handed because when the fingers of the right hand are wrapped around the particle in the same sense as the spin, the thumb indicates the direction of motion. When the spin axis is aligned the opposite way, the thumb of the left hand indicates the direction of motion and so the particle is said to be left-handed.

In general the handedness of a particle can be reversed simply by bringing the particle to rest and accelerating it in the opposite direction without disturbing the spin. Hence most particles necessar-

ily have both left- and right-handed components. The exceptions are massless particles, and the reason they are exceptions is that a massless particle must always move with the speed of light and can never be brought to rest. Therefore the handedness of a massless particle can never change. Among the quarks and leptons the only particles that might be massless are the neutrinos. Experimentally only left-handed neutrinos and right-handed antineutrinos have been observed; right-handed neutrinos and left-handed antineutrinos are presumed not to exist.

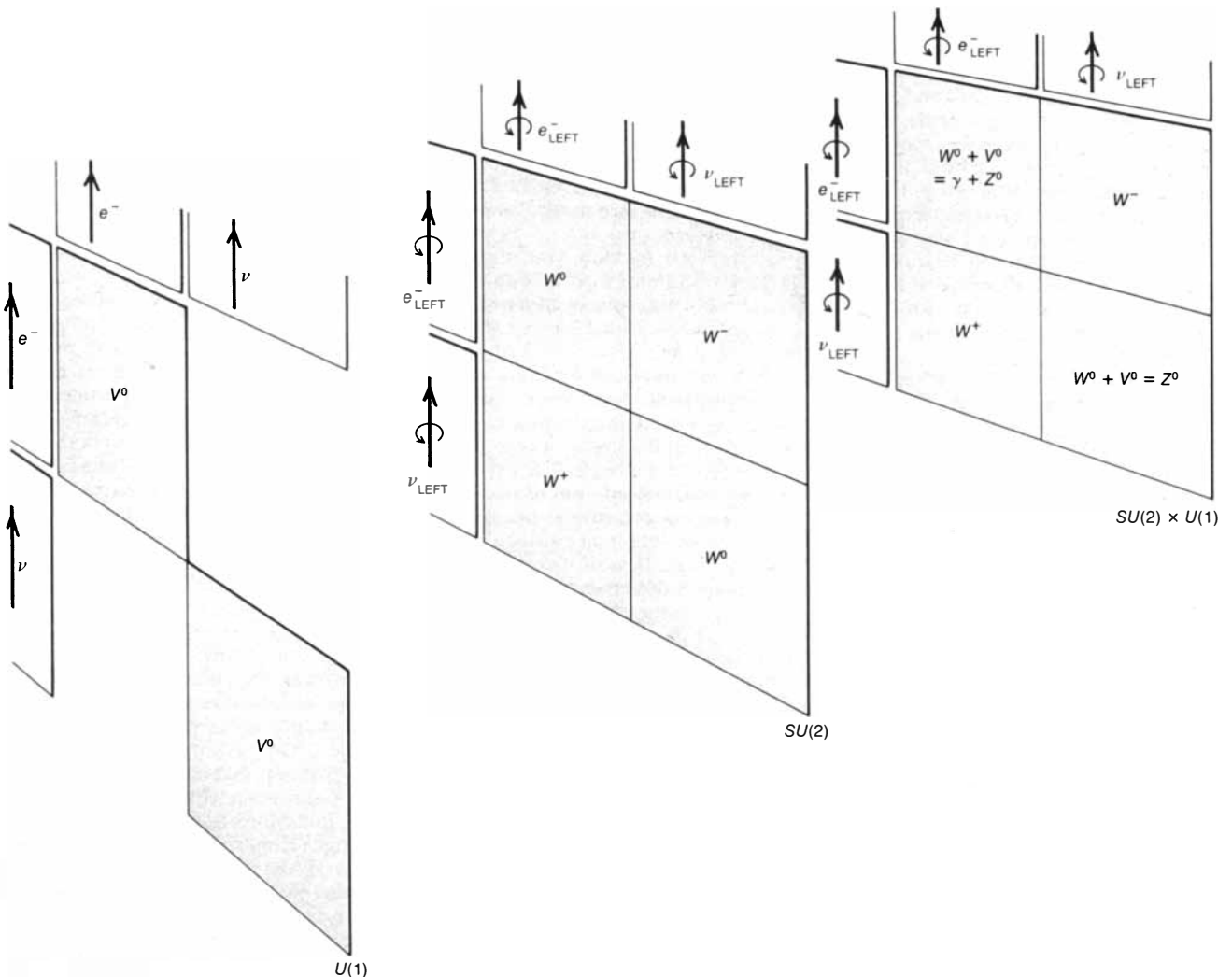
The introduction of handedness has the effect of almost doubling the number of distinguishable elementary particles, a number that is already fairly large. In the first generation of particles

there are two leptons (the electron and the electron-type neutrino) and two quark flavors (u and d). The three quark colors give a total of eight particles, and accounting for the corresponding eight antiparticles raises the total to 16. If all the particles had both left- and right-handed components, the inclusion of handedness would double the number yet again. Actually, because there is no right-handed neutrino or left-handed antineutrino, the total number of distinct particle and antiparticle states is 30. It is these 30 states that must be accommodated in a unified theory.

States of different handedness must be distinguished because the weak force acts differently on the left-handed and on the right-handed components of a particle. Like the other forces, the weak

force is associated with a charge, and the intrinsic strength of the weak interaction can be defined by means of a dimensionless coupling constant. The weak charge is unusual, however, in that it is assigned on the basis of handedness. Only left-handed particles and right-handed antiparticles bear a weak charge; the right-handed particles and the left-handed antiparticles are neutral with respect to the weak force and do not participate in weak interactions.

Because the weak charges of the left- and the right-handed electron (for example) are different, the weak charge cannot be conserved. The value of the weak charge depends on which way the electron is moving, and the value must change when the motion changes. The weak charge could be conserved only if



WEAK FORCE AND ELECTROMAGNETISM can be described jointly by a theory with a symmetry that is represented by the product of two groups: $SU(2) \times U(1)$. The $SU(2)$ part of the interaction induces all possible transformations of two objects or of a two-by-two matrix. The objects are the members of the weak doublets, represented here by the left-handed components of the electron and the neutrino. There are three $SU(2)$ particles, all of them massless: the W^+ and the W^- convert an electron into a neutrino and vice versa and

the W^0 mediates the identity operations electron to electron and neutrino to neutrino. The $U(1)$ part of the interaction is associated with another intermediary particle, the V^0 , which is capable only of identity transformations. In the combined $SU(2) \times U(1)$ theory the W^0 and the V^0 become mixed; the observed combinations are the photon (γ) and a carrier of the weak force called the Z^0 . The $SU(2) \times U(1)$ symmetry is spontaneously broken, which has the effect of giving the W^+ , W^- and Z^0 large masses while leaving the photon without mass.

the leptons and the quarks were all massless, since in that case none of the particles could come to a stop and reverse direction.

The weak force acts on doublets of particles. The theory that describes it is an $SU(2)$ theory, in which the two members of a doublet can be transformed into each other. For example, the left-handed neutrino and the left-handed electron make up one doublet; they are assigned weak charges of $+1/2$ and $-1/2$ respectively. The left-handed u quark and the left-handed d quark compose a second doublet (or three doublets if each color is counted separately), and they also have respective weak charges of $+1/2$ and $-1/2$. Four right-handed antiparticles form the remaining doublets: the positron, the electron-type antineutrino, the \bar{d} antiquark and the \bar{u} antiquark. Each right-handed antiparticle has a weak charge opposite to that of the corresponding left-handed particle. Six particles remain to be accounted for: the right-handed components of the electron, the d quark and the u quark and the left-handed components of the positron, the \bar{d} antiquark and the \bar{u} antiquark. They do not form doublets but remain isolated as singlets, and they have a weak charge of zero.

Three particles associated with the weak $SU(2)$ symmetry mediate transitions between the members of each doublet. The intermediary particles are the W^+ , with both a weak charge and an electric charge of $+1$; the W^- , with weak and electric charges of -1 , and the W^0 , which is neutral with respect to both the weak and the electromagnetic forces. The W^0 , like the photon and the G_1 and G_2 gluons, conveys a force between particles that carry charge, but it does not alter any of their properties. The W^+ and the W^- , on the other hand, transform the flavors of particles. A left-handed electron can emit a W^- and thereby be converted into a left-handed neutrino; in the process the electric charge changes from -1 to 0 and the weak charge goes from $-1/2$ to $+1/2$. The most familiar weak process is nuclear beta decay, in which a neutron (quark composition udd) emits an electron and an antineutrino and is converted into a proton (uud). Analyzed at a finer scale, the event begins when a d quark emits a virtual W^- and becomes a u quark; the W^- subsequently decays to yield the electron and the antineutrino.

In events such as these one can perceive some tantalizing relations between the weak force and electromagnetism. First, the W particles of the weak force carry weak charge and electric charge in the same amount. Second, in the structure of the weak singlets and doublets there is a curious fixed relation between weak charge and electric charge. The

electric charge of a particle is invariably equal to the sum of the particle's weak charge and the average electric charge of the singlet or doublet of which the particle is a member. This average charge I shall designate the $U(1)$ charge. For the singlets the $U(1)$ charge is merely the electric charge of the particle itself and the rule is little more than a tautology: it states that the electric charge is equal to the electric charge, since the weak charge of a singlet particle is always zero. For the weak doublets, however, the relation of the charges is more interesting. It is noteworthy that the relation remains valid both for the lepton doublets, where the charges being averaged are integers, and for the quark doublets, which are made up of fractional charges.

Like the other charges, the $U(1)$ charge is associated with a symmetry. The $U(1)$ symmetry that is extracted from the weak interactions in this way has a particle associated with it, which I shall call V^0 . Like the W^0 and the photon, the V^0 has neither electric nor weak charge. Indeed, because the weak, the $U(1)$ and the electric charges are related, the three carrier particles must also be related. It turns out that the W^0 and the V^0 are not observed in nature as pure states but instead appear only as mixtures. One such mixture of the W^0 and the V^0 is the photon; the other possible combination is identified as a particle labeled Z^0 . Both the photon and the Z^0 mediate interactions in which a particle is transformed into itself, so that its identity is unchanged. They differ, however, in that the photon couples only particles that have electric charge, whereas the Z^0 couples those that have weak charge, including the neutrinos. By deriving the photon and the W and Z particles from the same theory, the weak force and electromagnetism are partially unified. The combined theory is designated by the product of the groups it incorporates: $SU(2) \times U(1)$.

By analogy with the $U(1)$ of quantum electrodynamics and the $SU(3)$ of quantum chromodynamics, one might expect the $SU(2)$ and the $U(1)$ of the weak interaction to be exact symmetries of the theory. The weak charge would then be exactly conserved and the W and Z particles would be massless and of infinite range. As noted above, however, weak charge is not invariably conserved; furthermore, the weak force is observed to have an exceedingly short range of perhaps 10^{-15} centimeter. The reason is that the W^+ , the W^- and the Z^0 have large masses, approaching 100 GeV, or 100 times the mass of the proton. What becomes of the $SU(2) \times U(1)$ symmetry of the weak and the electromagnetic interactions under these conditions? Moreover, if the photon and the W and Z particles are closely related, how can

the one particle remain massless if the others acquire a large mass?

Answering these questions was an essential step in formulating the combined theory of weak and electromagnetic forces. The answer that is now favored is that the underlying force is indeed symmetrical, and in some hypothetical initial state all the carriers of the weak force would be massless. What is not symmetrical is the quantum-mechanical vacuum, the milieu in which the force operates. The structure of the vacuum spontaneously breaks the $SU(2) \times U(1)$ symmetry, giving a mass to the three carriers of the weak force but not to the photon. An analogous loss of symmetry can be observed in a crystal of salt. The ions in the crystal do not favor any one direction in space over any other; the ions are rotationally symmetrical. In the lattice structure of the crystal, however, the symmetry is broken, and certain directions, such as those parallel to lattice lines, have a special status.

The same analogy can illuminate another characteristic of a spontaneously broken symmetry: The appearance of the world depends on the scale at which it is examined. In the crystal three scales of distance can be distinguished. At distances of much less than 10^{-8} centimeter (the size of an atom) one sees the internal structure of the atom, which is fully symmetrical and unaffected by the organization of the crystal. At distances of about 10^{-8} centimeter the forces responsible for the cohesion of the atoms in the crystal become important, and the phenomena observed are extremely complex. At distances much greater than 10^{-8} centimeter the geometry of the crystal becomes fully apparent and the rotational symmetry of the atoms is conspicuously broken.

A similar hierarchy of distance scales can be defined in connection with the spontaneous breakdown of the $SU(2) \times U(1)$ symmetry, but the critical distance is smaller: about 10^{-16} centimeter. At distances much smaller than this the full symmetry is expressed. At such close range the massive W and Z particles are exchanged as readily as massless photons, and so the weak and the electromagnetic forces are effectively unified. Another way of expressing the same idea is to point out a relation between distance and energy. According to the uncertainty principle, the energy needed to probe or resolve a given distance is inversely proportional to the distance. An experiment that examined the structure of a particle at a range of less than 10^{-16} centimeter would have to be done at an energy of more than 100 GeV. At this energy W and Z particles can be freely created, and the mass difference between them and the photon is negligible; all the intermediary particles

are light compared with the energy of the experiment.

At a distance of about 10^{-16} centimeter the complex phenomena responsible for breaking the $SU(2) \times U(1)$ symmetry begin to intrude. W and Z particles are still observed, but they look quite different from the photon because their masses are comparable to the energy of the experiment. At still larger distances the symmetry between the photon and the W and Z particles is completely obscured; indeed, there is insufficient energy to create a real W or Z , and so they cannot be observed directly. Only the rare, short-range effects of their virtual exchange (such as the beta decay of the neutron) can be detected. This is the present domain of particle physics.

The concept of spontaneous symmetry breaking resolves the question of weak-charge conservation. At an energy much greater than 100 GeV, where the $SU(2) \times U(1)$ symmetry is observed directly, the mass of a quark or a lepton is negligible; the handedness of the particles is therefore essentially fixed, and so the weak charge is effectively conserved. At low energy, where the symmetry is spontaneously broken, the weak charge is not conserved but can disappear into the vacuum when a massive particle changes handedness.

This theory of the weak and electromagnetic interactions was worked out in the 1960's by Sheldon Lee Glashow and

by Steven Weinberg of Harvard University and by Abdus Salam of the International Center for Theoretical Physics in Trieste. Glashow was the first to deduce the form of the $SU(2) \times U(1)$ theory, but he did not know how to incorporate into it the masses of the W and Z particles. Weinberg and Salam separately found the $SU(2) \times U(1)$ form later and applied the notion of spontaneous symmetry breaking, thereby formulating a consistent theory.

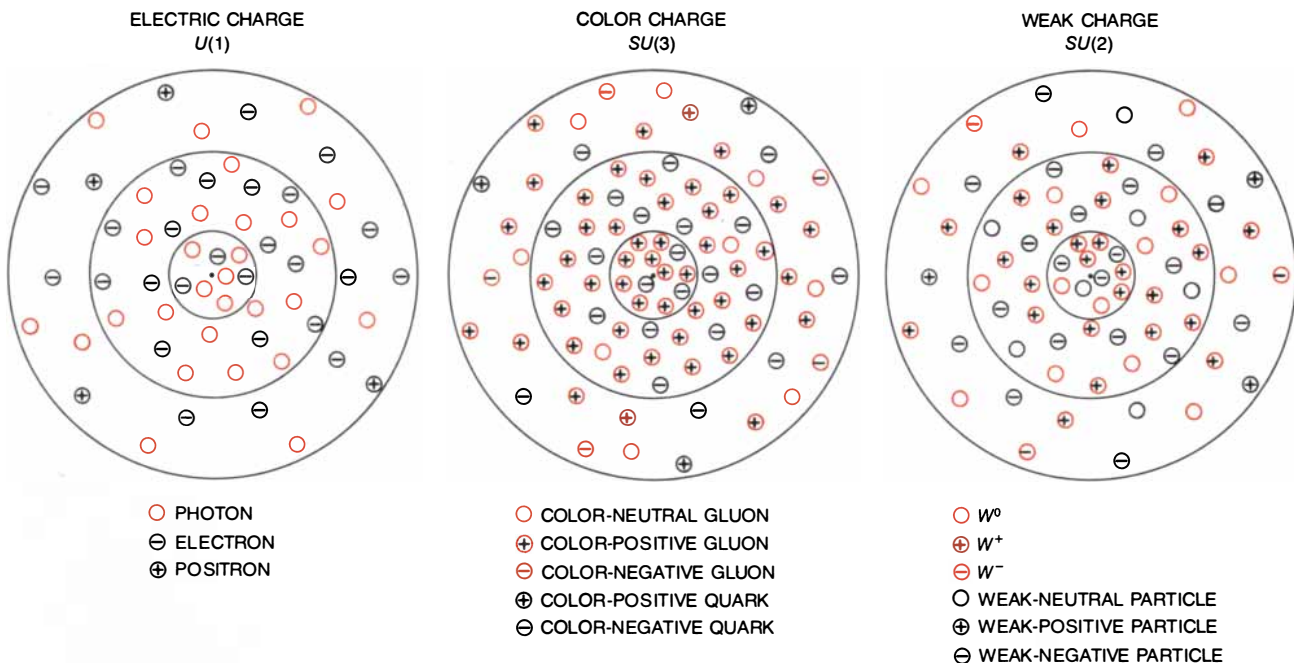
The $SU(2) \times U(1)$ theory is only a partial unification because it still includes two distinct forces, each with its own symmetry group and its own coupling constant. The ratio of the coupling constants is a free parameter, which must be chosen to fit experimental results. Another deficiency of the theory is that electric charge is only partly quantized. The construction of doublets of particles related by $SU(2)$ transformations requires that all differences between electric charges be integers, so that each particle can change its identity by emitting a W^+ or a W^- . The average charge of the doublets, however, is not quantized. The average charge is the $U(1)$ charge, which is defined in conjunction with the electromagnetic charge of quantum electrodynamics; as in that theory there is no fundamental reason for confining the charges to integer values. Indeed, in the doublets made up of quarks the interval between the

charges is an integer, but the actual charges are fractional.

The endeavor to construct a unified theory of the weak, the strong and the electromagnetic forces should not be seen as an attempt to do away with the color $SU(3)$ model or the $SU(2) \times U(1)$ model. The theories of the individual forces work too well for them to be discarded. What a unifying theory can provide is a superstructure in which the $SU(3)$ and $SU(2) \times U(1)$ theories can be embedded. The superstructure would take the form of a larger symmetry in which quarks and leptons would be closely related.

The search for such a larger symmetry must begin with the search for a larger group, one that includes both $SU(3)$ and $SU(2) \times U(1)$ as component structures. Many groups have this property, but one candidate has numerous advantages that distinguish it from all the others. This group is $SU(5)$, the group of all possible transformations of five distinct objects, or of a five-by-five matrix. It is the smallest simple group that can accommodate the constituent $SU(3)$ and $SU(2) \times U(1)$ symmetries, and I believe it may be the full symmetry group of nature. A unified theory based on $SU(5)$ symmetry was worked out in 1973 by Glashow and me.

In the simplest representation of the $SU(5)$ group the five objects are the right-handed components of the d quark



CLOUD OF VIRTUAL PARTICLES surrounds a central point charge and alters its response to a force. A positive electric charge is surrounded by virtual photons and virtual electron-positron pairs. The photons have little effect, but the charged virtual particles are polarized, so that the negative virtual charges cluster around the positive real charge, reducing its effective magnitude. A positive color charge is enveloped in a cloud of virtual gluons and virtual quark-antiquark pairs. The quarks and antiquarks become polarized much as electrons and positrons do, but the gluons act differently from pho-






tons. Whereas the photon is electrically neutral, some of the gluons have a color charge, and it is predominantly of the same polarity as the real charge. The color charge is therefore spread out in space and the net charge in any spherical volume becomes smaller as the radius of the sphere is reduced. The ultimate result of these effects is that the electromagnetic interaction becomes larger at close range, whereas the strong interaction becomes smaller. The weak interaction also has charged carrier particles, so that it too becomes feebler at small distances, although to a lesser extent than the strong interactions.

in each of the colors red, green and blue, the right-handed component of the positron and the right-handed component of the electron-type antineutrino (which has only a right-handed component). Each of the five particles is assigned a value for each of four independent charges: electric charge, weak charge and two color charges, which I shall take to be $R - G$ and $G - B$.






Twenty-four intermediary particles provide for all possible transitions between these five states of matter. Four of the particles are the photon, the Z^0 and the gluons G_1 and G_2 , which are directly associated with the four fundamental charges. Since these particles carry no charge, they take part only in those interactions where the identity of a particle does not change. Of the remaining 20 intermediary particles eight are already familiar. They are the W^+ and the W^- , which can convert a positron into an antineutrino and vice versa, and the six gluons that transform the colors of the quarks. With this complement of 12 carrier particles all interactions observed in nature up to now can be accounted for. The $SU(5)$ group includes 12 more intermediary particles, however, which are needed if the theory is to have the maximum possible symmetry. The 12 extra particles are labeled X , and they mediate the interconversion of leptons and quarks. Each X particle carries weak charge, color charge and electric charge; the electric charges have values of plus or minus $1/3$ and plus or minus $4/3$.

As with the distribution of color charges in $SU(3)$, the table of charge assignments in the $SU(5)$ theory has some intriguing regularities. For each kind of charge the sum of the charges assigned to the five particles is zero. For example, each of the three quark colors has an electric charge of $-1/3$, but these charges are balanced by the positron's electric charge of $+1$. A related observation is that all four varieties of charge are carried by at least some of the $SU(5)$ intermediary particles. The gluons have color, the W^+ and W^- have both weak charge and electric charge and the X particles carry all four forms of charge.































From these two facts it can be deduced that all the charges are necessarily quantized. All electric charges must be multiples of $1/3$; if a particle with some different charge were accepted into the family, the $SU(5)$ carrier particles could not be emitted or absorbed by it without violating the conservation of charge. Moreover, it is not just the minimum interval between charges that is fixed; the actual values of the charges are determined by the requirement that the total charge be zero. Here at last is an explanation of the quantization of electric charge. The same requirement explains the exact commensurability of the lepton and quark charges, which in turn implies the exact neutrality of the atom. In addition the intriguing coinci-

	ELECTRIC CHARGE	WEAK CHARGE	R-G CHARGE	G-B CHARGE
 $d^{\text{RED}}_{\text{RIGHT}}$	$-1/3$	0	$+1/2$	0
 $d^{\text{GREEN}}_{\text{RIGHT}}$	$-1/3$	0	$-1/2$	$+1/2$
 $d^{\text{BLUE}}_{\text{RIGHT}}$	$-1/3$	0	0	$-1/2$
 e^+_{RIGHT}	$+1$	$+1/2$	0	0
 $\bar{\nu}_{\text{RIGHT}}$	0	$-1/2$	0	0
	$= 0$	$= 0$	$= 0$	$= 0$

INTEGRATED FAMILY of elementary particles embraces both leptons and quarks. The family has five members: three right-handed quarks (the three colors of the d quark) and two right-handed antileptons (the positron and the antineutrino). Each particle is assigned values of electric charge, of weak charge and of two color charges, which are identified here as red minus green ($R - G$) and green minus blue ($G - B$). The third color charge is omitted because it is redundant. For each kind of charge the sum of the values assigned to all the particles is zero. Furthermore, in the unified theory the particles that bring about transformations within the family carry each kind of charge. It follows from these two facts that all the charges of the particles are quantized. The unified theory is associated with the symmetry group $SU(5)$.

	$d^{\text{RED}}_{\text{RIGHT}}$	$d^{\text{GREEN}}_{\text{RIGHT}}$	$d^{\text{BLUE}}_{\text{RIGHT}}$	e^+_{RIGHT}	$\bar{\nu}_{\text{RIGHT}}$
 $d^{\text{RED}}_{\text{RIGHT}}$	$G_1 + G_2 + \gamma + Z^0$	G_{R-G}	G_{R-B}	$X^{\text{RED}}_{-4/3}$	$X^{\text{RED}}_{-1/3}$
 $d^{\text{GREEN}}_{\text{RIGHT}}$	G_{G-R}	$G_1 + G_2 + \gamma + Z^0$	G_{G-B}	$X^{\text{GREEN}}_{-4/3}$	$X^{\text{GREEN}}_{-1/3}$
 $d^{\text{BLUE}}_{\text{RIGHT}}$	G_{B-R}	G_{B-G}	$G_1 + G_2 + \gamma + Z^0$	$X^{\text{BLUE}}_{-4/3}$	$X^{\text{BLUE}}_{-1/3}$
 e^+_{RIGHT}	$X^{\text{RED}}_{+4/3}$	$X^{\text{GREEN}}_{+4/3}$	$X^{\text{BLUE}}_{+4/3}$	$\gamma + Z^0$	W^+
 $\bar{\nu}_{\text{RIGHT}}$	$X^{\text{RED}}_{+1/3}$	$X^{\text{GREEN}}_{+1/3}$	$X^{\text{BLUE}}_{+1/3}$	W^-	Z^0

$SU(5)$ SYMMETRY encompasses all possible transitions between particles in the integrated family of five. The symmetries of the individual forces are incorporated into the $SU(5)$ theory as subgroups: the $SU(3)$ symmetry of the strong force is embodied in the three-by-three matrix at the upper left and the $SU(2)$ symmetry of the weak force appears in the two-by-two matrix at the lower right. The $U(1)$ symmetry associated with the Z^0 appears in the couplings of the photon (γ) and the Z^0 along the diagonal. The $SU(5)$ theory postulates 12 new intermediary particles, labeled X , which mediate transitions from a quark to a lepton or from a lepton to a quark. The interconversion of quarks and leptons is possible only in a unified theory.

		KIND OF CHARGE			
		ELECTRIC	WEAK	R-G	G-B
+	 $d_{\text{RIGHT}}^{\text{RED}}$	$-\frac{1}{3}$	0	$+\frac{1}{2}$	0
	 e_{RIGHT}^+	+1	$+\frac{1}{2}$	0	0
<hr/>					
	 $u_{\text{LEFT}}^{\text{RED}}$	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{2}$	0
+	 $d_{\text{RIGHT}}^{\text{GREEN}}$	$-\frac{1}{3}$	0	$-\frac{1}{2}$	$+\frac{1}{2}$
	 e_{RIGHT}^+	+1	$+\frac{1}{2}$	0	0
<hr/>					
	 $u_{\text{LEFT}}^{\text{GREEN}}$	$+\frac{2}{3}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$
+	 $d_{\text{RIGHT}}^{\text{BLUE}}$	$-\frac{1}{3}$	0	0	$-\frac{1}{2}$
	 e_{RIGHT}^+	+1	$+\frac{1}{2}$	0	0
<hr/>					
	 $u_{\text{LEFT}}^{\text{BLUE}}$	$+\frac{2}{3}$	$+\frac{1}{2}$	0	$-\frac{1}{2}$
+	 $d_{\text{RIGHT}}^{\text{RED}}$	$-\frac{1}{3}$	0	$+\frac{1}{2}$	0
	 $\bar{\nu}_{\text{RIGHT}}$	0	$-\frac{1}{2}$	0	0
<hr/>					
	 $d_{\text{LEFT}}^{\text{RED}}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{2}$	0
+	 $d_{\text{RIGHT}}^{\text{GREEN}}$	$-\frac{1}{3}$	0	$-\frac{1}{2}$	$+\frac{1}{2}$
	 $\bar{\nu}_{\text{RIGHT}}$	0	$-\frac{1}{2}$	0	0
<hr/>					
	 $d_{\text{LEFT}}^{\text{GREEN}}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$
+	 $d_{\text{RIGHT}}^{\text{BLUE}}$	$-\frac{1}{3}$	0	0	$-\frac{1}{2}$
	 $\bar{\nu}_{\text{RIGHT}}$	0	$-\frac{1}{2}$	0	0
<hr/>					
	 $d_{\text{LEFT}}^{\text{BLUE}}$	$-\frac{1}{3}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$
+	 e_{RIGHT}^+	+1	$+\frac{1}{2}$	0	0
	 $\bar{\nu}_{\text{RIGHT}}$	0	$-\frac{1}{2}$	0	0
<hr/>					
	 e_{LEFT}^+	+1	0	0	0
+	 $d_{\text{RIGHT}}^{\text{GREEN}}$	$-\frac{1}{3}$	0	$-\frac{1}{2}$	$+\frac{1}{2}$
	 $d_{\text{RIGHT}}^{\text{BLUE}}$	$-\frac{1}{3}$	0	0	$-\frac{1}{2}$
<hr/>					
	 $\bar{u}_{\text{LEFT}}^{\text{RED}}$	$-\frac{2}{3}$	0	$-\frac{1}{2}$	0
+	 $d_{\text{RIGHT}}^{\text{RED}}$	$-\frac{1}{3}$	0	$+\frac{1}{2}$	0
	 $d_{\text{RIGHT}}^{\text{BLUE}}$	$-\frac{1}{3}$	0	0	$-\frac{1}{2}$
<hr/>					
	 $\bar{u}_{\text{LEFT}}^{\text{GREEN}}$	$-\frac{2}{3}$	0	$+\frac{1}{2}$	$-\frac{1}{2}$
+	 $d_{\text{RIGHT}}^{\text{RED}}$	$-\frac{1}{3}$	0	$+\frac{1}{2}$	0
	 $d_{\text{RIGHT}}^{\text{GREEN}}$	$-\frac{1}{3}$	0	$-\frac{1}{2}$	$+\frac{1}{2}$
<hr/>					
	 $\bar{u}_{\text{LEFT}}^{\text{BLUE}}$	$-\frac{2}{3}$	0	0	$+\frac{1}{2}$

dence that all color-neutral systems of particles have integral electric charge follows from the organization of the family.

What about the remaining particles of the first generation? One of the most elegant features of the $SU(5)$ theory is that the five right-handed particles in the smallest family of $SU(5)$ can be combined in pairs to yield a family of 10 left-handed particles. These 10 states make up the next-simplest representation of the group. They are the left-handed components of the d quark, the u quark and the $\bar{\nu}$ antiquark (in three colors each) and of the positron. As in the assembly of antiquarks and gluons from the basic triplet of quark colors, this process should not be interpreted as a physical prescription for building the particles. A left-handed lepton or quark is not actually composed of two right-handed particles in a bound state. Nevertheless, the 10 possible ways of forming pairs of the five right-handed states yield all the correct charges for the left-handed particles.

The transitions available to any particle are also correctly given by this scheme of composition by pairs. They include both quark-lepton and quark-antiquark transitions. What is equally important, transitions that are not observed are not allowed by the structure of the group. Each family of particles is closed, or complete in itself. Every allowed transition gives rise to another particle in the same family and no other transitions are possible.

The family of five right-handed states and the family of 10 left-handed states hold a total of 15 particles. Two more families of slightly different form can be constructed to accommodate the remaining 10 right-handed particles and five left-handed particles, which are the antiparticles of the 15 states in the first two families. Hence all 30 elementary states of the first generation have a place in the theory, and there are no empty places. Equivalent representations of the higher generations can be construct-

FAMILY OF 10 PARTICLES is assembled by forming all possible pairs of the five states that make up the simplest family of $SU(5)$. All members of the family of five are right-handed, but when they are combined in pairs, they give rise to left-handed states. As in the building of antiquarks out of pairs of quarks, the process should not be interpreted as a physical assembly; a left-handed particle is not actually a bound state of two right-handed particles. Nevertheless, the assembly procedure gives all the correct charges of the left-handed states. With two more families of the same size and similar structure all the elementary particles of the first generation have a place in the theory and there are no vacant places.

ed by substituting the muon or the tau lepton for the electron, the *s* or *b* quark for the *d* quark, and so on.

Let me review what has been done so far. First the group $SU(5)$ was chosen as the smallest group in which $SU(3)$ and $SU(2) \times U(1)$ could be embedded. Next five right-handed components of particles were selected as members of the simplest $SU(5)$ family. The remaining components then had to fit into some other family of $SU(5)$, and they did. With nothing left out and nothing left over, they fit into the next-simplest family. Moreover, the composition of the derived family was specified by a simple procedure for combining particles in pairs. It is important to note that this assembly procedure did not have to work. In many groups other than $SU(5)$ it would not work. It represents the first, the simplest and in some respects the most remarkable success of the $SU(5)$ theory.

The most obvious significance of the $SU(5)$ unification is that the leptons and the quarks are no longer irreconcilably different. Instead they are members of a single family, and a quark can be converted into a lepton (or vice versa) as easily as one quark can be converted into a different quark or one lepton into a different lepton. A further consequence of the unification is that the weak, the strong and the electromagnetic forces should all have the same strength, or the same coupling constant. Neither of these expectations is satisfied in the world as it appears today. In the millions of elementary-particle interactions recorded by physicists not a single instance of quark-lepton conversion has been observed. Furthermore, the coupling constants of the three forces differ by large factors: the strong force is roughly 100 times as strong as electromagnetism, with the weak force somewhere in between. Thus if $SU(5)$ is a symmetry of nature, it is evidently a badly broken one.

The symmetry could be broken by a mechanism similar to the one that breaks the $SU(2) \times U(1)$ symmetry of the weak and electromagnetic forces. In this way the *X* particles would be given a large mass and the effects of *X*-particle exchange would be strongly suppressed. In $SU(5)$, however, the breakdown would have to come at a much higher energy or equivalently at a much smaller distance than in $SU(2) \times U(1)$. This distance is the unification scale, the distance at which the full symmetry of the theory becomes manifest.

With the $SU(5)$ theory one can speculate about what the world would look like at various scales of distance or energy. In an experiment that probed distances much smaller than the unification scale the $SU(5)$ gauge invariance of the

world would be readily apparent. All interactions, including the quark-lepton and quark-antiquark transformations, would be on an equal footing because all the $SU(5)$ intermediary particles (the photon, the gluons, the *W* and *Z* particles and the *X* particles) would be created with essentially equal probability. The masses of the *W*, *Z* and *X* particles would hardly distinguish them from the photon and gluons because the masses would be small compared with the energy of the experiment.

At distances close to the unification scale the complicated physics associated with the spontaneous breakdown of $SU(5)$ symmetry would be observed. *X* particles would be emitted, but their mass would make them quite different from all the other particles. At distances much greater than the unification scale (but still much smaller than 10^{-16} centimeter) the $SU(5)$ symmetry would be almost completely hidden. The *X* particles could no longer be created as real particles, and so the leptons and the quarks would become segregated in separate families, with little communication between them. On the other hand, the $SU(2) \times U(1)$ symmetry would remain intact, so that little distinction could be made between the weak and the electromagnetic interactions. At distances greater than 10^{-16} centimeter the $SU(2) \times U(1)$ symmetry would also be broken and there would be three distinct forces.

Spontaneous symmetry breaking can also explain the disparities among the coupling constants. The crucial element in the explanation is the effect of the virtual particles in the vacuum surrounding a point charge. In the case of the strong force, it will be remembered, the cloud of virtual gluons that envelops a quark effectively spreads out the color charge, with the result that the coupling constant decreases when it is measured at closer range. Virtual *W* particles have a similar effect on a weak charge, although the effect is somewhat smaller because there are fewer *W* particles that carry the weak charge than there are gluons that carry the color charge. In the $U(1)$ theory, on the other hand, the fact that the Z^0 has no charge gives rise to a quite different phenomenon. Because of the polarization of virtual electrons and positrons the $U(1)$ coupling constant increases at close range.

A simple inference can be made from these trends in the coupling constants. At long range the $SU(3)$ coupling constant of the strong force is the largest, but it also decreases the fastest; the $SU(2)$ constant of the weak force is smaller and decreases slower; the $U(1)$ constant is the smallest, but it increases as the distance is reduced. Hence it appears there may be some distance at

which all three coupling constants have approximately the same value.

At distances of less than 10^{-16} centimeter all the coupling constants are fairly small and the way they vary with range or energy can be calculated. Values can be determined at progressively smaller distances; the distance at which the three constants converge is the unification scale. If two of the constants are followed in this way, the value of the third constant can be predicted at any energy. Such calculations have been done using as input the strong coupling constant and the electromagnetic coupling constant; the latter, again, is a combination of the $U(1)$ and $SU(2)$ coupling constants. The results yield values of the unification scale and of the ratio of the $U(1)$ and $SU(2)$ constants, which had been an arbitrary parameter in nonunified theories.

In 1974 I made such a calculation with Helen R. Quinn (who is now at the Stanford Linear Accelerator Center) and Weinberg for a class of unified theories that includes $SU(5)$. We obtained a distance of about 10^{-29} centimeter for the unification scale and a value of .2 for the $U(1)$ -to- $SU(2)$ ratio. At the time the results were not encouraging because measurements of the ratio suggested a value of about .35. Since then refined measurements of the coupling-constant ratio have given lower values. The ratio is now thought to be equal to .22 plus or minus .02, in agreement with the theoretical result.

Another prediction of the $SU(5)$ model that can be tested at accessible energy was worked out in 1977 by Andrzej Buras, John Ellis, Mary K. Gaillard and Demetres V. Nanopoulos of the European Organization for Nuclear Research (CERN) in Geneva. They found that in the simplest version of the $SU(5)$ theory the ratio of the mass of the *b* quark to the mass of the tau lepton can be determined. As in the case of the coupling constants, the masses are expected to be equal at the unification scale, but at greater distances the *b* quark is heavier because of its color charge. The mass ratio at low energy was calculated to be about 3:1. The mass of the tau is known: it is a little less than twice the mass of the proton. The mass of the *b* quark is not as certain because the quark cannot be examined in isolation. The best estimate at present is about five times the proton mass, making the ratio of the masses 5:2.

The unification scale of 10^{-29} centimeter is an extraordinarily small distance. (If a single proton were inflated to the size of the sun, the unification scale would still be less than a micrometer.) Implicit in the $SU(5)$ unification is the assumption that no new physical principles will be encountered in the entire

range of distances between 10^{-16} and 10^{-29} centimeter; in particular, it must be assumed that the way the coupling constants vary with distance remains unchanged. Such an assumption is of course disturbing, but it is not entirely implausible. There is already a small scale of distance at which new phenomena are expected. At about 10^{-33} centimeter gravitation may become as strong as the other forces, and so any theory that describes events at this scale must include gravitation. I find it encouraging that although the unification scale is small, it is still 10,000 times greater than 10^{-33} centimeter.

A distance of 10^{-29} centimeter corresponds to an energy of about 10^{15} GeV, or roughly 10^{15} times the mass of the proton. The X particles must have a mass equivalent to this energy. For the sake of comparison, the heaviest particles that can now be created with particle accelerators have a mass of about 10 GeV. With a new generation of accelerators now being planned it is hoped to reach the 100 GeV or so needed to create W and Z particles. In order to create X particles the energy would have to be further increased by 13 orders of magnitude, which seems unlikely ever to be feasible.

Even if it is never possible to exhibit a

real X particle in the laboratory, the existence of such particles might be demonstrated by detecting events in which a virtual X particle is exchanged. Such exchanges would themselves be extremely rare, since they could take place only when two elementary particles happened to stray within less than 10^{-29} centimeter of each other. Even in the welter of commoner events, however, the exchange of an X particle would be readily discerned because X particles can do something no other particles can do: they can transform a quark into a lepton or a quark into an antiquark. This process puts in question the very stability of matter.

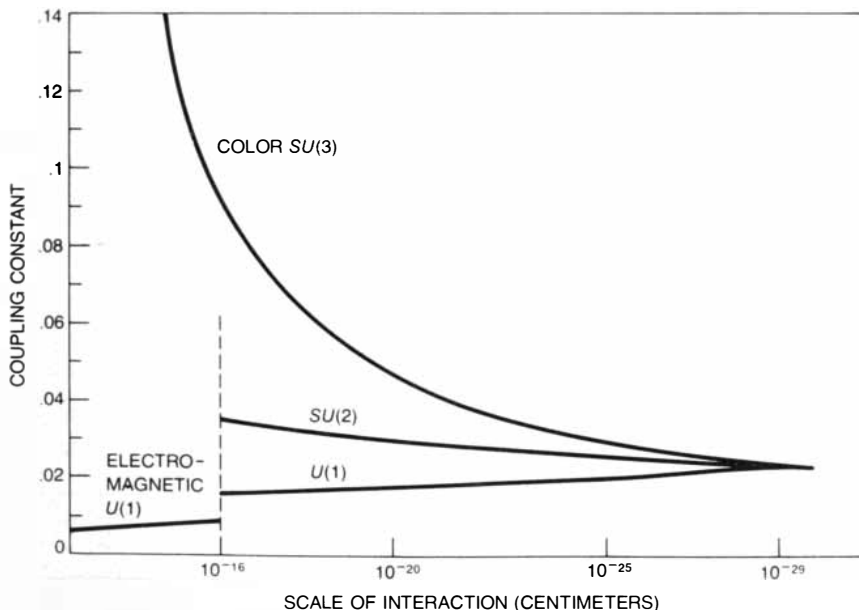
The interactions mediated by X particles differ from all other interactions in that they violate the conservation of a quantity called baryon number. The baryon number of any particle can be defined as one-third of the number of quarks minus one-third of the number of antiquarks. Hence the proton and all other particles made up of three quarks have a baryon number of +1, whereas the pi meson and other particles that consist of a quark and an antiquark have a baryon number of 0. Of course the leptons also have a baryon number of 0, since they include no quarks or anti-

quarks. In the strong, the weak and the electromagnetic interactions the total baryon number can never change. If the conservation of baryon number were an absolute law of nature, the proton could never decay because the proton is the lightest particle with nonzero baryon number. The $SU(5)$ unification predicts just such a decay.

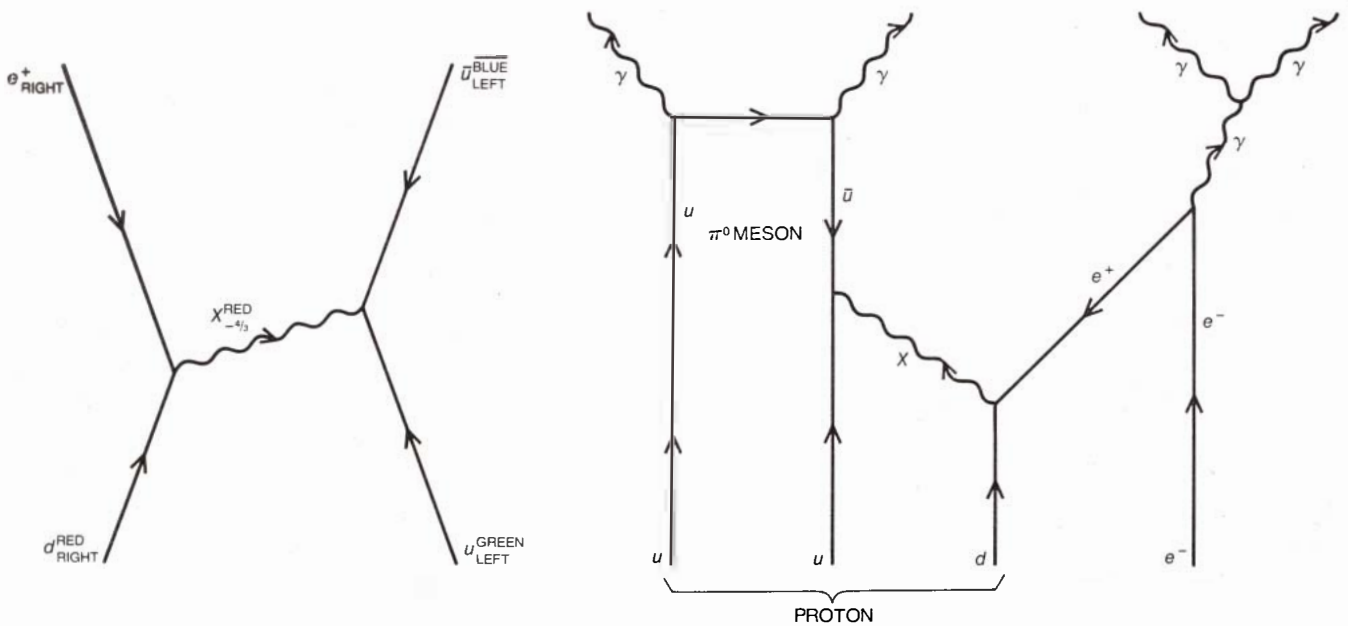
A possible failure of baryon-number conservation can be illustrated in a proton that forms the nucleus of a hydrogen atom. The proton consists of u , u and d quarks, with one quark in each of the three colors. If two of the quarks happen to approach to within 10^{-29} centimeter, an X particle can pass between them. For example, a right-handed red d quark can emit an X with an electric charge of $-4/3$ and color charges corresponding to the color red. The d quark, having lost its color charge and having changed its electric charge from $-1/3$ to $+1$, would thereby become a positron. Meanwhile the X particle could be absorbed by a left-handed green u quark, which would be converted into a left-handed \bar{u} antiquark with the color antiblue. The new \bar{u} antiquark would combine with the remaining u quark to form a neutral pi meson. The baryon numbers of both the positron and the pi meson are zero, so that the total baryon number changes from +1 to 0.

If this event were observed in the laboratory, the exchange of the X particle could not be directly perceived. All that would be seen would be the decay of a proton into a positron and a neutral pi meson. Moreover, that would not be the end of the sequence of events. The positron would subsequently encounter an electron (perhaps the electron of the hydrogen atom) and they would annihilate each other, giving rise to gamma rays, or high-energy photons. The u quark and \bar{u} antiquark of the neutral pi meson would eventually annihilate each other in a similar way, releasing more gamma rays. The end result is that a hydrogen atom decays into a state of pure radiation. This process represents a conversion of matter into energy far more efficient than that of nuclear fission or thermonuclear fusion. The fusion of hydrogen atoms to form helium releases less than 1 percent of their mass as energy, whereas the process outlined here liberates 100 percent of the mass.

The abrupt disappearance of a proton and hence of an atom is an event that must happen very rarely or it would have been noticed long ago. Indeed, a low rate is to be expected because particles seldom come within the range where X -particle exchange is possible. Quinn, Weinberg and I employed our calculation of the unification scale to estimate the decay rate of the proton. Our estimate has since been refined by many others, including Buras, Ellis, Gaillard



UNIFICATION OF FORCES in the $SU(5)$ theory is expected to become apparent at extremely high energy or equivalently at a very small distance. The intrinsic strength of each force is measured by a dimensionless coupling constant associated with the underlying symmetry. Because of the effects of virtual particles surrounding a charge, the strong $SU(3)$ interaction declines in strength as the separation between particles is reduced. The $SU(2)$ interaction also becomes weaker at close range, but at a lesser rate. The $U(1)$ interaction, on the other hand, becomes stronger. Extrapolation from the values of the constants measured at comparatively long range suggests that the curves converge at a distance of about 10^{-29} centimeter, equivalent to an energy of 10^{15} gigaelectron-volts. At this range and energy all the forces would have the same strength and all transitions between the elementary particles would be equally likely. At distances greater than 10^{-16} centimeter the $SU(2) \times U(1)$ symmetry is broken and the $SU(2)$ and $U(1)$ forces cease to exist as separate entities. The $U(1)$ symmetry of electromagnetism remains, but the weak force appears only through the virtual exchange of W and Z particles.



DECAY OF THE PROTON is the inevitable consequence of the $SU(5)$ transitions that convert a quark into a lepton or a quark into an antiquark. The nature of such a conversion is shown in the diagram at the left. First a right-handed red d quark emits a red X particle with an electric charge of $-4/3$; as a result the quark is transformed into a right-handed positron. Next the X particle is absorbed by a left-handed green u quark, which thereby becomes a left-handed antiblue \bar{u} antiquark. The diagram at the right shows the outcome of the process when it takes place in a proton that forms the nucleus of

a hydrogen atom. The proton has the quark composition uud . As a result of the X -particle exchange the d quark becomes a positron, which can eventually collide with the electron of the atom (or with some other electron), annihilating both particles in a burst of high-energy photons. What is left of the proton after the X -particle exchange is a u quark and a \bar{u} antiquark, which together constitute a neutral pi meson. The neutral pi meson also decays into high-energy photons. Thus the end result of the exchange is the conversion of the entire mass of the hydrogen atom into electromagnetic radiation.

and Nanopoulos of CERN, Terrence J. Goldman and Douglas A. Ross of Cal Tech and William J. Marciano of Rockefeller University. The present estimate is that the average lifetime of the proton is about 10^{31} years.

As a test of the $SU(5)$ unification it is obviously impractical to wait 10^{31} years for a given proton to decay. The age of the universe since the big bang is only about 10^{10} years. The search for proton decay is not hopeless, however. An average lifetime of 10^{31} years implies that in a collection of 10^{31} protons a decay ought to be observed once a year. In 1,000 tons of matter there are about 5×10^{32} protons and neutrons, so that roughly 50 of them can be expected to decay each year. Hence the strategy for detecting events that violate the conservation of baryon number is to monitor everything that goes on in at least 1,000 tons of matter for several years and to distinguish decaying protons and neutrons from commoner events.

Several groups of investigators are planning experiments on this scale. The experiments will be done deep underground or underwater in order to reduce to a minimum the number of cosmic rays passing through the sample of matter. The cosmic rays can give rise to interactions that might be mistaken for proton decay. One experiment is to be installed in a salt mine near Cleveland,

another in a silver mine in Utah and a third in an iron mine in Minnesota. Experiments of a slightly smaller scale are planned for two tunnels under the Alps, and smaller experiments are already under way in gold mines in South Dakota and India.

The energy needed to create real X particles may be forever beyond the capabilities of manmade machines, and indeed there may be no process anywhere in the universe today that can generate such a high energy. In an earlier era, however, X particles may have been commonplace. About 10^{-40} second after the big bang the size of the universe was comparable to the unification scale. The universe was then so hot (about 10^{18} degrees Kelvin) that all particles had energies comparable to the X mass. Consequently the $SU(5)$ symmetry was just beginning to break down, and quark-lepton conversions were as frequent as any other interactions. No fundamental distinction could be made between quarks and leptons or between the strong, weak and electromagnetic forces: there was one kind of matter and one force.

A remnant of that era of manifest symmetry may persist in the universe today; in a sense the universe today is the remnant. An old puzzle of astrophysics asks why the universe is built out of matter rather than antimatter. Actually, perhaps the simplest expectation is that there would be equal quanti-

ties of matter and antimatter, which would ultimately annihilate each other to leave a universe consisting of nothing but radiation. The $SU(5)$ unification offers a conjectural explanation for the apparent predominance of matter. It is possible that the free interchange of X particles in a brief period after the $SU(5)$ symmetry was broken created more quarks than antiquarks and hence more baryons than antibaryons.

The intriguing speculation that processes violating baryon number might be responsible for the excess of baryons was first made by the Russian physicist Andre Sakharov in 1967. More recently Motohiko Yoshimura of Tohoku University suggested that the baryon-number violation predicted by unified theories has the right properties to account for the observed excess. The idea has since been elaborated by many others, including Ellis, Gaillard, Nanopoulos, Weinberg, Savas Dimopoulos and Leonard Susskind of Stanford University and Sam B. Treiman, Anthony Zee and Wilczek of Princeton. They have shown that an excess of baryons can arise only if the processes that violate the conservation of baryon number look different when they are run backward in time. This condition is satisfied by the $SU(5)$ theory. Thus one item of evidence for the $SU(5)$ unification, albeit indirect and circumstantial evidence, is the very existence of matter.

Speech Recognition by Computer

Designing a machine that listens is much more difficult than making one that speaks. Significant improvements in automatic recognition may come only with a better understanding of human speech patterns

by Stephen E. Levinson and Mark Y. Liberman

Modern computers have prodigious powers, but they would be still more useful if more natural ways of communicating with them were possible. The evolution of spoken language has made it well adapted to the needs of human communication. It is fast and nearly effortless. It requires neither visual nor physical contact and it places few restrictions on the mobility of the body or on the use of the hands. A machine capable of recognizing human speech could combine these advantages with the quite different powers of the computer. Such a machine could provide universal access to large data bases through the telephone network. It could provide for the control of complex machines by vocal command and make possible sophisticated prosthetic devices for the handicapped.

After more than 40 years of research, however, the automatic recognition of natural or conversational speech remains a utopian goal. Current speech-recognition devices have small vocabularies and little ability to deal with fluent sequences of words; usually they must be trained to recognize only one speaker's voice. Even so, the advantages of automatic speech recognition are so great that devices capable of recognizing isolated words or short phrases from a vocabulary of between 10 and 30 words are commercially available and are economically practical in some applications. In research laboratories there are speech recognizers with vocabularies of up to 1,000 words, systems that recognize limited-vocabulary sentences with brief pauses between the words and systems that recognize connected speech with fair accuracy if the vocabulary is small, the syntax is limited and the speaker is careful.

The interaction of technology and economics will undoubtedly lead to speech-recognition systems of greater capability. We cannot accurately predict the pace of such development. We are certain, however, that mere elaboration and extrapolation of current technology will not lead to the development

of machines that match the human capacity for recognizing speech. Major progress depends on new discoveries.

Why is the problem of recognition so hard? The core of the difficulty is the complex and variable way linguistic messages are encoded in speech. Spoken language enables people to express their thoughts in sound and to recover messages from the sounds produced by others. This curious two-way mapping between concepts in the mind and vibrations in the air presupposes the participants have some common conceptual framework, so that the message received is at least approximately equivalent to the one that was sent. It is not enough, however, to share knowledge of the things one might want to say. Monolingual speakers of English and of Finnish may have many potential messages in common without being able to understand each other's utterances at all. In order to speak and to understand, people must share a system for encoding messages in sound and for decoding speech sounds to yield meanings. In other words, they have to know the same language.

Speech communication with a computer can be understood in an analogous way. The computer "knows" (in some extended sense of the word) about a domain its users also know about. It is useful for information in this domain to be exchanged, and speech happens to be the chosen medium of communication.

Consider a conversation between a computer and its users concerning the inventory of a warehouse. The computer "knows" how many of each item are on hand and where each article is stored. Its data base also lists costs and suppliers. People probably think about the warehouse and its contents in many ways, but the structure of the computer data base is sufficiently similar to one mode of human thought for certain kinds of communication to be possible. The users have questions that the computer can answer, at least in principle, such as, "Do we have any blue pencils in

stock?" The users also have things to say that the computer can profitably "understand," such as, "There is no more room in bay 13." If such communication is to be accomplished through the medium of speech, the computer and its users must agree on how to encode such messages in sound and how to reverse the process. They have to "know" the same language.

We are mainly interested in languages such as English, which are called natural because they are defined implicitly by the everyday usage of ordinary people. Computers now function with formal languages such as FORTRAN, which are defined by an explicit set of rules consciously established by specialists. At least for now, computers do only what they are specifically programmed to do. They do not live in the world of people and learn from everyday experience. Hence for a computer to "know" a natural language, it must be provided with an explicit and precise characterization of the language, or at least with a characterization of what the programmer takes the language to be. In all existing and currently conceived speech-recognition systems the formal description of a natural language covers only a fragment of the language, and the formalism reconstructs the fragment in ways that are probably quite different from the implicit knowledge of a native speaker. Even imperfect linguistic abilities on the part of a computer, however, are enough to make possible useful communication with people.

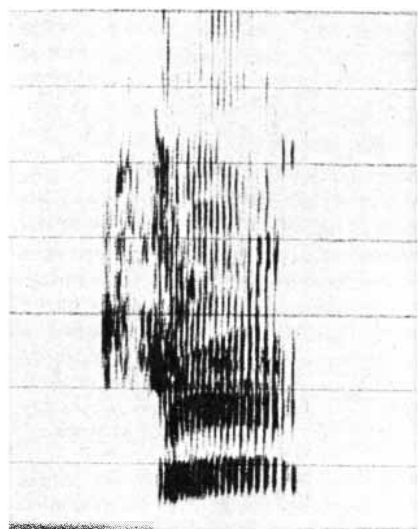
It will help in understanding various approaches to natural-language recognition if we begin by considering some aspects of language and speech in their own terms. We shall then discuss methods of recognizing isolated words and review some procedures for analyzing connected speech. Finally we shall describe a speech-recognition system developed at Bell Laboratories that attempts to combine the major elements of human speech communication into a single operating unit.

At the center of human language is the word. Sequences of words are generally arranged into phrases according to principles of combination known as syntax; moreover, such sequences are usually intended to mean something. The fact that words are ordinarily part of coherent discourse can help in the recognition of the words themselves by providing a context in which some words are likelier than others. Arranging for a computer to act as if it could "understand" word sequences is formidably difficult. The problem involves not only relations among words but also knowledge and reasoning about the nature of the world.

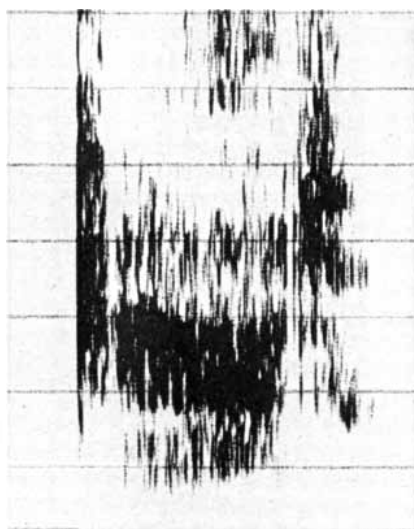
Although a capacity for understanding language may be the ultimate goal, the enterprise of speech recognition is really founded on the identification of words. The aspect of words that concerns us here is their sound. In this sense a word is an equivalence class of noises: the set of all sounds, however distinct in other ways, that represent (in the context of their utterance) the same lexical unit. The problem in word recognition is to find a mathematically defined space in which such a set of sounds can be effectively delimited. Because the amount of variation within the set of sounds corresponding to a given word is quite large, because the acoustic distinc-

tion between words can be quite small and because an adult speaker may know 100,000 words or more, the problem is a difficult one.

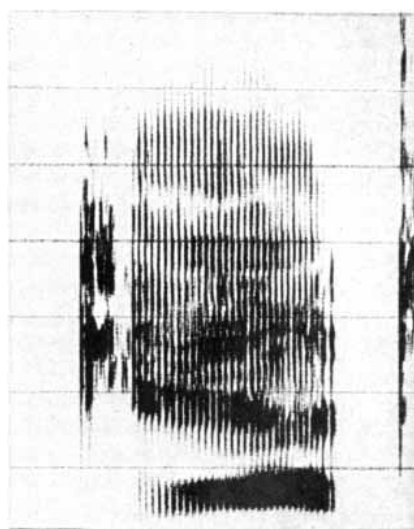
In order to understand the sources of variation in the sound of a word and the nature of the distinction between one word and another, it is necessary to grasp two things. First one must understand the basic medium of spoken communication: the ways in which acoustic disturbances of the air can be produced by the human vocal apparatus and perceived by the human auditory system. Second, one has to recognize that speech sounds are elements of a phonological system peculiar to a given language. The



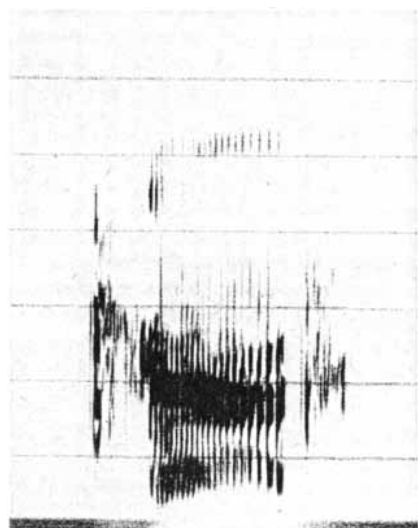
"CAT" SPEAKER 1 MICROPHONE



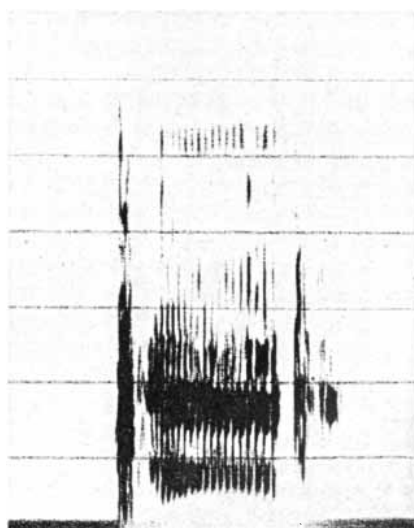
"CAT" SPEAKER 1 WHISPERED



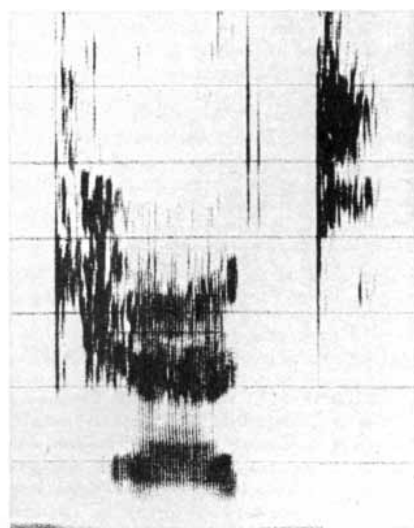
"CAT" SPEAKER 2 MICROPHONE



"CAT" SPEAKER 1 TELEPHONE



"PAT" SPEAKER 1 TELEPHONE



"CAT" SPEAKER 3 MICROPHONE

VARIABILITY OF HUMAN SPEECH, illustrated here by means of sound spectrograms, is one of the principal difficulties encountered in building an automated system for speech recognition. Spectrograms of distinct but acoustically similar words may be more alike than spectrograms of the same word pronounced under various conditions by different speakers. Automatic speech recognition must

be able to attend only to relevant spectral differences (when they exist) and must disregard apparent differences that are linguistically irrelevant. The sound spectrogram represents a series of amplitude spectra over time. Time varies along the horizontal axis and frequency varies along the vertical axis. The darker the mark on the graph, the greater the amplitude of the waveform at that frequency and time.

```

NFRAM= 396
NO. OF WORDS= 5
CANDIDATES FOR WORD NO. 1 20 FRAMES
WHAT 1 .180
CANDIDATES FOR WORD NO. 2 29 FRAMES
IS 1 .270
CANDIDATES FOR WORD NO. 3 24 FRAMES
NINE 1 .343
ONE 1 .278
SIX 1 .370
SEVEN 1 .242
THE(ALT) 1 .314
CANDIDATES FOR WORD NO. 4 24 FRAMES
SOME 1 .296
FLIGHT 1 .341
ONE 1 .414
THREE 1 .390
CANDIDATES FOR WORD NO. 5 28 FRAMES
NINE 1 .378
OH 1 .390
TIME 1 .291
ONE 1 .411
SEVEN 1 .396
TIMES 1 .315
WHAT IS SEVEN SOME TIME
METRIC= .1279442E 01
WHAT IS THE FLIGHT TIME
METRIC= .1396801E 01
THE FLIGHT TIME IS FIVE HOURS AND TWENTY-FIVE MINUTES

```

Machine: This is the Bell Laboratories flight information system. May I help you?

User: I want one first-class seat on flight number three one to Denver on Sunday.

Machine: I am sorry the flight is not available on Sunday.

User: I would like to leave on Saturday.

Machine: Flight number three one leaves New York at nine forty-five A.M., arrives in Denver at one ten P.M.

User: **What is the flight time?**

Machine: **The flight time is five hours and twenty-five minutes.**

User: How many stops are there on the flight?

Machine: This flight makes one stop.

User: I want a nonstop flight.

Machine: I am sorry there are no nine A.M. nonstop flights.

User: I want to go at four P.M.

Machine: Flight number one six leaves New York at four fifty P.M., arrives in Denver at seven P.M.

User: What stops are there on the flight?

Machine: This flight makes no stops.

User: Is a meal served on the flight?

Machine: Dinner is served.

User: How much is the fare?

Machine: The round-trip first-class fare is three hundred sixty dollars.

User: What plane is on the flight?

Machine: The aircraft is a Douglas DC ten.

User: I prefer the Boeing seven four seven.

Machine: The aircraft is a Douglas DC ten.

User: Please repeat the departure time.

Machine: The departure time is four fifty P.M.

User: I will take flight number six five.

Machine: I am sorry there is no flight six five to Denver.

User: Please repeat the flight number.

Machine: The flight number is one six.

User: I want flight one six.

Machine: Your reservation on flight number one six to Denver on Saturday is confirmed. Thank you.

TELEPHONE CONVERSATION between man and machine proceeds according to steps that can be displayed on the screen of a computer terminal. The computer counts the number of words in each sentence to be recognized and also divides the entire utterance into "frames" by taking a spectral sample every 15 milliseconds. "NFRAM" is the number of frames in the sentence. It exceeds the number of frames occupied by individual words because the speaker must pause briefly between the words. The candidate words listed for each position in the sentence have been found by comparison with word templates stored in the computer. Only those candidates appear that are grammatically possible in a given position and similar in spectral structure to the spoken word. Not all the candidates to be considered are listed. The numbers following each candidate word are measures of the distance between the word's template and the utterance. The shorter the distance is, the more similar the template is to the utterance. "METRIC" is the unrounded sum of the distance measures for a particular string of words. If the smallest possible METRIC (which necessarily consists of the most likely word in each position) is not allowed by the internal grammar, the grammatically correct string with the smallest METRIC is substituted. A synthetic-voice response to the question by the user is given over the telephone. The complete conversation is transcribed in the computer printout.

phonological system limits the ways in which the various words of the language can differ and controls in part the ways in which the pronunciation of any specific word in the language can vary.

During speech a flow of air from the lungs passes through the larynx, or voice box, into the throat and out through the mouth. If the velum (the flap of soft tissue at the rear of the palate) is lowered, the airflow also proceeds out through the nose; if the velum is raised, the nasal passages are blocked. The airflow can also be obstructed by closing the lips, by pressing the tongue against the palate or by closing the glottis, which consists of two parallel folds of soft tissue (the vocal cords) within the larynx.

The flow of air through the vocal tract can give rise to sound in three main ways. First, the vocal cords can be made to vibrate in somewhat the same manner as the double reed of an oboe or a bassoon. When the vocal cords are brought together, they stop the passage of air from the lungs, and so pressure builds up below them. The pressure forces the vocal cords apart, but the velocity of the rushing air then reduces the pressure in the space between them. The reduction in pressure and the elasticity of the tissues bring the vocal cords together again, in position for another buildup of pressure. The rate at which this cycle is repeated is the fundamental frequency of the voice, which is heard as pitch.

The second way of generating sound in the vocal tract is to form a constriction in the airway narrow enough to cause turbulence. For example, forcing air past a close contact between the upper teeth and the lower lip causes a turbulent flow that is perceived as the sound "f." Unlike the periodic sounds created by vibration of the vocal cords, the sounds generated by turbulent flow are aperiodic, or noise-like. It is possible for the vocal tract to create both periodic and aperiodic sounds at the same time. Combining vocal-cord vibration with the noise source of an "f" gives rise to the sound perceived as a "v."

A third kind of sound generation takes place when pressure built up behind a closure is abruptly released. Such bursts of acoustic energy occur in the pronunciation of consonants such as "p," "t" and "k."

These three sources of speech sound are shaped acoustically by the changing physical shape of the vocal tract. If the vibrations of the vocal cords were somehow vented directly to the outside air without first passing through the throat, mouth and nose, they would sound rather like a door buzzer and not like speech at all. On passing through the throat, mouth and nose cavities, however, the quality of the buzz is changed profoundly. It is the shape of the vocal tract, including the positions of the larynx, the

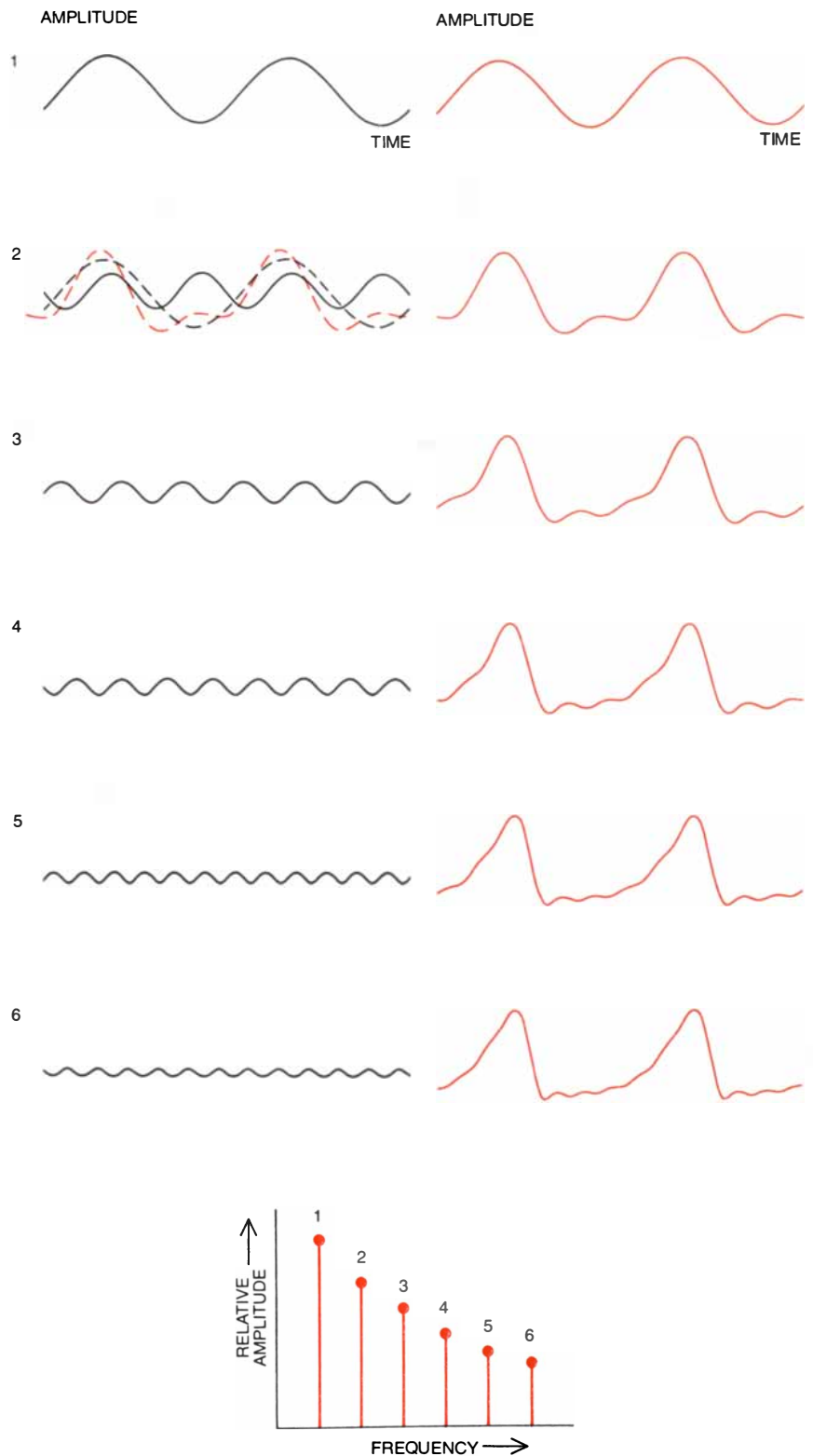
tongue, the lips and the velum, that distinguishes (for example) the “ee” sound in “me” from the “oo” sound in “you.”

One way of understanding this acoustic transformation is the mathematical technique called Fourier analysis. In 1822 the French mathematician Jean Baptiste Joseph Fourier showed that any periodic waveform can be represented as the sum of an infinite series of sine waves. A periodic waveform is one that is repeated at uniform intervals. If the interval of repetition is t seconds, the fundamental frequency of the waveform is $1/t$ hertz. In the Fourier series for a periodic waveform, the frequencies of the component sine waves are harmonics, or integral multiples, of the fundamental frequency of the waveform being analyzed, and they must be assigned appropriate amplitudes and phases. The Fourier transform is a generalization of the Fourier series; it allows analysis of aperiodic waveforms. Thus the noisy hiss of an “f” sound can be represented as a sum of sinusoidal components all along the frequency continuum.

The most obvious way to represent sound waves is to graph the variation of air pressure with time. Fourier’s result implies that the same information can also be displayed by a graph that shows amplitude and phase as a function of the frequency of the sinusoidal components. Because phase differences are of little perceptual significance, a speech sound can be represented in practice by its amplitude spectrum, a graph that shows the amplitude of the sine-wave component at each frequency.

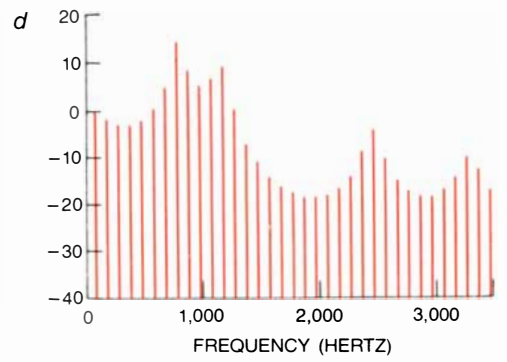
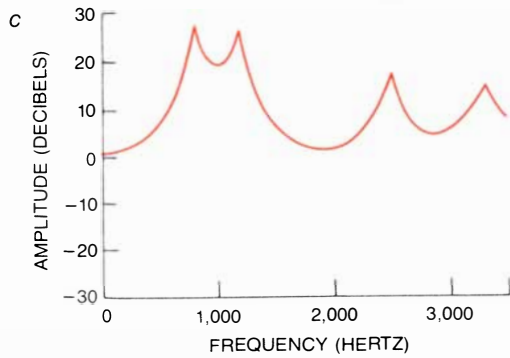
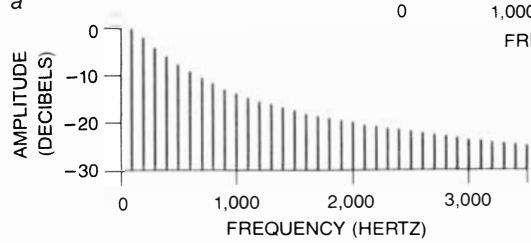
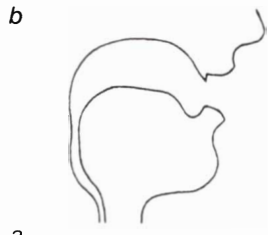
What is the acoustic effect of the shape of the vocal tract on the sound emitted? When the sounds are represented by their amplitude spectra, the effects are clear [see illustration on next page]. The vocal tract acts as a filter on the spectrum of the sound source, enhancing some frequencies and diminishing others. The selective filtering can be described by a mathematical expression called a transfer function; a separate transfer function is associated with each position assumed by the organs of the vocal tract. The transfer function usually has several well-defined frequency peaks, called formants, in which most of the energy from the sound source is concentrated.

It is now possible to state with some precision why it is hard for a computer to make the translation from sounds to words, from an acoustic characterization of an utterance to a linguistic characterization of the intended message. One source of difficulty is that the organs of speech do not take up a series of fixed configurations corresponding to units of the message. Instead parts of the vocal tract are in constant motion along smooth trajectories. Some investigators

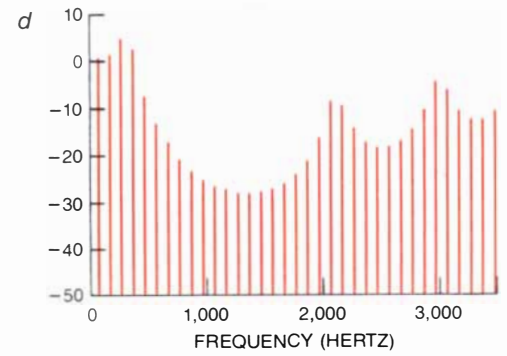
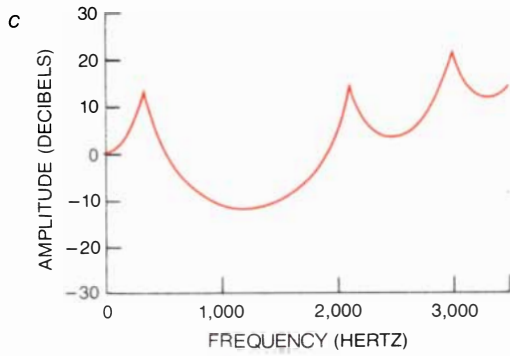
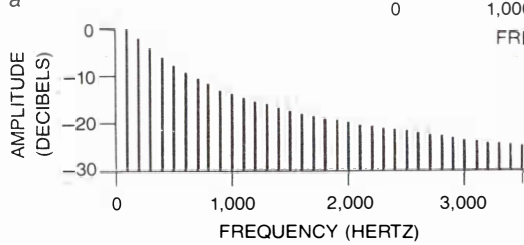
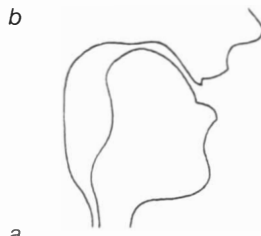


PRINCIPLE OF SUPERPOSITION allows temporal variation in the sound pressure of the signal to be represented as a spectrum of sound amplitude, or energy, at different frequencies. The amplitude spectrum is generally a more useful way of displaying acoustic information. The waveform (here a glottal wave) can be treated mathematically as a pattern that repeats indefinitely in the past and the future at a fundamental frequency. As the French mathematician Jean Baptiste Joseph Fourier showed, any such waveform can be decomposed into a series of sine waves at integral multiples of the fundamental frequency, with various amplitudes and phases. When the sine waves are combined by adding the amplitudes at each point, the result is the original waveform. When the amplitude of each sine wave that makes up the decomposition is graphed as a function of its frequency, the result is an amplitude spectrum.

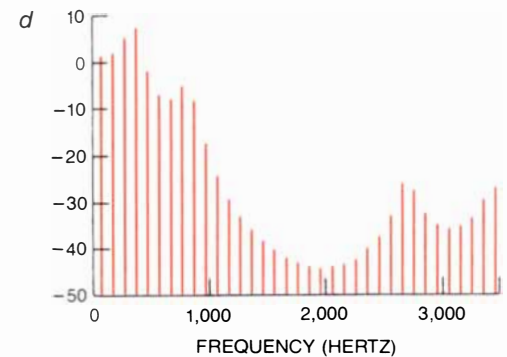
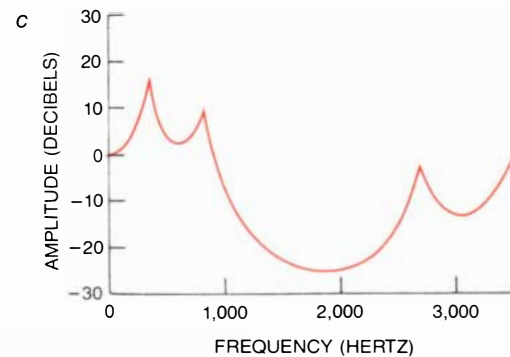
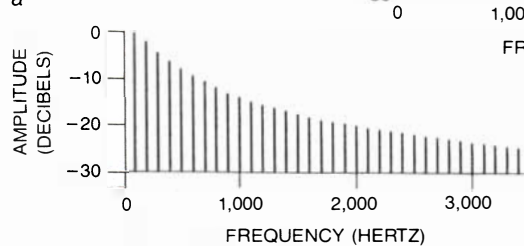
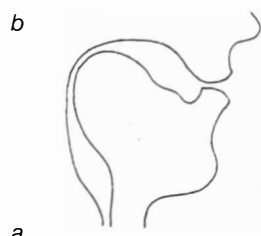
AH



EE



OO



VOWEL SOUNDS result from various configurations of the mouth, lips, tongue and velum (soft palate). The resulting shape of the vocal tract can be modeled by a series of resonating cavities that enhance energy at some frequencies and diminish it at others in predictable ways. Such filter-response characteristics can be represented by a transfer function (c) for each position of the model of the vocal tract (b). When the input sound energy is periodic (which is almost the

case for vocal-cord vibration), both the input spectrum (a) and the output spectrum (d) are line spectra. In a line spectrum the sound energy is concentrated at harmonics, or integral multiples, of the vocal-cord frequency. An aperiodic sound source such as a whispered vowel has no discrete lines in its spectrum, but the shape of the output spectrum still matches that of the transfer function. Model vocal-tract configurations for vowels "ee," "ah" and "oo" are shown.

believe these motions “flow” through a sequence of target positions defined by linguistic units such as consonants and vowels. Others think even the simplest linguistic units are inherently dynamic. In any case the result is a complex and continual motion, which is inherited by the emitted sound in the form of a constantly changing amplitude spectrum. Such patterns of changing sound quality can conveniently be represented by a sound spectrogram, a graph in which time proceeds from left to right, frequency increases from bottom to top and amplitude increases from white through shades of gray to black.

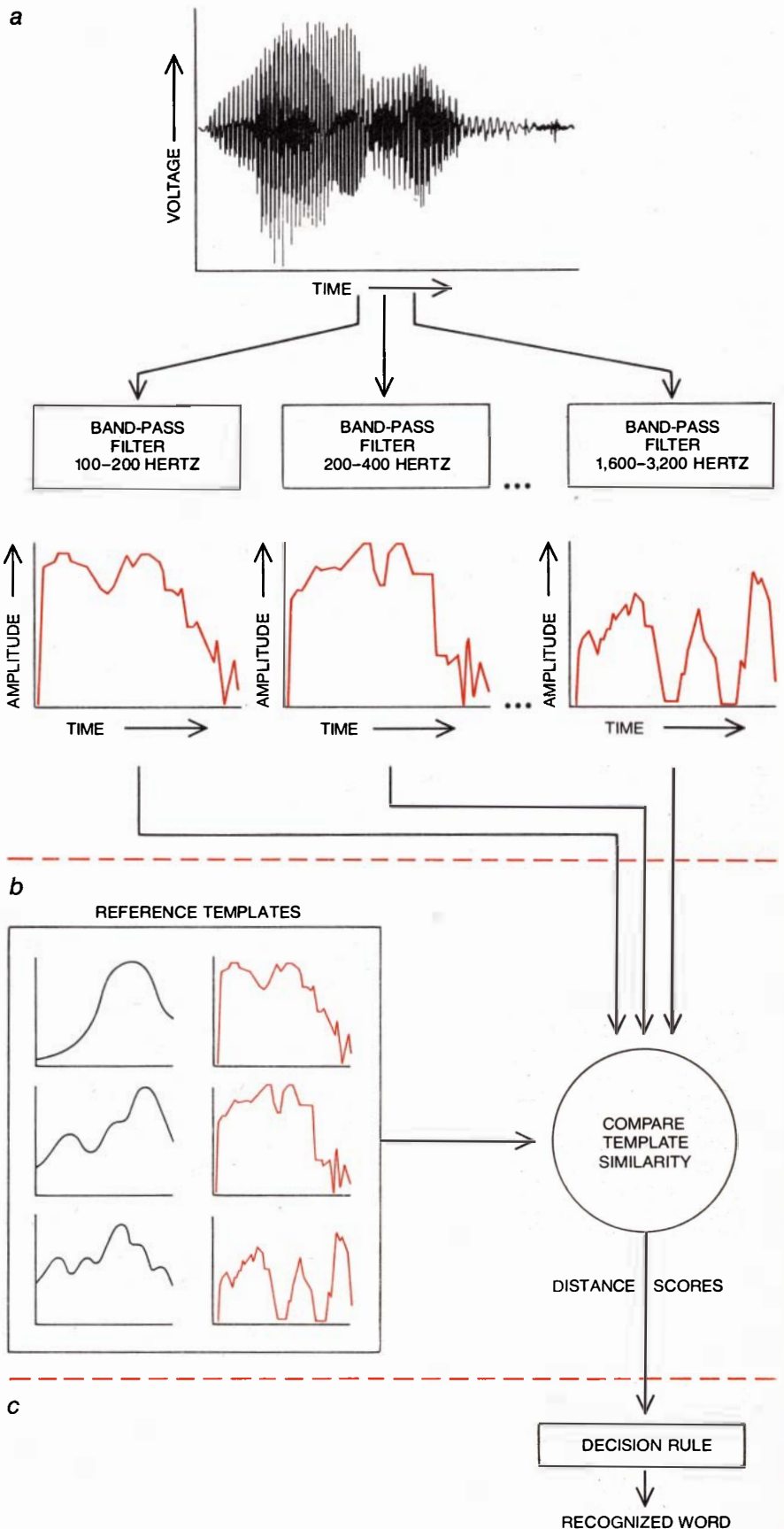
Motions of the vocal tract that correspond to linguistic units usually overlap and combine with their neighbors. For instance, in saying “coo” the lip-rounding of the vowel “oo” usually precedes the tongue motion of the initial consonant. Hence the acoustic effects of the two motions are combined from the beginning of the word. In fluent speech such amalgamation also applies between one word and the next. The effects are sometimes quite plain to the ear. When the “t” of “cat” combines with the “y” of “your,” it makes the phrase “You gave the cat your dinner” sound like “You gave the catcher dinner.”

Other variations in the sound of a word result from its position in a phrase, from its degree of emphasis and from the rate at which it is pronounced. The size and shape of the vocal tract vary from individual to individual, and habits of speech differ widely according to age, sex, geographic region and social background. Furthermore, the signal that reaches a speech-recognition device is influenced by various circumstances in addition to the sounds made by the speaker, such as room acoustics, background noise and the characteristics of the transmission channel.

For these reasons it is hard to divide a speech signal into chunks corresponding to the elements of the message the signal conveys, and it is difficult to translate pieces of acoustic information into information about the identity of pieces of the message. People find speech easy to understand, and so the needed information must be present in the signal. The trick is to find it.

A natural starting place is the recognition of words. Words are generally distinct from one another as elements of a linguistic system, and they constitute natural and relatively stable patterns for an automated speech-recognition system. Although speech is more than a sequence of words, it is at least such a sequence, so that a crucial function of a speech recognizer is identifying words. If a speech-recognition system can recognize words accurately, it will succeed; if it cannot, it will fail.

Most speech recognizers now in use are not capable of recognizing words in



STANDARD METHOD OF WORD RECOGNITION employs the basic principles of pattern recognition to discriminate among acoustic patterns. The speech waveform is measured and analyzed (a), in this case by filters that divide the signal into frequency bands, each band being an octave wide. The output of each filter is the energy in its band. The outputs are compared with stored reference templates, and distance scores are assigned to each template (b). A decision procedure then classifies the input utterance on the basis of the distance scores (c).

connected speech. Instead recognition is carried out on isolated words by a process of acoustic pattern recognition. Generally the user must "train" the machine by speaking into a microphone all the words the system is to recognize. In some cases the training is limited to one utterance of each word by a few of the speakers who will use the system. In other applications every potential user must say each word several times. The result of this training process is a set of stored "templates," which represent typical acoustic patterns for the words in the vocabulary.

When a word is presented for recognition, the machine analyzes the acoustic signal, compares the results of the analysis with the stored templates and decides which one of the templates most closely resembles the spoken word. The machine may also list other possible matches in decreasing order of similarity. Once a classification has been made the machine can respond to the user's utterance or issue an appropriate signal to some other device. Each stage of the template-matching procedure (analysis of the speech signal, comparison with the template and classification of the signal) can be carried out by a variety of techniques.

The aim of all methods of analyzing the speech signal is to characterize the temporal variation of the signal's amplitude spectrum. Perhaps the simplest method of estimating the spectrum is the zero-crossing count. This method consists in counting the number of times the voltage analogue of the speech signal changes its algebraic sign (from plus to minus or from minus to plus) in a fixed interval. The number of such axis crossings is related to the frequency.

One refinement of the zero-crossing method filters the speech signal into three frequency bands. The zero crossings are measured separately in each band to give rough estimates of the first three formant frequencies. Such measurements are useful in classifying vowel sounds, and for small vocabularies of easily distinguished words these measurements alone are sufficient for discrimination. The zero-crossing method is economically attractive because it can be accomplished by simple electronic devices.

A more elaborate procedure for spectral estimation is the filter-bank method. The speech signal is divided by filtering into between 20 and 30 frequency bands, covering the frequency range of human speech. The output of each filter is a measure of the energy in that frequency band. The energy levels are suitable for direct comparison with those of a template. The Fast Fourier Transform provides a general, computationally efficient method for estimating the amplitude spectrum of a signal from its time-domain waveform. This algorithm

provides one of several ways to obtain filter-bank information in purely digital form.

Recently a new method for estimating the amplitude spectrum of speech, called linear predictive analysis, has been introduced. Actually statisticians have employed the method for some time under the name autoregressive analysis. The method predicts the amplitude of a speech wave at a given instant from a weighted sum (or linear combination) of its amplitudes at a small number of earlier instants. The coefficients, or weights, that give the best estimate of the true speech wave can then be mathematically converted into an estimate of the amplitude spectrum. For the analysis of speech linear predictive analysis is particularly appropriate because it is mathematically equivalent to treating the vocal tract as a pipe of varying circular cross section, or in other words as a sequence of resonant cavities. The model is quite faithful for non-nasalized, voiced speech. Because it is a model of the vocal-tract resonances and not of vocal-cord vibration, the linear-prediction spectrum is smooth. None of the pitch harmonics are in evidence. Consequently the formant structure of the speech wave, which is important for speech recognition, is brought clearly to the fore.

During the comparison, or template-matching, stage the phonological structure of a word can be exploited in an indirect way. A spoken word consists of a sequence of vocal gestures, which gives rise to a time-varying pattern of sound. The parts of the sound pattern rarely have the same durations in different utterances of the same word, but their sequence is more nearly constant. For example, the word "fable" begins with an "f" noise, which is followed by a pattern of moving formants that show the lips opening out from the "f" and closing again for the "b" while the tongue is moving through the first vowel; next there is a "b" lip closure, and finally there is another pattern of spectral motion as the lips open and the tongue moves into the final "l." On different occasions the timing of these patterns may vary considerably, but they must all be

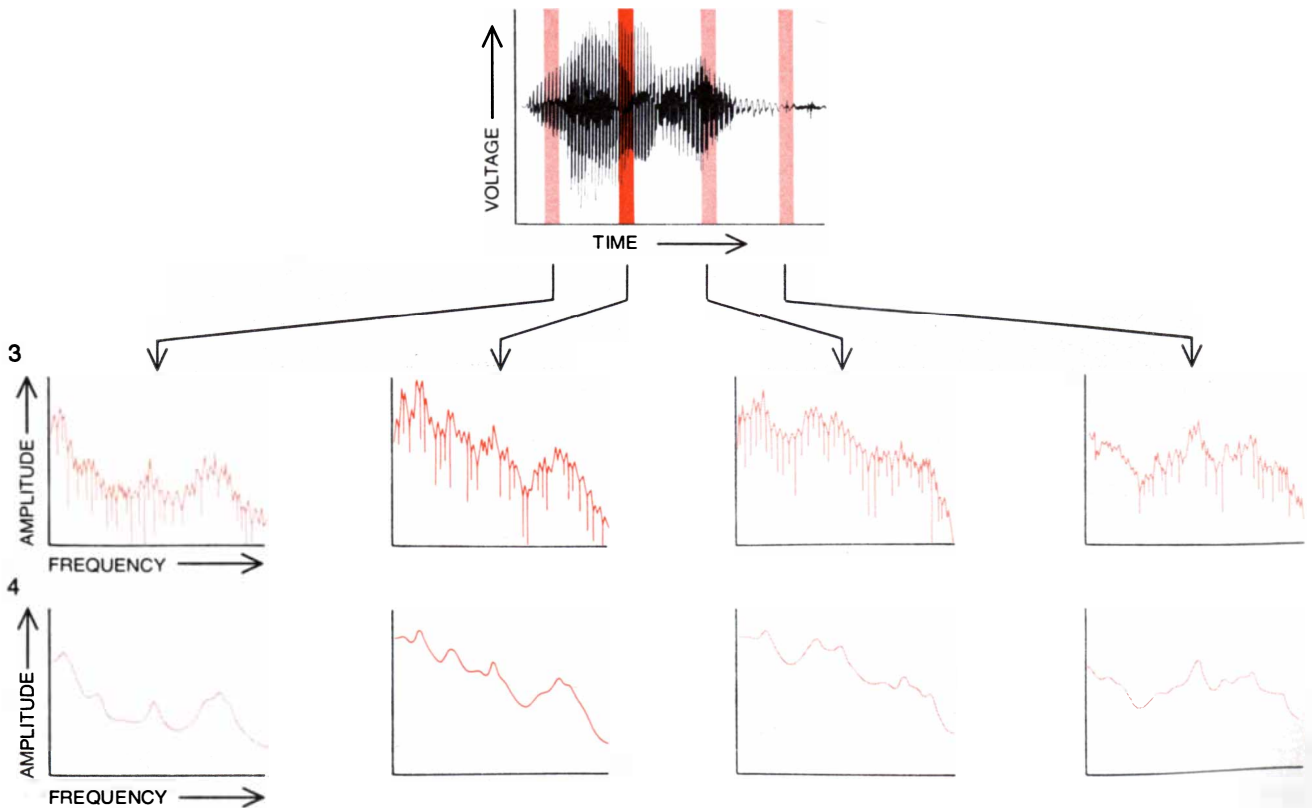
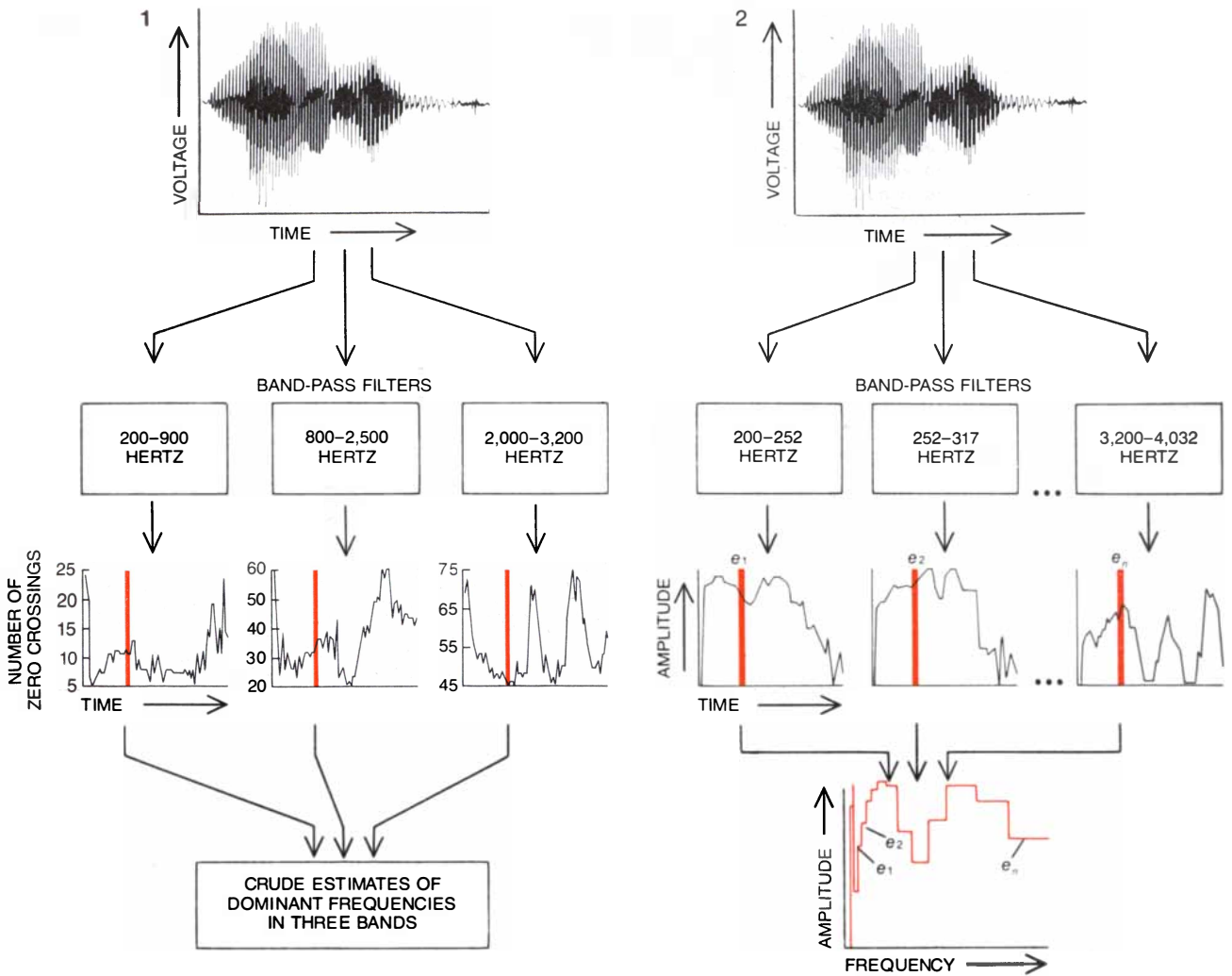
present in the described order if the utterance is to count as a reasonable rendition of the word "fable."

Because of differences in timing, the various parts of a word may be badly out of alignment with the corresponding parts of the template it is to be matched against. Since the order of events is fairly constant, the misalignment can be corrected by stretching the template in some places and compressing it in others so that a mathematically optimum match is found. Nonuniform temporal alignment is accomplished by means of a procedure called dynamic programming. Dynamic programming was developed by Richard E. Bellman of the University of Southern California School of Medicine for solving problems in the design of servomechanisms. It is a technique for mathematical optimization that is often carried out with the aid of a computer, but it should not be confused with computer programming itself.

Comparison implies some estimate of the degree of similarity between the sound of the input and the sound represented by the stored template. The final aspect of processing common to all word recognizers is a decision strategy, which is usually based on a statistical measure of closeness of fit. Each template is assigned a point in an abstract space; the position of the point is defined by the spectral characteristics of the template. The utterance to be classified is represented as a point in the same space. The recognizer calculates the distance in the space between the utterance and each of the templates. It then picks either the template closest to the utterance or the equivalence class of templates that is closest to the utterance in a statistical sense.

The performance of automatic recognition systems in identifying isolated words is poor compared with that of people. Even for the most powerful word recognizers the number of errors rises rapidly as the vocabulary increases to more than a few hundred words. The error rates get worse still when unknown speakers and acoustic conditions are introduced. In a recent experiment isolat-

METHODS OF ESTIMATING the amplitude spectra of short intervals of a word (here the word "language") all seek to highlight linguistically relevant information in a computationally efficient way. Zero-crossing counts exploit the fact that as the frequency increases, the number of times the voltage analogue of the acoustic signal changes its sign increases as well. In the band-pass-filter method the signal is divided into several frequency bands and the amount of energy in each band is measured. These measurements yield an amplitude spectrum for the interval. The Fast Fourier Transform is a general, computationally efficient algorithm for estimating the amplitude spectrum of the signal from its time-domain waveform. It is one of several ways of computing filter-bank information in digital form. The rough appearance of the spectrum is caused by pitch harmonics or other fine structure in the spectrum. The fourth method of spectral estimation, called linear predictive analysis, employs a model of the vocal tract to generate successive frequency spectra. Its advantage is that a smooth, continuous spectrum is generated for each sample. The spectra in dark color are all constructed from the same interval of the time-varying signal. Several other methods of spectral estimation are also in use.



ed words from a 26,000-word vocabulary were spoken by a variety of speakers unknown to the listener; the words were identified with an error rate of less than 3 percent. Human word-recognition abilities are also remarkably tolerant of background noise: conversation can be understood even at a noisy party. No existing recognition system can approach this level of performance.

In attempts to recognize continuous speech the disparities between human and computer performance are even more evident. People generally find it easier to recognize words in context, but for an automated system the recognition of fluent speech is far more difficult than the recognition of words in isolation. One of the crucial problems is coarticulation, which causes the blending at the boundaries between words and makes the spectral patterns to be recognized highly complex and unstable. In fluent

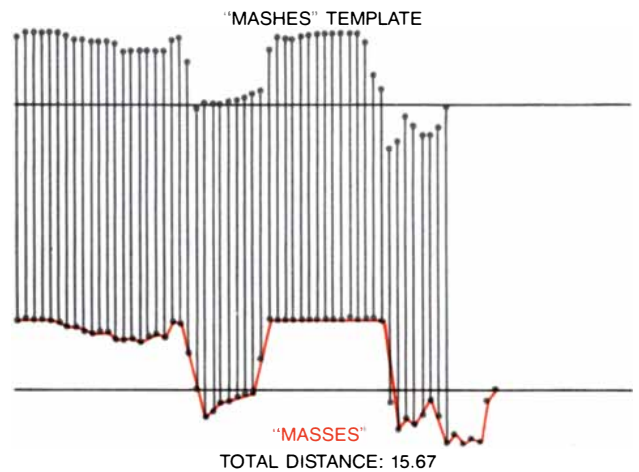
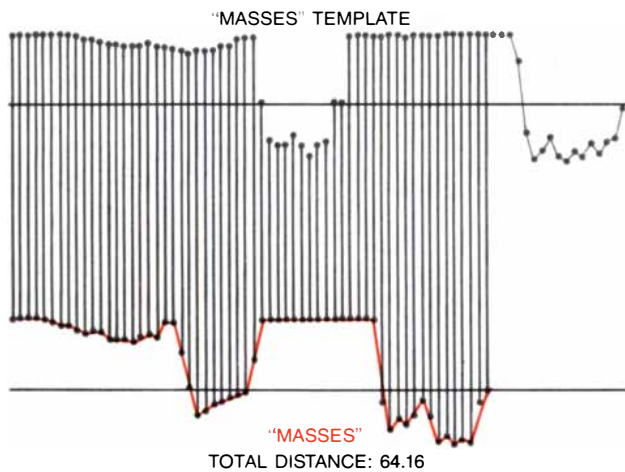
speech there are no clear acoustic signs of word boundaries and direct template-matching becomes extremely difficult. In essence every template must be aligned with every possible interval of the utterance by means of a variant of the dynamic-programming method.

The computational burden is somewhat reduced by the requirement that the intervals be contiguous, so that the end of one word meets the beginning of the next. Still, the combinatorial complexity of the process increases too fast for it to be considered a practical solution to the general problem of recognizing continuous speech. Direct template-matching can be useful only where the range of possible utterances is small. With present technology the technique can work in real time (that is, as fast as the utterance is spoken) for sequences as many as five words long, drawn from a vocabulary of about 20 words.

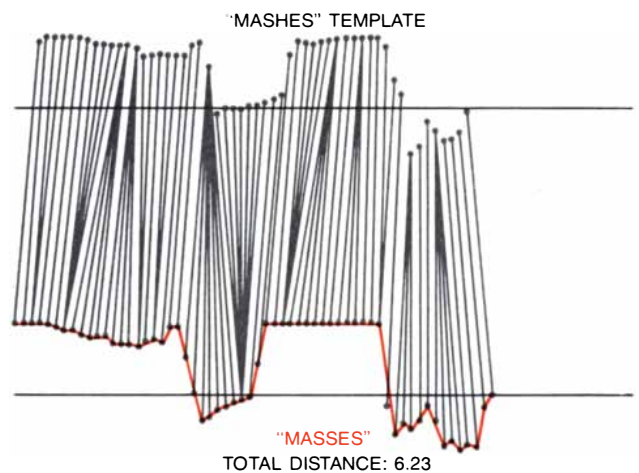
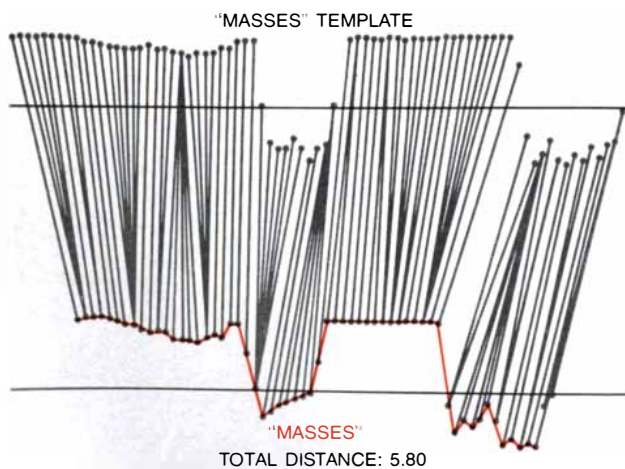
Instead of looking for every possible pattern everywhere in the signal, a continuous-speech-recognition system can search for linguistic units in a more constrained way, such as in sequence from the beginning of the utterance to the end. The speech signal is divided into intervals that correspond to specific acoustic patterns, and the intervals are classified in a way that matches the categories of a potential linguistic message as closely as possible. We shall call such techniques segmentation and labeling. The processes of segmentation and labeling can be carried out in many ways, and the intervals to be found can correspond to words or to smaller linguistic units such as syllables, phoneme pairs or phonemes.

The easiest way to achieve automatic segmentation and labeling is to require the user to pause briefly between words. The pauses that appear as intervals of

DIRECT MATCHING



MATCHING BY DYNAMIC PROGRAMMING



COMPARISON STAGE of word recognition is carried out by compressing and stretching stored templates according to an optimization process called dynamic programming. For each stored template, dynamic programming seeks to associate every frame of the input word with some frame of the template in such a way that a distance measure of overall fit between the input and the template is minimized. The nonuniform time alignment of the stored template with

the spoken word allows for variations in the rate of speech and in the relative lengths of the vowels and consonants in a word. Here matching the templates (black) to the input (color) without dynamic programming yields a misidentification, indicated by the distance scores, that is corrected when the compression and expansion procedure is applied. Dynamic programming is often done with the aid of a computer but should not be confused with computer programming.

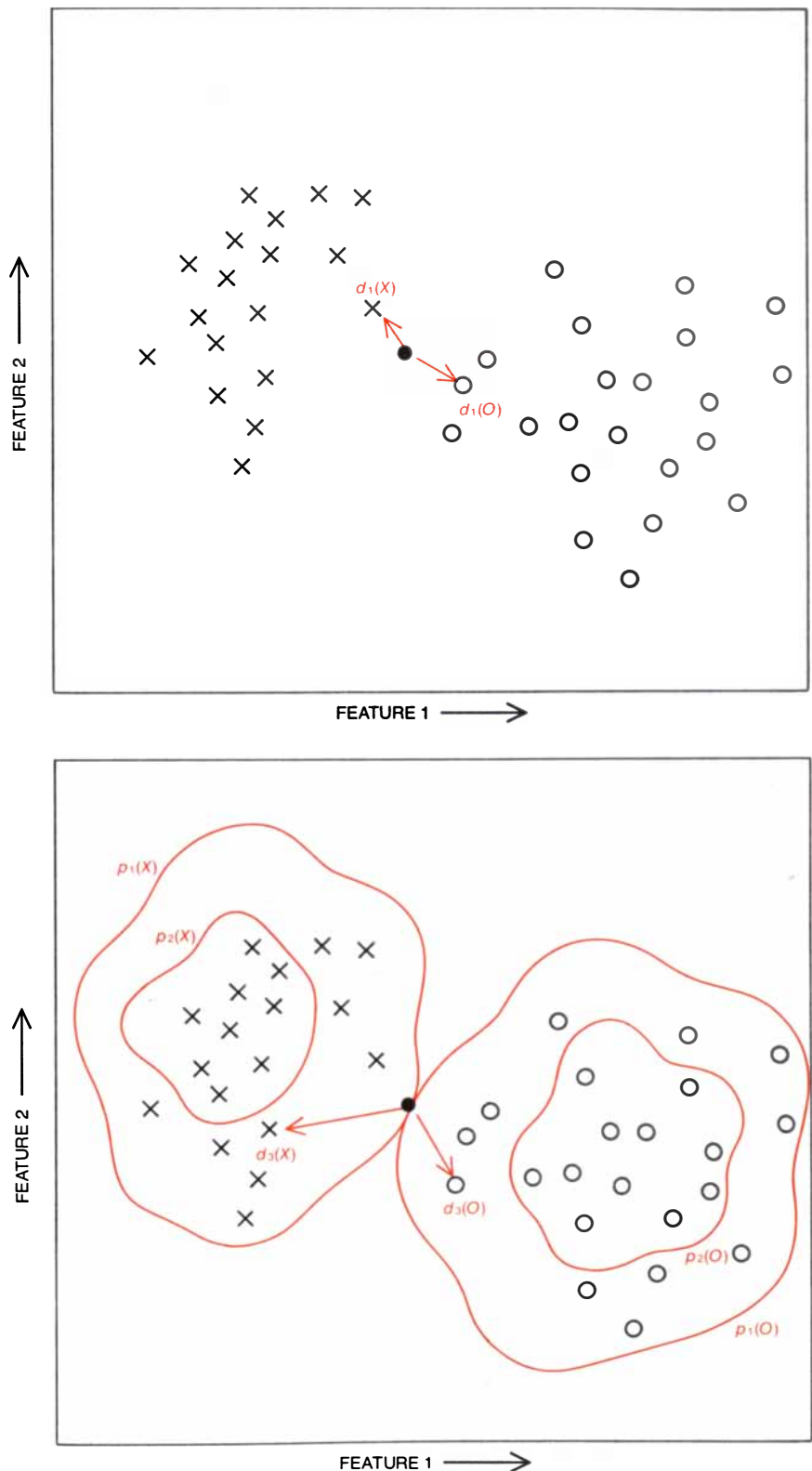
low sound energy are a reliable indication of word boundaries. Once the words have been segmented they can be analyzed independently. Although this method works well, it does not really address the question of fluent speech recognition. Other methods of segmentation and labeling are available.

Discontinuities in the spectrum, peaks and valleys in the energy of certain frequency bands and other acoustic signs provide clues to articulatory events: the closing or opening of the vocal tract or the beginning or ending of laryngeal vibration. This suggests that segmentation and labeling might be carried out on the basic phonological units of which words are constructed.

Blending and the diffusion of acoustic information across boundaries affect the acoustic shape of the smaller speech units even more than they affect words. As a result such units are difficult to identify by template-matching, and segmentation errors would probably be at least as frequent for the smaller speech units as they would be for words. Nevertheless, there may be reason to favor segmentation into smaller units as the vocabularies of speech recognizers become larger.

There are some 300,000 words in English, far too many for all of them to be tested by template-matching. Moreover, it is difficult to allow for the effects of blending at the boundaries of words when word templates are employed. English syllables number some 20,000, which is still too many for them to be identified easily and reliably. In addition the effects of blending at boundaries are even more disruptive to template-matching with syllables than they are with words. In contrast, there are only about 40 English phonemes (basic linguistic elements such as consonants or vowels), and the phonemes can be further decomposed into about a dozen phonological features that specify distinctive characteristics of vocal-tract shape and larynx control. Such features can also be combined directly into syllable-like units. As the set of linguistic units is reduced in number, however, the relation of the units to patterns of sound becomes more abstract, more complex and less well understood. Segmenting and labeling such small speech units by currently available techniques leads to high error rates. Still, if constraints imposed by the linguistic code can compensate for the errors, or if more reliable methods of analysis can be found, the small number of the basic phonological units will give them a decided advantage as the fundamental elements of a recognition system.

There is one difficulty that is common to all segmentation and labeling procedures: the probability of error is much higher in making a number of independent classifications than it is in making



CLASSIFICATION OF AN INPUT SOUND consists in finding the shortest distance in a space of acoustic features from the input (represented by a dot) to a stored template or class of templates (represented by X's and O's). The simplest decision strategy picks the closest template (upper graph), and so the input is classified as the sound "ah" (an X). When several templates represent linguistically equivalent sounds (as when the computer must recognize the voices of several speakers), the decision strategy may take account of entire classes of templates. One method calculates the distance from the input to the third-nearest neighbor in each class (lower graph); here the input is classified as the sound "aw" (an O). Under certain conditions it is possible to draw equal-density contours along which the number of template samples per unit area is constant. The highest-density contour that passes through the input can then be found; because $p_1(X)$ is greater than $p_1(O)$ the input is classified as the sound "ah" (an X).

a single classification. In a three-word phrase, even if the probability of recognizing the correct word in any given position is .8, the probability of recognizing the entire phrase correctly is only about one-half ($.8 \times .8 \times .8$).

One way of offsetting this effect is to introduce constraints imposed by the linguistic code, such as allowable sequences of words in a sentence or allowable sequences of syllables in a word. An area of mathematics called formal-language theory provides several methods for specifying and using such constraints. By applying some of the elementary principles of formal-language theory it is possible to write

precise and efficient descriptions, or formal grammars, of linguistically possible sequences of sounds and words. One can also write computer programs that utilize these grammars to recognize formally correct linguistic sequences.

One simple way of exploiting grammatical structure makes use of a mathematical construction called a state diagram. A state diagram defines every possible sentence the machine can recognize. Each path from the starting point of the diagram to the end points represents an acceptable sentence. From acoustic measurements the recognizer assigns a probability to each transition in the diagram. A probability can then

be calculated for each path by forming the product of the probabilities of all the transitions that make up the path. The sentence chosen is the one represented by the path with the highest probability. This technique can significantly reduce the error rate in sentence recognition: it can choose a word with a relatively low probability in a given position in order to enhance the likelihood that the overall transcription is correct.

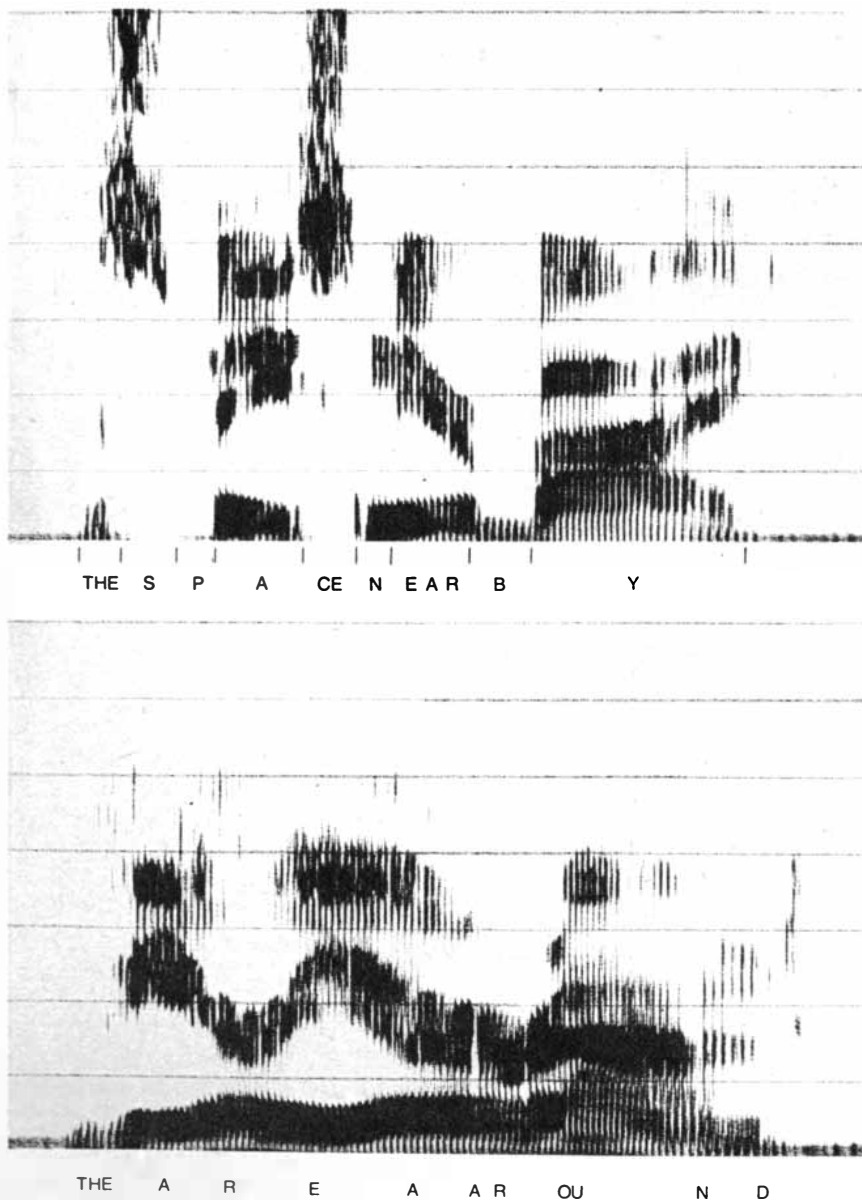
Such a reduction in the error rate was demonstrated in a phoneme-based system for the recognition of fluent Japanese, which was tested at Bell Laboratories and at the Nippon Telegraph and Telephone Electrical Communication Laboratories. The segmentation and labeling of phonemes was correct only 60 percent of the time. Syntactic processing, however, led to a 70 percent success rate for the recognition of sentences with an average length of 25 phonemes. Although 70 percent recognition is not adequate for reliable communication, the result is remarkable in view of the small probability of finding a correct sentence without syntactic processing: it comes to about one chance in three million.

A state diagram can also improve the efficiency of continuous-speech recognition by nonlinear time alignment. Instead of matching every template to every interval in the input sentence, the recognition system tests only those templates that fit admissible sequences described by the state diagram. This procedure eliminates much wasted computation, since only a small subset of the words in the vocabulary can appear at a given position in a sentence. A device employing syntax-directed time alignment can recognize connected sentences of more than 20 words composed from a vocabulary of more than 100 words.

So far we have described the phonological symbols that correspond to the acoustic reality of speech and the grammatical organization of the symbols into words and phrases. These form the linguistic code of speech. The purpose of the linguistic code is to convey meaningful messages: semantic information. Hence semantic information imposes additional constraints on the way the symbols of a language can be combined to form messages.

A machine that processes the semantic information encoded in speech attempts a much more complex and subtle task than a machine that merely recognizes words. In order to deal with meaning a machine not only must recognize acoustic patterns but also must manipulate abstract representations of reality. In other words, it must simulate at least some important aspects of human intelligence.

At Bell Laboratories we have incorporated a rudimentary semantic proc-



SEGMENTATION OF ACOUSTIC PATTERNS into words or other linguistic units is difficult because of the temporal smearing of speech sounds. Certain sounds, however, exhibit more spectral discontinuity than others. The alternation of the consonants and vowels in the phrase "The space nearby" shows up in relatively clear discontinuities. The smooth sequence of blended vowels in the phrase "The area around" makes segmentation more difficult.

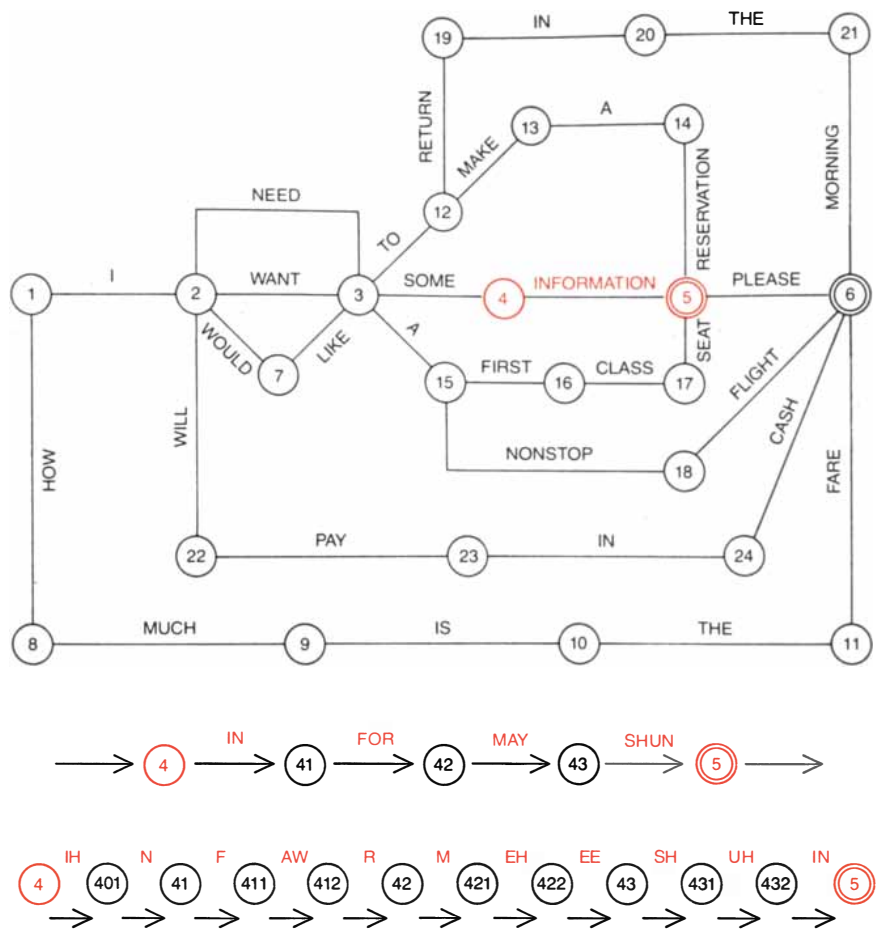
essor in a system designed to emulate the entire process of human speech communication. The user communicates with the system by telephone. The computer, which is intended to function as an airline ticket agent, responds in a synthetic voice. The integration of the necessary functions into a single device has enabled us to study the interaction of the subsystems and their control.

As a complete simulation of human communication the Bell Laboratories machine is the most advanced system known to us. The individual components, however, are less advanced than those of experimental systems in other laboratories. There are speech-recognition systems that work with vocabularies much larger than the 127 words our machine recognizes, and there are systems with a more flexible syntax. There are more sophisticated semantic processors that accept typed input instead of speech. There are processors that respond faster than ours does. A question that is asked in 10 seconds receives a reply on our system after about 50 seconds. We hope to improve the performance of all the building blocks of our system.

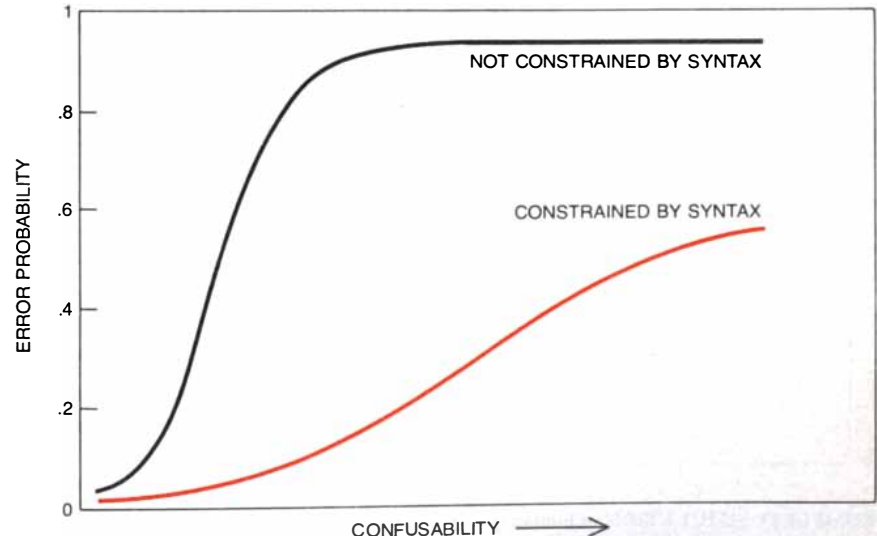
In the airline-information system the acoustic processor and the syntactic parser are coupled, so that the acoustic processor tests each hypothetical word-identification made by the parser for agreement with spectral information. The rest of the system, with the exception of two memory units that are shared by all the components, is devoted to semantic processing.

The semantic processor incorporates a world model, whose state can be changed as a conversation progresses, and a memory module, which cannot be altered. The world model is based on a set of concepts, each of which can take on a number of values. Among the concepts are "destination," "departure day" and "departure time." During a particular conversation these categories might be assigned the values "Boston," "Tuesday" and "5:00 P.M.," whereas another state of the world model might correspond to the values "Chicago," "unknown" and "unknown." The semantic processor determines a new state from its present state, from the words in the input sentence and from the transitions in the state diagram that were employed in generating the sentence. The need for the latter two sources of information reflects the fact that semantic content is a function both of the words and of their relations in a sentence.

The memory units store two kinds of information: facts and procedures. Facts are of two types. Airline schedules are stored as a portion of the *Official Airline Guide*, but relations among the concepts in the *Airline Guide* must also be stored. If the system is asked the elapsed time of a flight, it can calculate



FINITE-STATE GRAMMAR is computationally the most straightforward means of imposing syntactic (or word-order) constraints on the recognition of sentences. The grammar diagrammed here would force the computer to classify every sequence of acoustically possible words as one of the 26 sentences that can be traced through the state diagram, starting at state 1 and ending at state 5 or 6. For instance, one possible sentence is "I would like a first-class seat, please." The principles of the diagram can also be extended to levels of analysis lower than the level of the word, such as the syllabic and phonemic analyses of the word "information." The grammars of experimental recognition systems allow for billions of sentences.



CONFUSABILITY OF A SPEECH SIGNAL is a complex function of the size of the input vocabulary, the acoustic similarity of the elements to be discriminated, the number of speakers to whom the system must respond and the amount of noise in the communication channel. Errors tend to become more frequent as confusability increases. Syntactic constraints can significantly reduce the effect. This error pattern is as true for human listeners as it is for machines.

the time from the listed departure and arrival times; for it to do so, however, the time zone of each city must be available. (In the *Official Airline Guide* all times are local.)

Procedures are special-purpose programs that use stored facts to derive new information from an input and from the current state of the world model. For example, one program is a perpetual calendar, which can find the day of the week for any given date. The conversion is needed because a question may specify only a departure date, whereas the *Official Airline Guide* is organized by day of the week.

When an internal instruction calls for a reply to the user, the system activates a linguistic encoder. The semantic analyzer tells the encoder what concepts are to be communicated from the world model. Then the encoder retrieves grammar and vocabulary from memory and transforms the concepts into a sequence of symbols. The speech synthesizer

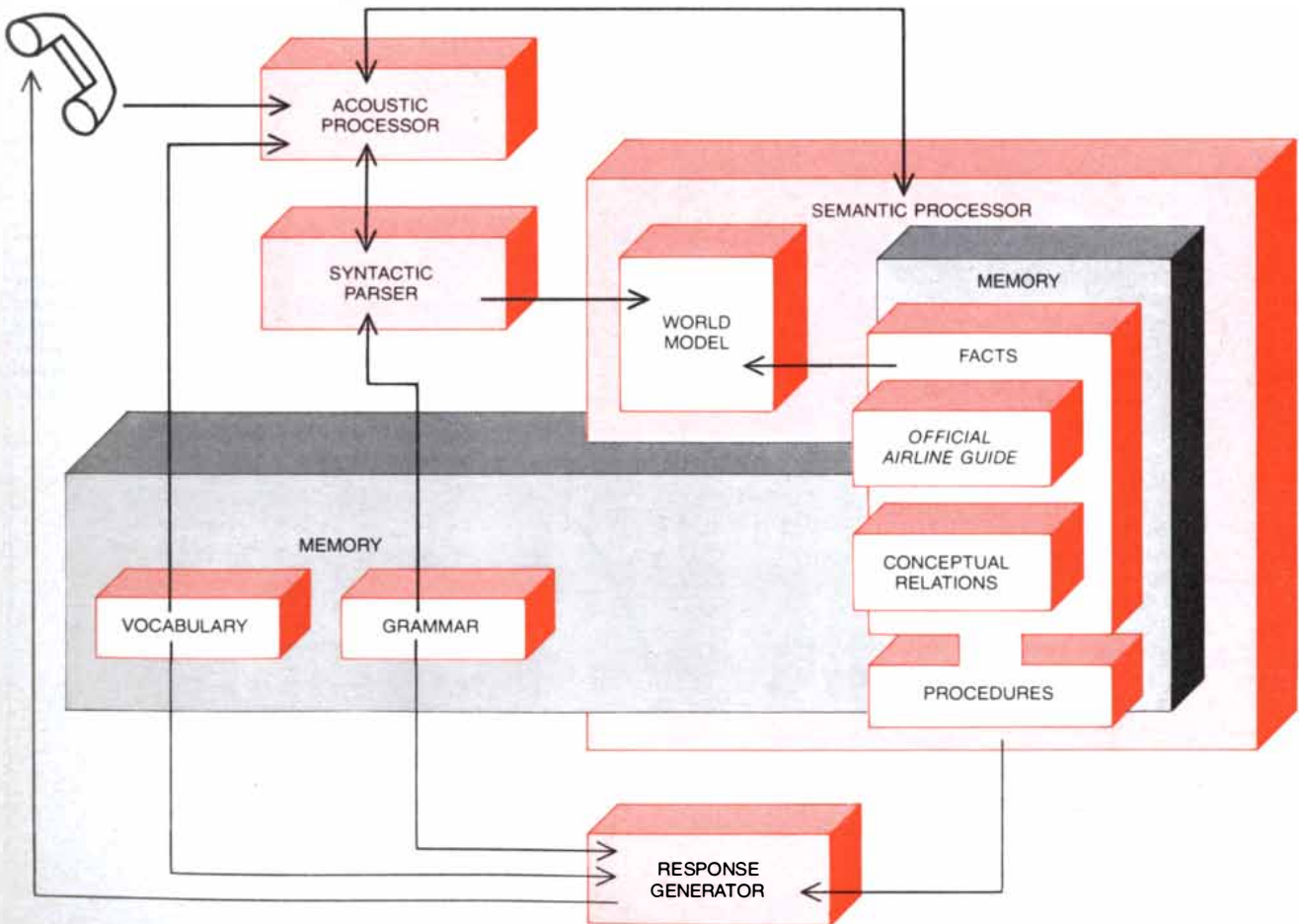
transforms the sequence into speech [see "The Synthesis of Speech," by James L. Flanagan; *SCIENTIFIC AMERICAN*, February, 1972].

In what ways can the art of speech understanding be advanced? We see two basic aims. For the near term it is important to seek a better grasp of the fine structure of speech communication. This should include detailed information about the kind of signal analysis done by the human ear and a better understanding of the relation between sound symbols (such as phonemes and syllables) and actual sounds. More efficient ways of exploiting this information must be developed and incorporated into recognition systems.

For the long term several areas of investigation may bring significant advances. We have stressed that the speech code includes a number of coexisting kinds of structure, such as phonology, syntax and semantics. A general theory

of such complex codes is needed, particularly so that the interactions of the levels can be coordinated and controlled. It is also desirable to gain a better understanding of the processes by which people acquire a language. Although present speech recognizers are "trained," the training is rudimentary and cannot be altered through "experience." We believe this lack of adaptive abilities is a serious disadvantage. The best design strategy is not to program a computer directly with the wealth of descriptive detail that constitutes a natural language but rather to give it the basic set of expectations and abilities that are needed to learn a language.

It is hard to predict how well these investigative strategies can ultimately succeed in approximating natural speech communication. Whatever the rate of progress, this goal will continue to be pursued. Some success is guaranteed, and wisdom will be required in its application.



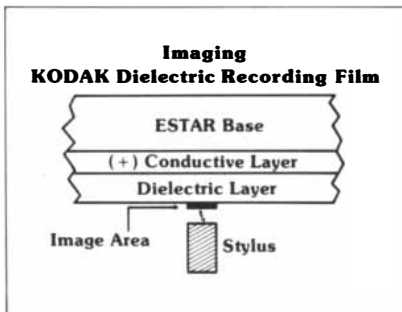
COMPLETE SIMULATION of human speech communication is attempted by an automated system constructed by the authors and their colleagues at Bell Laboratories. Functional relations among major parts of the system are shown here in a block diagram. The user asks for information about airline timetables over the telephone, and the computer replies in a synthesized voice. Heavy arrows trace the flow of information related to speech recognition. The genera-

tion of a response is traced by lighter arrows. Memory modules concerned with semantic processing include facts and procedures related to flights and reservations. Nonsemantic memory stores vocabulary templates and grammatical rules used both in speech recognition and in speech synthesis. The semantic processor also includes a world model, which is constantly updated with data based on the user's questions and on the information in the semantic memory.



Nice film without silver

Silver had better be saved for such ornaments of high civilization as elegant teaspoons, radiography of fractures, and the amplification of a light photon's kick by 10^9 .



When energy with which to write in tiny dots can be supplied like this, you don't need that kind of amplification. Estar base is nicer than paper for car-

rying the conductive layer and the dielectric layer that receives the toner to visualize the electrostatic image. (No light sensitivity, no chemicals.)

In the great hunt for formations likely still to be holding hydrocarbon treasure, the geophysical industry's plotters have been drawing pictures something like the young lady is holding, but on dielectric recording paper. Paper gets pretty conductive when the humidity is up. That makes it hard to hold a fine-structured, crisp, electrostatic image. Estar base doesn't have that problem. And the film is quite transparent, not just translucent. Therefore that fine detail can be reproduced in the diazo machine

without the services of somebody's copy camera.

On the chance that there are instrument people quite outside the geophysical industry who haven't heard about Kodak dielectric recording film SO-222 and are feeling some excitement, David R. Gadbaw of our Graphics Markets Division is standing by with a fresh supply of publication G-117SA. To obtain your copy, drop him a note at Kodak, 343 State Street, Rochester, New York 14650.



© Eastman Kodak Company, 1981

SCIENCE AND THE CITIZEN

Overestimated Undercount

The U.S. census of 1980 has been described as the largest sociological research project ever undertaken. It employed more field-workers, cost more money, generated more publicity and gathered more information about more people than any other census in the nation's history. It was also the most controversial survey of its kind to date. The outcome will determine not only the constitutionally mandated apportionment of congressional representation but also the allocation of a record \$50 billion per year of Federal and state revenue-sharing payments.

The issue that provoked the most controversy was the size of the undercount: the number of people missed on Census Day (April 1, 1980) and their geographic distribution. In the past three censuses the margin of error has remained more or less constant in absolute terms, and a comparable discrepancy was expected in 1980. Moreover, it seems that disproportionate numbers of those missed by the census enumerators were poor people, blacks, members of other minority groups and residents of inner-city neighborhoods. Because other estimates of the population are considered more reliable than a direct count, pressure has been brought to bear on the Bureau of the Census to make adjustments in the final census figures.

Preliminary study of the raw census data now suggests that the 1980 census was remarkably complete: the undercount is turning out to be much smaller than had been projected on the basis of the past three censuses. According to one demographer, Nathan Keyfitz of Harvard University, the first count of the 1980 census "makes obsolete nearly everything that has been said about the undercount and possibilities of adjusting. Expectations of incompleteness based on earlier experiences have to be drastically revised, and so do hopes of easily making corrections." Keyfitz explains his assessment in the first of a series of articles on "census politics" in the social-science journal *Society*.

After the 1970 survey the Bureau of the Census released data showing that the undercount that year amounted to some five million people. Vital statistics independent of the census, such as records of births, deaths and immigration, indicated that the resident population of the U.S. on April 1, 1970, was 208.5 million, whereas the census counted only 203.2 million. No official adjustment was made to the 1970 census, but the foundation was laid for demands that the next 10-year count be adjusted one way or another. In the past year or so these demands have become more insis-

tent, and they have begun to receive some support in the courts. Keyfitz comments: "Supposing that the 1980 figures would be five million or more short, and that this undercount would contain a disproportionate number of members of minority groups, the sentiment for adjustment... has seemed justified." It is this argument, however, that the preliminary counts apparently do not support.

Information available before last year's census indicated that the number of residents in the U.S. on April 1, 1980, was approximately 227 million. This total, Keyfitz says, "is about as solid a figure as can be obtained, and it is all that we have for checking the completeness of the census. If the calculated omission of each of the previous three censuses, about five million, is subtracted from 227 million, that brings the number down to 222 million, and this is what the 1980 census was expected to count." Instead the current best estimate is that the 1980 total will be close to 226.5 million, a finding that can be interpreted to mean the undercount was reduced from about 2.5 percent a decade ago to less than .25 percent in the present enumeration.

"It will be some time," Keyfitz notes, "before the analysis of this material is complete, so that we can know just how the drastic change from past censuses has come about." He believes, however, that "some features other than sheer efficiency of enumeration are present. The extensive publicity, together with the knowledge that one's town and state benefit financially through each person's being enumerated, engaged civic loyalty in a way that increased the response."

The same factors could have resulted in overenumeration, as could the inclusion in the census totals of illegal or undocumented aliens. Indeed, the mere fact that the total national population seems to be given correctly is no assurance that the census was fair. Some groups, and most notably some races, may be overrepresented and others may be underrepresented. Important questions remain concerning such differential completeness. Still, the net reduction in the undercount nationally is likely to make the differentials smaller, at least in absolute numbers.

Keyfitz' article concludes: "Given the superior quality of this census, in which the undercount is barely large enough to be estimated with any degree of confidence, it is ironic that 1980 is the first year in which the courts insist on adjustment. They are apparently making their decision on the basis of the large sums of money legislated during the 1970's and riding on the 1980 census, combined with the (supposedly) known undercount of 1970. This strange combination of elements out of the past

presents the Bureau with some of the most puzzling dilemmas that statisticians have had to face."

Social Disease

Viruses that cause tumors have long been known in animals, but there are only two human cancers for which a viral origin has been fairly well established. They are Burkitt's lymphoma, a malignancy of the lymphoid tissues, and nasopharyngeal carcinoma; the agent implicated in inducing both of them is the Epstein-Barr virus. For some time an infectious and perhaps viral origin has also been suspected for Hodgkin's disease, another lymphoma. The evidence for this interpretation is largely epidemiological, that is, it comes from the analysis of patterns of incidence and prevalence. It now appears that the demographic patterns resemble those of another disease whose viral cause was originally deduced on epidemiological grounds: poliomyelitis.

The epidemiology of poliomyelitis in the first half of the 20th century was highly distinctive. The age of peak incidence seemed to be delayed as living conditions improved. Increased risk also seemed to be associated with higher social class and with smaller family size. The significance of these observations became apparent when it was understood that poliomyelitis is a rare, grave consequence of a common infection and that the likelihood of contracting the disease increases with the age at which a person is first infected.

Similarities between the epidemiology of Hodgkin's disease today and that of poliomyelitis in the prevaccine era were noted in 1977 by Nancy Gutensohn and Philip Cole of the Harvard School of Public Health. The analogy suggested to them that Hodgkin's disease might also be an age-related consequence of a viral infection. If it is, the incidence of Hodgkin's disease among young adults should be associated with factors in the childhood environment that tend to delay exposure to infectious agents. Gutensohn and Cole have now tested the hypothesis in an epidemiological survey. They report the results in *The New England Journal of Medicine*.

Gutensohn and Cole interviewed 225 patients whose Hodgkin's disease had been diagnosed between the ages of 15 and 39. All the subjects lived in a defined area of eastern Massachusetts. For each patient two people of the same age and sex who lived in the same area were chosen at random as control subjects. Gutensohn and Cole obtained information on topics such as number of brothers and sisters, birth order in the family, number of neighborhood playmates,

THE TASTE BEYOND 12-YEAR-OLD SCOTCH

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.



history of infectious diseases, type of housing and parents' social class (mother's education and father's occupation).

On the basis of the data gathered the patients and controls were placed in categories and a crude relative incidence was calculated for each category. For example, if the risk of contracting Hodgkin's disease was taken as being 1 for only children and for those with one sibling, the risk for people with five siblings was found to be only .48. Gutensohn and Cole then adjusted each crude relative incidence by compensating for "confounding" factors. The adjustments were necessary because not all the variables were entirely independent of one another. Family size is correlated to some extent with type of housing, religion, number of playmates and other factors. Compensating for the influence of these correlations raises the five-sibling relative incidence to .56. The implication is that the difference between 1 and .56 reflects a specific effect of family structure. Similar effects were found for type of housing (almost twice as many Hodgkin's-disease patients as controls had lived in a one-family house), number of playmates (the more playmates, the lower the incidence of Hodgkin's disease) and social class (the risk increased with maternal education and paternal occupation level).

One of the most intriguing findings was a correlation between Hodgkin's disease and infectious mononucleosis. Twice as many patients (14 percent) as controls (7 percent) had a history of infectious mononucleosis, a usually minor disease that is common in teenagers and young adults but is rarely seen in children. What makes the correlation particularly noteworthy is that infectious mononucleosis is caused by the Epstein-Barr virus. Moreover, mononucleosis fits the epidemiological pattern that was seen in poliomyelitis and is suspected in Hodgkin's disease: early Epstein-Barr infection does not give rise to significant symptoms, so that mononucleosis is commoner in those with a more protected childhood.

Of course, the discovery of a correlation does not necessarily indicate a causal connection. Still, even after controlling for factors, such as family size and housing, that Hodgkin's disease apparently has in common with mononucleosis, the data show a significant association between mononucleosis alone and Hodgkin's disease. By itself a history of mononucleosis was enough to produce an adjusted relative incidence of 1.8. Earlier studies had shown a similar association between the two diseases.

Gutensohn and Cole conclude that, taken all in all, the evidence supports the hypothesis that Hodgkin's disease is an unusual age-related response to a virus or a group of viruses. Perhaps the infection is mild and brief in children but persists in older people; such a chron-

ic infection might somehow trigger the cancer. Among possible causal agents the Epstein-Barr virus is clearly a leading candidate, not only because of the association with mononucleosis but also because elevated levels of antibodies to the virus have been reported in Hodgkin's-disease patients as a group.

On the other hand, Epstein-Barr viral DNA has not yet been found in tumor tissue from Hodgkin's-disease patients (whereas it has been found in Burkitt's lymphoma and nasopharyngeal-carcinoma tumors). Moreover, not all Hodgkin's-disease patients have high levels of antibodies to the Epstein-Barr virus. And in spite of the clear association between mononucleosis and the cancer, only a seventh of the patients in the Gutensohn-Cole study had a history of mononucleosis. Thus even if the viral hypothesis for Hodgkin's disease is correct, viruses other than Epstein-Barr may be implicated.

Bare Bottom

At the most fundamental level the properties of matter are sufficiently few that the discovery of a new property cannot pass as routine, even when the discovery has long been expected. Evidence for such an expected but nonetheless exciting discovery has recently been derived from experiments done with a particle accelerator at Cornell University. The property of matter that seems to have appeared is called naked beauty or bare bottom.

The antecedents of the recent experiments can be traced back to 1974; it was then that another new property, charm, was first observed. The existence of charm had been predicted a decade before, but the discovery could hardly be called expected. The first evidence for the existence of charm was the particle called psi (or J), which was discovered simultaneously but independently by workers at the Stanford Linear Accelerator Center and at the Brookhaven National Laboratory.

The psi is a hadron, one of the particles that are made up of the more fundamental entities called quarks. Before the discovery of the psi three kinds of quark, labeled up, down and strange, could account for all known hadrons. A hadron can be made by combining quarks in either of two ways: three quarks can be bound together to make a baryon or a quark can bind to an antiquark to make a meson. By the early 1970's hadrons corresponding to all the allowed combinations of the original three quarks had been observed. Hence when the psi appeared, there was no vacancy for it in the table of quark combinations. It was immediately proposed that the constituents of the psi include a new kind of quark; since a fourth quark bearing the property of charm had already been suggested, the charm label was adopted.

The psi is now known to be a meson made up of a charm quark and a charm antiquark. It is closely analogous to the artificial atom that can be made by binding an electron to a positron, or antielectron. In each case there is a spectrum of energy levels that correspond to various modes of motion of the two particles. At least five such excited states of the psi system have been observed.

The psi and the electron-positron atom are analogous in another way. The constituents of the atom have equal and opposite electric charges, with the result that the atom as a whole has a charge of zero. Similarly, in the psi all the properties of the charm quark and of the charm antiquark are opposite, so that they cancel exactly. The psi has zero charm: it does not overtly exhibit the very property whose existence it established.

If there is a fourth kind of quark, however, it must form combinations not only with its own antiquark but also with all the other quarks and antiquarks. For example, there must be a meson composed of a charm quark and a down antiquark and there must be a baryon composed of a charm quark and two down quarks. Combinations of this kind exhibit naked charm. The first such hadrons were observed about a year after the psi itself; they are called D mesons. Several other mesons and baryons with naked charm have since been found.

The history of beauty or bottom is closely parallel to that of charm. In 1977 Leon Lederman and his colleagues at the Fermi National Accelerator Laboratory found evidence for a very massive particle; with some 10 times the mass of the proton, it was the heaviest particle then known. The particle was named the u , and it was soon interpreted as a meson whose constituents include still another kind of quark: the beauty or bottom quark. In analogy with the psi, the u consists of a bottom quark and a bottom antiquark. The u too can be imagined as an atomlike composite with a spectrum of excited states. Several excited states have been found.

As a combination of a quark and a matching antiquark, the u can have no manifest beauty, or bottomness. If the interpretation of the structure of the particle is correct, however, the bottom quark must also form hadrons with other quarks and antiquarks, where the properties of the constituent quarks would not all cancel. The search for such particles, which would be characterized by bare bottom, has proved to be much more difficult than the search for naked charm was.

The most straightforward way of making an u particle is to arrange for an electron and a positron to collide with each other at an energy equivalent to the mass of the u . The electron and the positron are both annihilated, so that their energy becomes available

THE BEOCENTER 7000

THERE'S NOTHING REMOTELY LIKE IT

INTRODUCING THE REMOTE CONTROL BEOCENTER 7000.

It's a totally integrated FM/Phono/Cassette unit. Complete with an infrared remote control unit and microcomputer "brain" that endows the Beocenter 7000 with features and capabilities that are nothing short of a major breakthrough.

THERE'S NOTHING REMOTELY LIKE IT.

For starters, the remote control unit lets you control your music from your listening position for optimum adjustment. Just a touch controls on/off, volume, and selection of phono, cassette, or your choice of six preprogrammed FM stations; even tape record, fast forward, and rewind.

But without moving from your chair, the Beocenter 7000 lets you do so much more...

INTRODUCING COMPUTER LOGIC.

The Beocenter 7000's microcomputer "thinks" for itself. And opens up a whole new realm of possibilities.

For example, you can program it to automatically turn itself on tomorrow at 2:00 pm and record a concert from an FM broadcast.

Or you can tell the phonograph to wake you up with your favorite rhapsody at 6:55 am.

The interaction between tuner, turntable, cassette deck and timer satisfies almost any possibility.



2-WAY COMMUNICATION.

The microcomputer keeps you constantly informed on the status of the system, via an illuminated LED display.

While a digital clock gives you the time on command, the display also confirms the radio station which is playing or that the phono section is engaged.

Record or play back a tape, and the clock read-out is replaced by a precise 4-digit tape counter and recording level meter.

If a mistake is made in the programming sequence, the display even indicates that the microcomputer is patiently waiting for correct instructions.

TURNTABLE AND CASSETTE DECK: FUNCTIONAL GENIUSES.

The turntable simplifies your duties to the single touch of a button. Simply touch "Phono" and it automatically discovers if a record is on

the platter, determines its size, sets the speed and ever so gently lowers the stylus to the lead-in groove.

If there is no record on the platter, the Beocenter refuses the command and politely informs you on the display section.

The advanced cassette deck rests under the aluminum cover with the secondary control panel. It's capable of handling all forms of magnetic tape, including metal, and again, it's remarkably automatic.

Not only does it automatically sense and select the correct bias, but it also demagnetizes the head after every recording and automatically refuses any signal that would damage a tape. Truly a thoughtful, intelligent unit.

EXPERIENCE THE BEOCENTER 7000.

Granted, the Beocenter is not inexpensive. But neither is it merely a plaything. It's a totally integrated, thinking, communicating...almost breathing...audio center.



Bang & Olufsen

Bang & Olufsen of America, Inc.
515 Busse Road
Elk Grove Village, Illinois 60007
Attention: Sue Albert

Pleasant Reflections



Montreal

Quebec

Modern skylines. Underground complexes with stores, boutiques, restaurants and theatres under one roof. New life styles alongside old world charm. Come on up

Canada

SO MUCH TO GO FOR

Tourisme Québec, C.P. 20 000, Québec G1K 7X2

NOW
IN PAPERBACK

"One of the most remarkable self-portraits of a scientist that I have ever read."—Frank E. Manuel, *The New Republic*


"Dyson is that rarity, a scientist concerned about public knowledge; his intention is not to instruct us in science but to teach us how to think about it so that we might help control it.... A memoir of the momentous ideas, events, and personalities that shaped his life, his book is endlessly absorbing, brimming with provocative intellectual juices and sauces, and utterly readable."—Henry Kisor, *Chicago Sun-Times*

Disturbing the Universe

Freeman Dyson

Portions of this book originally appeared in *The New Yorker*

\$4.95

HARPER  COLOPHON BOOKS

for the creation of new particles. This was the method employed in the Cornell experiments. The electron-positron collisions took place in the Cornell Electron Storage Ring (which is abbreviated CESR and pronounced "Caesar").

The electrons and positrons in CESR orbit the ring in bunches and collide at two points on the circumference. Each of these interaction regions is surrounded by a set of nested cylindrical detectors, which count and identify the particles emitted in the aftermath of a collision. Signals interpreted as evidence for bare bottom were obtained from both detectors. The findings are reported in *Physical Review Letters* by two groups comprising some 70 investigators.

Any state of matter created from the annihilation of an electron and a positron must include equal numbers of particles and antiparticles. In the case of the psi and the upsilon this constraint is not a serious one, since each particle is itself made up of a quark and the corresponding antiquark. Hence a psi or an upsilon can be created as a single particle. A particle with bare bottom, however, can appear only in conjunction with its associated antiparticle. It is not surprising that such a pair of particles is more massive than a single upsilon; for this reason the upsilon cannot decay into a pair of bottom mesons. It turns out that the first and the second excited states of the upsilon also have insufficient mass to make up a bottom-antibottom pair. The Cornell workers found their evidence for bare bottom in the decay of the third excited state of the upsilon.

Neither the upsilon itself nor its immediate decay products survive long enough to reach the detectors that surround the interaction regions. All that can be seen directly are the stable or nearly stable particles that make up the ultimate decay products. At CESR the telltale sign of bare bottom was a sharp enhancement in the number of high-energy electrons emitted when the energy of the colliding beams was tuned precisely to the energy of the third excited state. A similar enhancement was found in the spectrum of emitted muons, which are particles similar to electrons but heavier.

From the energy, the angular distribution and other characteristics of the detected electrons, a likely sequence of events at the interaction zone could be reconstructed. When the electron and the positron annihilate each other, the resulting electromagnetic energy materializes in the form of an upsilon particle in the third excited state. The bottom quark and bottom antiquark constituting the upsilon subsequently split apart and a new up quark and up antiquark form spontaneously. The bottom quark pairs with the up antiquark, forming a B^- meson; similarly, the up quark and the bottom antiquark make up a B^+ meson. The total mass of the two B mesons

must be somewhat less than the mass of the u third excited state. The B mesons also subsequently decay. It seems they preferentially give rise to D mesons. Thus bare bottom gives way to naked charm.

Program Power

Some 200 years ago Wolfgang von Kempelen, a Hungarian inventor, toured Europe with a chess-playing machine that took on all comers and defeated virtually all of them. Called the *Maezel Chess Automaton*, the machine had many admirers, including Edgar Allan Poe, who speculated in an essay on how it worked. A few years later the "mechanism" was revealed. Inside the machine was a legless Polish army officer named Worowski, a master chess player.

The evolution of the electronic computer over the past 35 years has stimulated much honest work on the design of chess-playing machines, but until a few months ago none played as strong a game as the fake automaton. A milestone was reached in January, however, when the computer program called *Belle* demonstrated that it could play chess at the master level against strong human opposition. At the Virginia Open Chess Tournament in Fredericksburg, Belle scored four wins and one loss to tie for the fourth through the sixth places in a field of 66 players.

Belle played as if it had a chess rating of 2,300, which would put it in the top 1 percent of the roughly 30,000 Americans who compete in chess tournaments. A rating of 2,200 or higher qualifies a player as a master, but Belle has not yet actually earned that designation. An official rating is based on all the tournament games in a player's career, and after a sudden improvement it may take some time for the rating to catch up. Belle's present cumulative rating is 2,113, which is in the expert category.

Belle put on another master-level performance at a tournament in Westfield, N.J., also in January. By winning two games and drawing two the program placed second in a field of 14 masters and experts. Belle played as if it had a rating of 2,400, but the tournament was an unofficial one, meaning that the results did not affect the players' ratings. In competition against other chess-playing computers Belle has achieved preeminence. In September it won the Third World Computer Chess Championship in Linz, Austria, where 18 programs were entered.

Belle was created by Ken Thompson and Joe Condon of Bell Laboratories. The secret of its success is special hardware designed explicitly to carry out chess calculations. It is a rule of thumb in computer science that an operation can be executed fastest when it is hardwired into the computer rather than

Easy Does It



Pincher Creek

Alberta

Take it slow through the land of the big blue sky. Drink in the mountains, the meadows, the valleys, the tidy Hutterite farms with their gentle people. Come on up.

Canada

SO MUCH TO GO FOR

Travel Alberta, Dept. 009, Box 2500, Edmonton, Alta. T5J 0H4

computique

computique

computique

TEXAS INSTRUMENTS
INCORPORATED
electronic calculators

TI-30SP	17.95
TI-35SP NEW	24.95
Bus. Anal. I	21.95
BA II Exec. NEW	44.95
Bus. Card	39.95
Invest. Anal. NEW	49.95
TI-55	34.95
TI-57	54.95
TI-5100	39.95
TI-5015	64.95
TI-5040	89.95
TI-5135 NEW	79.95
TI-5142 NEW	124.95
TI-808-41 NEW	62.95
TI-58C	104.95
TI-59	209.95
PC 100C	169.95
TI 58/59 Libraries	35.00
Speak & Spell	59.95
Speak & Read NEW	CALL
Speak & Math NEW	CALL
Lang. Tutor NEW	169.95
Home Comp. 99/4	
Console only	499.95

THE HP-85!



Complete Enhancements, Peripherals and Accessories

Your **hp** HEWLETT PACKARD Headquarters

HP-33C NEW	89.95	HP-41C NEW	CALL
HP-34C NEW	134.95	HP-85 NEW	CALL
HP-38C NEW	134.95	HP-41CV NEW	CALL
HP-32E	54.95	HP-83 NEW	CALL
HP-37E	59.95		

(714) 549-7373
INFORMATION LINE

(800) 432-7066
TOLL FREE (Within CA)

(800) 854-0523
TOLL FREE (Outside CA)

BPF IMAGINATION MACHINE 495.00

apple computer II PLUS (16K) 999.95

apple // APPLE III NEW CALL

6200 NEW EXECUTIVE 99.95

5813 NEW SCI PROGRAMMABLE 34.95

5100 NEW LCD ALPHANUMERIC 89.95

5102 NEW FIN. PROG 79.95

7000 NEW MEMO WRITER 124.95

CASIO * ML 90 49.95

FX 68 27.95

F CHESS CHALLENGER "7" 99.95

SENSORY CHESS CHALLENGER "8" NEW 129.95

800 (16K) STARTER SYSTEM 899.95

ATARI 1000 109.95

CODE-A-PHONE 1750 NEW 269.95

ATARI X-1000 59.95

TIMER NEW 33.95

CONTROLLER COMMAND 33.95

ALSO: CANON, TOSHIBA, NSC, SEIKO, MATTEL, PEARL CORDER, ITT, GTE, AND MANY OTHERS.

ALL AT GREAT PRICES!!

WE WILL MEET OR BEAT ANY COMPETITOR'S ADVERTISED PRICE ON MOST ITEMS IF HE HAS THE MDSE. ON HAND. VISA, MASTERCARD, MONEY ORDER, PERS. CK. (14 WRKG. DAYS TO CLR.), COD ACCEPTED; MIN. \$4.95 SHIPPING U.S.A.; AIR ON REQST.; CAL. RES. ADD 6% SALES TX.; ALL MDSE. SUBJ. TO AVAIL.; PRICES SUBJ. TO CHANGE.

MAIL & PHONE
ORDERS ONLY

computique

WRITE OR CALL FOR
FREE CATALOG

Send orders to:
SA-A
3211 SO. HARBOR BLVD.
SANTA ANA, CA 92704
NEWPORT
(714) 549-7373

WEST LOS ANGELES
(213) 8200423
TARZANA
(213) 705-7507

MID-WILSHIRE
(213) 385-7777
LAWNDALE
(213) 370-5795

PASADENA
(213) 795-3007
BREA
(714) 990-6600

PROFESSIONAL DISCOUNTS

Call Of The Wild



Georgian Bay

Ontario

More trees and streams and lakes than the mind can comprehend. Some singularly beautiful. Others lost in the panorama of the view. Something for the senses. Come on up.

Canada
SO MUCH TO GO FOR

For information call collect: 416-965-4008

specified as part of a program. Belle's hardware enables it to examine and evaluate possible positions of the chess pieces at an average rate of 160,000 positions per second. If it spends three minutes on a move (which is the average time in tournament play), it evaluates about 29 million positions.

In choosing a move Belle makes an exhaustive search: it looks at all possible continuations of the game to a certain depth. The move chosen is the one calculated to minimize the opponent's net gain, and so the method of evaluation is called a minimax search. The possible moves are organized as a tree diagram. If Belle investigated every branch of the tree, it would not get far. The branching factor for chess (the average number of moves that can be made from a given position) is 35. An exhaustive search to a depth of only three moves for each player would require the evaluation of more than 1.8 billion positions, or 62 times the number of positions Belle examines in the time allotted in tournament play.

A mechanism called the alpha-beta algorithm substantially reduces the number of moves that must be considered, but it nonetheless yields the same move that would have been selected if the program had explored all the branches of the tree diagram. The algorithm abandons the investigation of a move when it finds the opponent can respond with a countermove that is better than the best response to a move Belle has already examined. Other mechanisms for pruning the branches of the tree diagram are also incorporated into the program. For example, Belle remembers the positions it has considered, so that it evaluates only once a position to which more than one branch leads. In addition it looks first at a direct assault on the opponent's king, since that is the best plan if there is no satisfactory defense. Any move leading to the capture of an enemy piece is also given a high priority.

Although Belle now plays at master strength, its brute-force approach to the game is not at all like that of a human master. The master usually investigates only a few moves in depth, and in most positions he ultimately chooses the first move he explores. The time spent comparing various continuations serves not so much to choose a move as it does to confirm his intuition about which move is the best one.

Many chess games can be divided into three stages: the opening, the middle game and the end game. In the opening each side seeks control of the squares in the center of the board, brings its pieces into play and finds shelter for its king. In the middle game the sides engage in close combat and work to accumulate small advantages in the configuration of the pawns, in the balance of material and in other factors that can influence the outcome. In the end game, when



Journal of Field Archaeology

Taxonomists

in all branches of botany, zoology, and medicine are needed to identify archaeological finds. The *Journal of Field Archaeology* is preparing a list of taxonomic specialists who are able to classify plant (seeds, pollen, fibers, wood, etc.), animal (insects, shells, bones, hair, teeth, leather, etc.), and human remains from current archaeological excavations. If you are willing to participate in such inter-disciplinary work, please send your name, address and the materials you are competent to identify to:

Curt W. Beck
Vassar College, Box 92
Poughkeepsie, N. Y. 12601

**IT ONLY
LOOKS
HARD...**

... but it's
really
easy!



The secret's in our kits—imported from Europe—where these historic ship models are hand-crafted in walnut and rich woods by some of the finest model makers in the world. Solid brass, cast bronze and detailed wood fittings are used throughout. The hard work is done for you—most parts are *pre-cut*. Assembly is simple as you follow step-by-step instructions and large scale plans. Everything you need to complete is included, and we guarantee satisfaction.

Even if you have never built a ship before, send for our full-color catalog. See how easy it can be to build a museum quality model you will display with pride in home or office. Only \$1.00.

Mail This Coupon Today!

MODEL EXPO, INC.
23 Just Road, Fairfield, N.J. 07006

Dept. SA-41

Please send full-color catalog; \$1.00 enclosed.

Name _____

Address _____

City _____

State _____

Zip _____

each side is reduced to one or two pieces and a few pawns, the side that stands better tries to convert its advantages into a winning position.

Belle's power of calculation puts it on an equal footing with a human master in the middle game but not in the opening game or in the end game. The program nonetheless plays a strong opening game because of the computer's capacious memory, in which 350,000 opening positions are stored. If Belle did not play the opening by rote but instead had to rely on its evaluation function, a human master would almost always take a commanding lead at the outset.

Neither the minimax search nor a huge memory is much help to the program in the end game. Thompson estimates that Belle handles end-game positions with the skill of a player whose rating is 2,000. A chief goal of the end game is to march a pawn across the board to the eighth rank, where it can be promoted to a queen. Attaining this goal often requires lengthy, intricate maneuvers that a minimax search does not reveal. In the end game, pattern recognition and long-range planning are more important than rapid calculation.

On the other hand, the end game calls for one kind of calculation that even a poor human player does routinely but that Belle is not equipped to carry out.

The calculation determines whether or not an unobstructed pawn can simply be advanced until it reaches the eighth rank. The player first counts the number of squares in front of the pawn; this is the number of moves it would take to queen the pawn if the opponent did nothing to interfere. Next the player determines whether or not in that number of moves the opponent can block the pawn's advance.

Belle has erred in many end games because of its failure to appreciate the strength of an unimpeded pawn. Such a failure was responsible for its only loss in the Virginia Open; it lost to Geoffrey McKenna, who is rated 2,222. McKenna emerged from the opening and the early middle game with a considerable advantage because he had two bishops against Belle's two knights operating on an open board, where the long range of the bishops could be exploited effectively. The program played resourcefully, however, achieved an even game and eventually gained a slight advantage. In the end game Belle brought on its own defeat. The program arranged a forced combination in which it temporarily gave up a rook but ultimately won back not only the rook but also a bishop and a pawn. The exchange left McKenna with only a king and a pawn to oppose Belle's king, two pawns and a knight. In spite of

Belle's advantage in material, however, nothing could stop McKenna's pawn from reaching the eighth rank. Belle resigned on the 69th move, seven moves after the pawn was promoted to a queen.

Roman Times

It has occasionally been suggested that without the Arabic system of numeration the modern exact sciences would be unthinkable. Roman numerals seem too cumbersome for any but the simplest calculations. In particular multiplication and division have been considered extremely awkward with large numbers expressed in Roman notation. James G. Kennedy of the Hughes Aircraft Company now points out that the clumsiness of Roman numerals has been somewhat exaggerated. Indeed, the algorithms for multiplication and division are actually simpler in the Roman system than they are in the Arabic one.

The Arabic system of numeration probably originated not in the Arabic lands but in India; its most important element, the zero, was a late addition. It is a decimal place-value system, in which each digit in a number, reading from right to left, is understood to be multiplied by a higher power of 10. The Roman system is not based on place value; instead each symbol represents a



"IN JUST 10 DAYS, I'LL SHOW YOU HOW TO DO

REAL MATH

$$\int_a^b f \quad \frac{df}{dx} \quad \sum_{n=1}^{\infty} a_n$$

ON YOUR SCIENTIFIC CALCULATOR!*

- Quick.
- Guaranteed.
- Easy.
- Fun, too!

INTRIGUED BY CALCULATORS? Then you can step up your math skills *fast!* Use my *new method* in guidebook form. It's called **CALCULATOR CALCULUS** and comes with this guarantee: *If after 10 days you're not astounded at the problems you're solving on your own calculator, return the guidebook for an immediate refund.*

But the point is - you won't want to send it back. For this is the *easiest, fastest shortcut* ever! The day you receive your copy in the mail you'll want to put it to work. It's that exciting and helpful.

My name is Dr. George McCarty. I teach math at the University of California. I wrote this guidebook to cut through the confusion. It does just that — with worked-out examples, simple exercises and practical problems — all designed to work with precision and magic on your own calculator.

POWER METHODS. Need to evaluate functions, areas, volumes — solve equations — use curves, trig, polar coordinates — find limits for sequences and series? It's all here!

If you're in the biological, social or physical sciences, you'll be doing Bessel functions, carbon dating, Gompertz' growth curves, half-life, future value, marginal costs, motion, cooling, probability, pressure — and plenty more (even differential equations).

Important numerical techniques? Those algorithms are here, too -- rational and Padé approximation, bracketing, continued fractions, Euler's method, Heun's method, iteration functions, Newton's method, predictor-corrector, successive substitutions, Simpson's method and synthetic division.

LOOK AT WHAT USERS SAY. Samuel C. McCluney, Jr., of Philadelphia writes:

"**CALCULATOR CALCULUS IS GREAT!** For ten years I have been trying to get the theory of calculus through my head, using home-study courses. It was not until I had your book that it became clear what the calculus was all about. Now I can go through the other books and see what they are trying to do. With your book and a calculator the whole idea becomes clear in a moment, and is a MOST REFRESHING EXPERIENCE. I program some of the iterative problems you suggest and it

always GIVES ME A THRILL to see it start out with a wild guess and then approach the limit and stop."

Professor John A. Ball of Harvard College (author of the book *Algorithms for RPN Calculators*) adds:

"I wish I had had as good a calculus course."

Professor H. I. Freedman of the U. of Alberta, writing in *Soc. Ind. Appl. Math Review*, states:

"There can be no question as to the usefulness of this book...lots of exercises...very clearly written and makes for easy reading."

C. B. of Santa Barbarasays: "Your book has given me much instruction and pleasure. I do not hesitate to recommend it. 'CALCULATOR CALCULUS' inspires the reader to understand everything down to the last detail. You have put your heart into the teaching."

I WANT YOU TO TRY THIS. So, if you act now, you can invest in 'CALCULATOR CALCULUS' for only U.S. \$14.95 (to USA or foreign: add \$1 for shipping, or \$4 by AIR; in Calif. add 90c tax).

***GET MY COMPLETE KIT,** with a TI-35 calculator, a 200 p. Student Math Book, AND the guidebook, ALL for \$39.95 (to USA only: add \$2 for shipping, or \$5 by AIR; in Calif. add \$2.40 tax). As pennynwise Ben Franklin said, "An investment in knowledge pays the best dividends." GET STARTED NOW — Tax deductible for professionals.

NO RISK WHATEVER! Send for it today. Be sure to give me your complete mailing address with your check or money order. If you want to charge it (Visa or MC), tell me your card no. and exp. date. *Prompt shipment guaranteed.*

Thank you!


EduCALC Publications, Box 974, Dept. A-4, Laguna Beach, California 92652
 In California, call 714-497-3600; elsewhere
 TOLL FREE 24-hour Credit Card orders: 800—854-0561, Extension 845; Dept. A-4

fixed magnitude regardless of its position. On the other hand, the earliest form of the Roman system apparently was decimal. In that primitive version a series of I's represented any number from 1 through 9, and a new symbol was introduced for each higher power of 10: X for 10, C for 100 and M for 1,000. The symbols V, L and D, which stand for 5, 50 and 500, are thought to have been added by the Etruscans. They make the representation of a number more compact, but they also make arithmetic somewhat more complicated.

Kennedy presents his algorithms for the multiplication and division of Roman numerals in *The American Mathematical Monthly*. In multiplication the first step is to rewrite the numbers in a simple place-value notation. Seven columns are set up, headed by the symbols M, D, C, L, X, V and I, and tallies are marked in each column corresponding to the number of times that symbol appears in the multiplicand. For example, if the multiplicand is XIII (13), one tally is marked in the X column and three tallies are marked in the I column. The multiplier is written in the same way.

The multiplication itself is done by forming partial products according to two simple rules. In most cases the partial product generated by any one tally in the multiplier is simply the set of tal-

lies that represents the multiplicand, shifted to the left an appropriate number of columns. If the multiplier digit is I, the multiplicand is not shifted at all; the multiplicand is shifted one place left for V, two places for X, three places for L and so on. The second rule, which is merely a modification of the first one, is applied only when an Etruscan character (V, L or D) is multiplied by another Etruscan character. In such cases the tallies representing the multiplicand digit are written twice in the appropriately shifted column and an additional tally is written one column to the right. Without the modified rule $V \times V$ would equal X; with the modification it is equal to XXV (25), the correct result. Once a partial product has been formed for every tally in the multiplier, the tallies in each column are accumulated and replaced by the Roman symbol at the head of the column. The resulting string of symbols is the product of the multiplication.

A procedure only slightly more complicated is needed for multiplying Roman numerals in "subtractive notation," where 4 is written IV and 90 is written XC. Some of the tallies must then be marked as negative quantities. Division problems can be solved in the same place-value format. The essence of the division algorithm is to shift the divisor to the left as many columns as possible

without allowing it to exceed the dividend. The shifted divisor is then subtracted from the dividend as many times as possible and one tally is marked in the quotient for each subtraction.

If these algorithms do not seem particularly simple, it should be kept in mind that the procedures for multiplying or dividing Arabic numbers would also be quite involved if they were written out in explicit form. Moreover, the Arabic operations require, as an adjunct to the algorithms, a multiplication table giving the 100 products of all the possible pairs of Arabic digits. No comparable table is needed with Roman numerals, where all arithmetical operations can be defined in terms of shifting rules, addition and subtraction.

The formal simplicity of the Roman-numeral algorithms does not argue for a return to Roman notation. The fact that the procedures are simpler does not necessarily mean they are easier to use. The existence of such algorithms cannot even be employed to gauge the computational capabilities of those who reckoned in the Roman system. Kennedy writes: "Arithmetic was commonly done by Romans on the abacus or counting table. What methods they used for written calculations, if any, have apparently been lost, and the methods described here may be a modern invention."

Wear the Expanding Universe on your chest.

LET THE WORLD KNOW ABOUT YOUR INTEREST IN ASTRONOMY with T-shirts and bumper stickers from the ASTRONOMICAL SOCIETY OF THE PACIFIC

Since 1889, the A.S.P. has brought together professional astronomers, amateurs, educators and interested laypeople. Although we began as a Pacific Coast organization, today our members can be found in every state and in 54 foreign countries.

Membership in our non-profit society is open to everyone with an interest in astronomy; benefits include: *Mercury* magazine, which brings you authoritative non-technical articles on the latest discoveries about the Universe; a directory of astronomy activities and services in your area; and a membership card good for discounts at planetariums and stores around the country.

For a free illustrated brochure and a discount catalog which includes many other astronomical items besides the bumper stickers and T-shirts, send coupon (below) to:

A.S.P., Dept. S.
1290 24th Ave.
San Francisco, CA 94122

Please send your brochure and catalog.

Name: _____

Address: _____

Zip _____



TURN OFF YOUR LIGHTS
TURN ON TO ASTRONOMY

ASTRONOMY IS LOOKING UP

ASTRONOMERS
DO IT AT NIGHT

SUPERNOVAE ARE A BLAST



TENBA INC. believes that there is only one way to build a bag for the serious photographer. It must be easy to work out of, light to carry, protective of its equipment, and completely reliable. We take no shortcuts, and make no compromises in quality.

Why You Should Invest In Tenba Quality

- Quick, positive, one-hand access and closure
- Everything at your fingertips
- More gear in less space
- Your interior-arranged as you like it
- Complete shock-proof padding
- Rugged dependable materials and construction
- Lightweight
- Complete weather protection

TENBA is a Tibetan word meaning *strong, unshakable* and *reliable*. From its first bag, TENBA, INC. has upheld the commitment to its name.

The goal of the TENBA Pro-pak is to facilitate you, the photographer, in reaching your objectives. We consider your photographic achievements, in part, a measure of our success.

For complete information see your dealer or write or call TENBA.

Construction

1 Weather flap over zipper 2 #10 coil nylon zipper 3 Outside raised top compartment 4 Inside top flap pocket 5 Waterproof cotton-backed vinyl liner 6 #10 coil nylon zipper 7 Welded D-ring 8-9 Box-X bar tack 10 Non-skid pad 11 Small D-rings (one in each corner, rear two are for attaching Waist Belt, and/or Back Pack harness. All four can be used to lash tripod to bottom.) 12-13 3000 lb test nylon web 14 Wood liner 15 Closed-cell ETHAFOAM 16 1000 lb test nylon web 17 1000 denier CORDURA nylon 18 Bar tacks 19 Velcro pocket closure 20 1000 lb test nylon web 21 Tourniquet buckles 22 Box-X bar tacks 23 Nylon adjustable tuck catch 24 #5 molded nylon zipper 25 Bar tacks 26 Top double stitched onto bag and ends reinforced with bar tacks 27 Velcro hook 28 Velcro loop 29 Mid divider 30 Side divider

The Quality Bag For The Working Pro © 1981 SCIENTIFIC AM

The Origin of Genetic Information

Laws governing natural selection of prebiotic molecules have been inferred and tested, making it possible to discover how early RNA genes interacted with proteins and how the genetic code developed

by Manfred Eigen, William Gardiner, Peter Schuster and Ruthild Winkler-Oswatitsch

Charles Darwin saw in the diversity of species the principles of evolution that operated to generate the species: variation, competition and selection. Since Darwin's time an understanding of molecular biology and of the geophysics and geochemistry of the prebiotic earth has been gained that would have been unimaginable in the 19th century. Does that make it possible now to follow evolution back into the era before there were organisms?

A first answer is no. The prebiotic fossil record, as far as is known, decayed or was wiped clean by later generations of life. The intellectual fossils that remain—the genetic code, the genetic messages of present organisms and the known reaction pathways of biochemistry—convey information so fragmentary that one could never describe prebiotic evolution in as much detail as, say, the evolution of primates.

Fragmentary information, however, has never been a barrier to the discovery of laws of nature. Newton discovered the universal laws of motion from observations of a few planets; Mendeleev discovered the structure of the periodic table in the chemistry of only a few elements; today physicists infer laws that describe the interactions of elementary particles from observations of small numbers of events. One does not need a detailed history of prebiotic conditions and events in order to discover the evolutionary laws that led to the first life on the earth. One must only hope there is enough fossil evidence left to guide one's thinking and demand enough predictive power of a theory to make it subject to experimental testing. In this sense the answer to the question raised above is yes: Definite statements can be made about the natural laws that governed the origin and prebiotic evolution of life.

In this article we describe what must be added to Darwin's ideas to describe evolution before organisms came into existence. First we shall show that his ideas do apply to evolution far below the level of organisms. To explain how the complexity of higher organisms and

the variety of species came about, Darwin proposed that the more complex evolves from the less complex through natural selection. Why should this principle not apply as well to the complexity of large molecules? We shall give necessary and sufficient conditions for natural selection to proceed at the molecular level. The outcome of such selection is a "regularity of events" that follows inevitably whenever certain conditions are fulfilled.

Whereas competition is the basis of natural selection among organisms, competition alone would not have worked in prebiotic times to select the fittest molecular assemblies; certain forms of cooperation were also essential. The evolutionary interplay of molecular competition and molecular cooperation reflected the need to process and utilize the first genetic information so as to stabilize and then improve it. It is impossible to re-create the actual stages of genetic improvement because enormous numbers of chance mutations were tested and discarded during early evolution. Nevertheless, one can now understand the natural laws that governed those stages. The laws can be tested in diverse ways: through experiments with bacterial viruses, through chemical studies of the components of nucleic acids and proteins and through comparative analysis of nucleic acids and proteins that have survived three or four billion years of molecular evolution.

The Earth before Life Began

Before the drama of life could unfold the stage had to be set and certain minor actors had to take their places. The stage was somewhere on the primitive earth, which had a temperature not much different from what it is today. The composition of the earth's surface was also much as it is today if one considers the mere abundance of elements, but it was vastly different in the ways the elements were combined. Experiments have shown that almost any source of energy, such as lightning, shock waves, ultraviolet

radiation or hot volcanic ash, would have led to significant conversions of early surface materials into a great variety of substances that would now be considered organic. The early solar system also included a very large amount of cometary and meteoritic material, which may have contributed substantially to the earth's surface. The effects of solar radiation on this ultracold material left over from the condensation of the solar system could have produced organic molecules as large as some biological polymers.

All conceptions of the "primordial soup" from which life arose agree in that it included not only the particular sugars, amino acids and other substances that are now essential biochemical reactants but also many other molecules that are now only laboratory curiosities. It was therefore necessary for the first organizing principle to be highly selective from the start. It had to tolerate an enormous overburden of small molecules that were biologically "wrong" but chemically possible. From this background the organizing principle had to extract those molecules that would eventually become the routinely synthesized standard monomers of all the biological polymers, and it had to link them dependably in particular configurations.

The total amount of potential organic material was immense. If the carbon now found in coal, carbonate rocks and living matter were uniformly distributed in all of the present ocean water, it would make a carbon solution as concentrated as a strong bouillon. Geophysical processes such as weathering, evaporation and sedimentation must have acted then as they do now to create a diversity of environments. Evidently at least one of these environments was suitable in temperature and composition for the origin of life.

The primitive soup did face an energy crisis: early life forms needed somehow to extract chemical energy from the molecules in the soup. For the story we have to tell here it is not important how they did so; some system of energy stor-

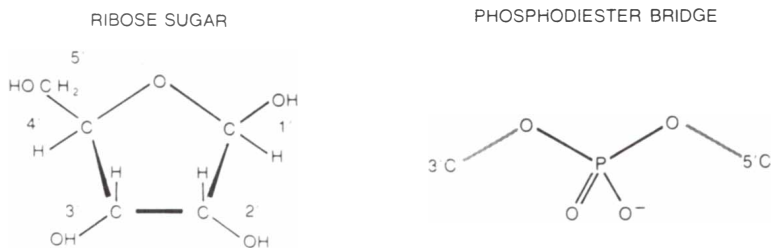
age and delivery based on phosphates can be assumed. Nonmetabolic replenishment of the phosphate energy reservoir (perhaps by some kind of conversion of solar energy to chemical energy) had to last until a mechanism evolved for fermenting some otherwise unneeded components of the soup. Fermentation would have been adequate until the advent of photosynthesis provided a continuing energy source.

The First Genes

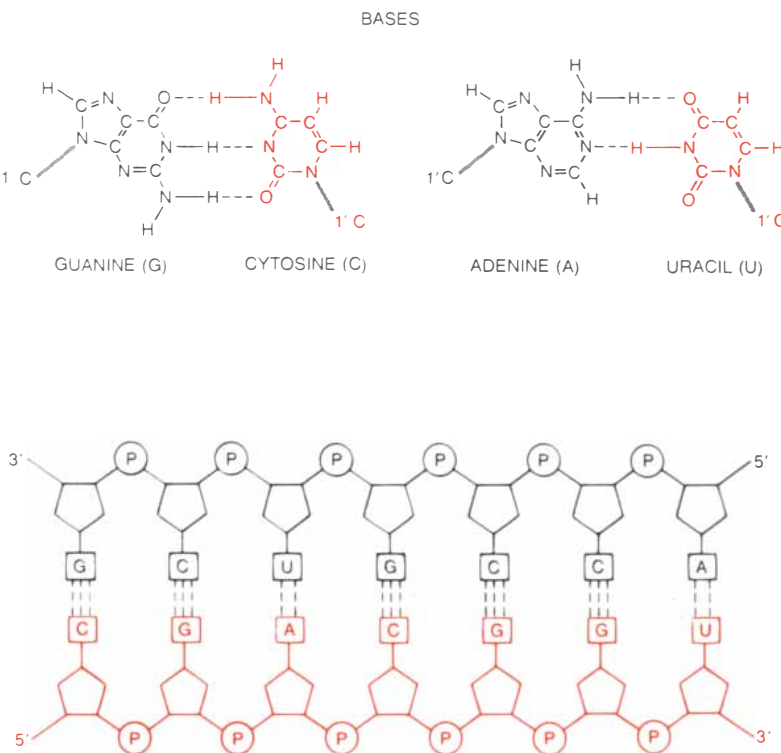
In cells genetic information is stored on DNA, transcribed onto messenger RNA and then translated into protein; in viruses the information is carried on either DNA or RNA strands. Both nucleic acids are threadlike molecules made up of nucleotides. Each nucleotide has three components: a chemical group called a base, a sugar (deoxyribose in DNA, ribose in RNA) and a phosphate group. The linked sugars and phosphates form the backbone of the molecule; the genetic information is encoded in particular sequences of bases. In DNA the four bases are the purines adenine (*A*) and guanine (*G*) and the pyrimidines thymine (*T*) and cytosine (*C*); in RNA uracil (*U*) takes the place of thymine. The bases are complementary, and they pair in accordance with specific rules: *A* pairs with *T* (or *U*) and *G* pairs with *C*. The complementarity is the basis of replication and transcription. In replication a strand of DNA or RNA serves as a template along which complementary nucleotides are assembled according to the base-pairing rules (by the various enzymes known as replicases and polymerases) to form a complementary strand carrying a duplicate copy of the information. In transcription a DNA sequence gives rise by a similar assembly process to a complementary strand of messenger RNA.

Knowing the chemical properties of DNA and RNA, what can one deduce about the identity of the first prebiotic information carriers? The deoxyribose nucleosides from which DNA is assembled are more difficult to deal with chemically than their ribose counter-

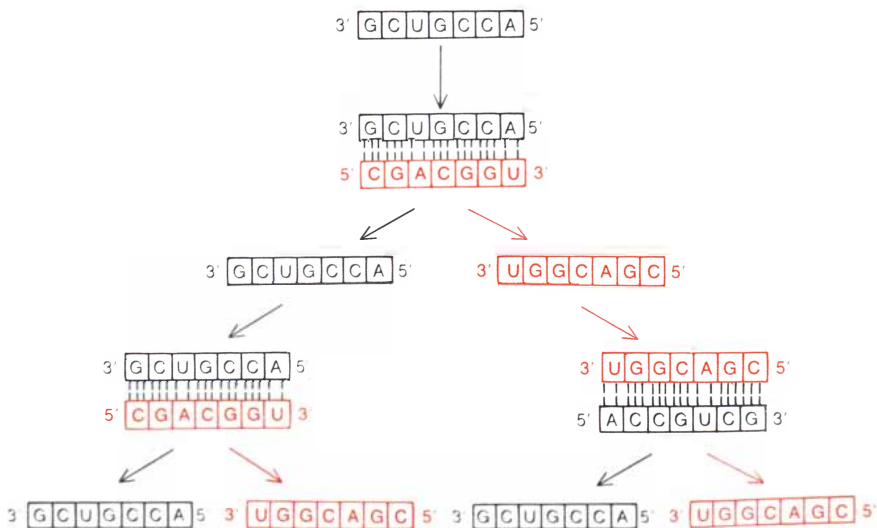
a



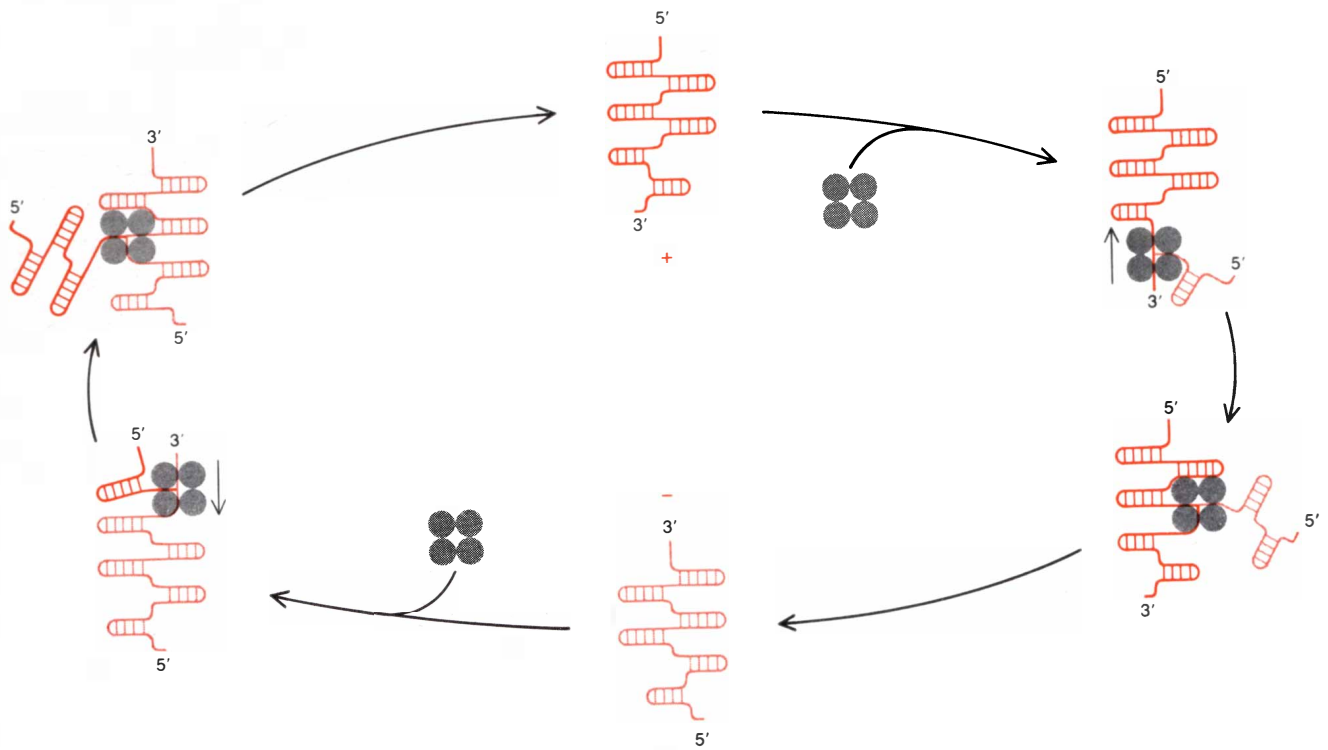
b



c

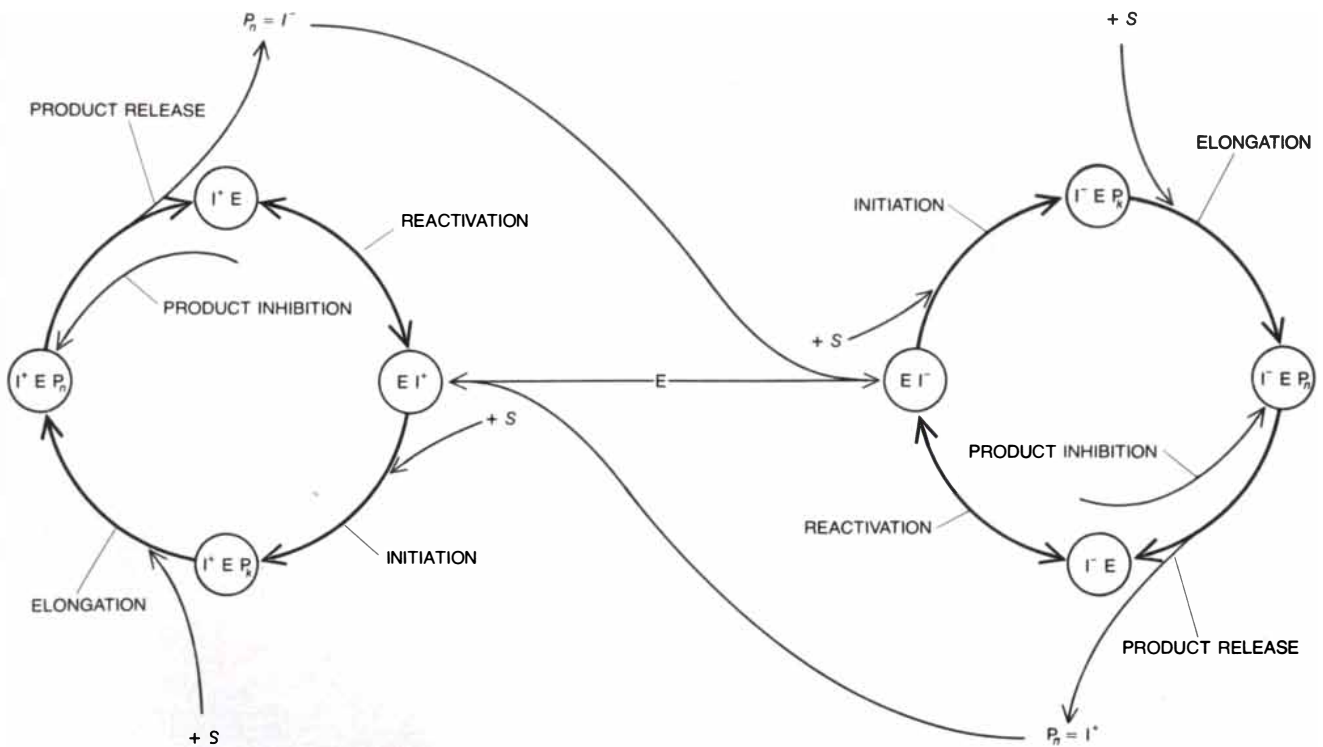


FIRST CARRIERS of genetic information were molecules of RNA: linear chains of nucleotides linked by phosphodiester bridges. Each nucleotide in RNA consists of a ribose sugar bonded to one of four bases (*a*); information is encoded in the particular sequence of the bases along the RNA strand. The bases are complementary: by means of hydrogen bonding (*broken lines*) adenine forms specific pairs with uracil and guanine pairs with cytosine. Information encoded in RNA is transferred by synthesis of a complementary replica strand (*color*), with the original strand as the template (*b*). A two-generation history of RNA information transfer is shown for an arbitrary sequence of seven nucleotides (*c*).



SINGLE-STRAND RNA of the bacterial virus Q_{β} is replicated by a specific replicase, an enzyme made up of four subunits. The "plus" strand of viral RNA is normally held in a specific folded configuration by base pairing within the strand (*top middle*). The strand unfolds as the replicating enzyme moves along it from the 3' end toward the 5' end (*top right*). The enzyme links complementary

monomers (nucleotide precursors in an energized form), following the instructions of the template according to the base-pairing rules to make a complementary "minus" strand (*bottom middle*). The new strand folds immediately, thus preventing the formation of a double strand, which would put a stop to replication. The minus strand is replicated in turn (*left*) to form a duplicate of the original plus strand.



INTERCONNECTED CYCLES of plus-strand and minus-strand synthesis are characteristic of RNA replication. The process has four phases: initiation of replication, elongation of the replica strand, release of the replica and reactivation. E is the replicating enzyme; I is the information-carrying molecule, or template RNA strand;

P_n is the product, or replica RNA strand, and S stands for substrate, the set of nucleotide monomers. The cross-catalytic nature of the process regulates the relative concentrations of plus and minus RNA strands in such a way that the growth rate of the ensemble of strands is the geometric mean of the rates for each kind of strand.

parts in RNA. Indeed, the synthesis of the monomers of DNA in cells proceeds via ribose intermediates, and DNA replication itself is initiated with RNA primers. In present organisms genetic information is processed by complex protein-RNA machinery. For such machinery to have evolved, the information carriers themselves must have had structural features that made them targets of recognition. Single-strand RNA can fold to form a great variety of three-dimensional structures, in contrast to the uniform double helix of DNA. In the present cellular machinery, wherever both functional and instructional properties are required, RNA is found. There is no reason to think it was otherwise during life's early stages. Nor is there any reason to think there was a process whereby information stored in any other form could have been transferred onto nucleic acids.

The search for the likely chemical identity of the first genes thus leads quickly to the base sequences of RNA. One can safely assume that primordial routes of synthesis and differentiation provided minute concentrations of short sequences of nucleotides that would be recognized as "correct" by the standards of today's biochemistry: the sequences had the same bases, the same covalent bonds and the same stereochemistry, or spatial arrangement of chemical groups. These sequences were present, however, with myriads of others that would be regarded today as "mistakes," with different stereochemistry, misplaced covalent bonds and nonstandard bases. What was so special about the sequences that resembled today's RNA?

There is a simple answer. Those RNA strands with a homogeneous stereochemistry and with the correct covalent bonding in the backbone of the strand could reproducibly lead to stable secondary structures, or foldings of the molecule, as a result of the formation of hydrogen bonds between pairs of complementary nucleotides. This was an important advantage, making the strands more resistant to hydrolysis, the cleavage by a water molecule that is the ultimate fate of polymers in water solution.

The primitive RNA strands that happened to have the right backbone and the right nucleotides had a second and crucial advantage. They alone were capable of stable self-replication. They were simultaneously both the source of instruction (through the base-pairing rules) and the target molecules to be synthesized according to that instruction. Here at the molecular level are the roots of the old puzzle about the chicken or the egg. Which came first, function or information? As we shall show, neither one could precede the other; they had to evolve together.

The chemical species and processes of prebiotic times surely had a variety of features in common with present-

The Vintage Years



FROM THE CHRISTIAN BROTHERS® OF CALIFORNIA

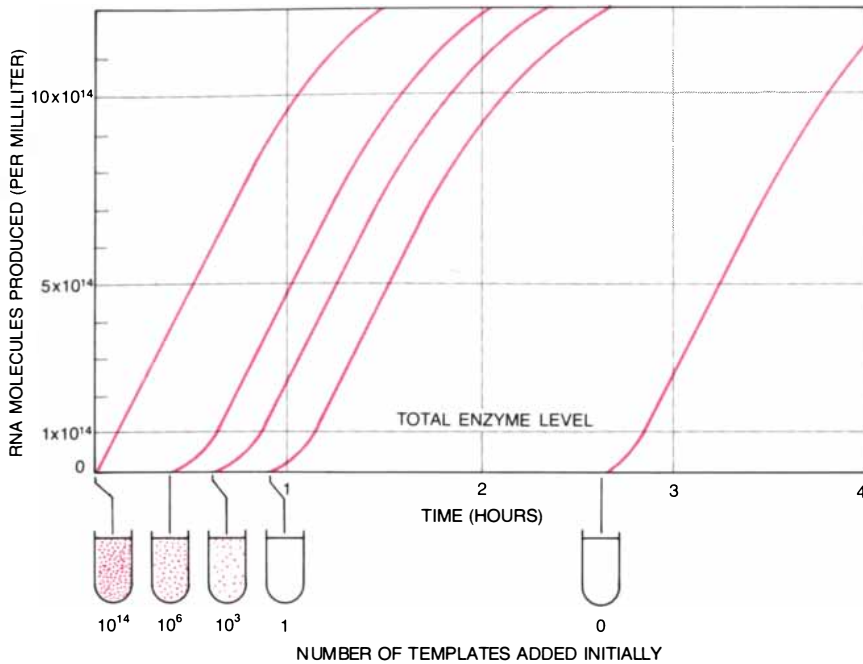
1975 Napa Valley Cabernet Sauvignon.
Our vintage-dated Cabernet

Sauvignon is made exclusively from Napa Valley grapes. Well-aged in oak and blended with Merlot and Cabernet Franc for added complexity, you'll find it to be a richly rewarding wine. Its intense varietal character makes it an excellent choice to put down in your own cellar... for it will continue to mature for many years. We invite you to enjoy our 1975 Cabernet Sauvignon.

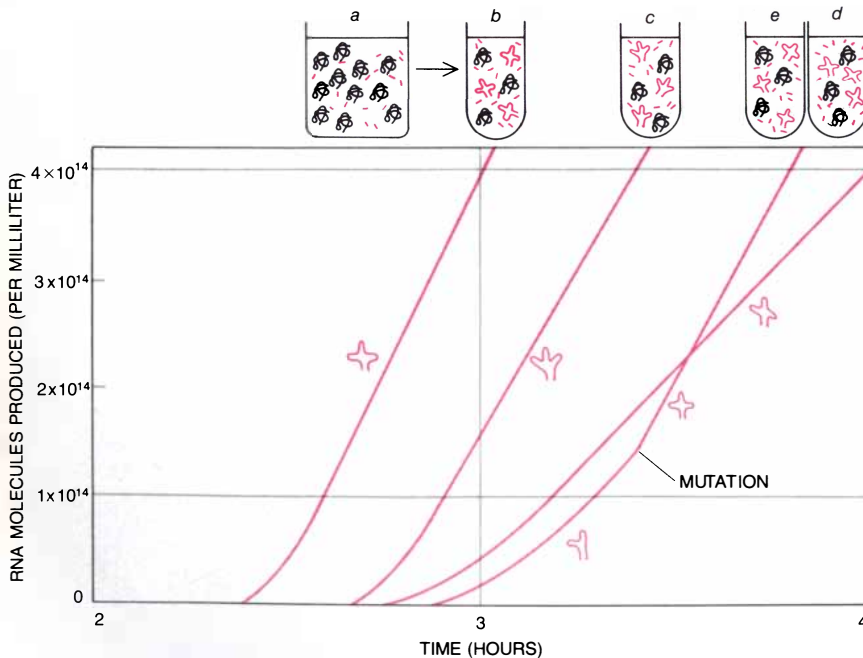
Brother Timothy F.S.C.
Cellarmaster



WORLDWIDE DISTRIBUTORS: FROMM AND SICHEL, INC., SAN FRANCISCO, CALIFORNIA, U.S.A.



RATE OF MULTIPLICATION of RNA strands is studied by incubating mixtures of Q_{β} RNA template strands, nucleotide monomers and Q_{β} replicase. When there is more template than enzyme, the growth of template molecules is linear (until it levels off at a high concentration because of product inhibition). Such growth is characteristic of successive replications where enzyme must be reactivated at each round. When there is more enzyme than template, growth is exponential; as soon as a cycle of replication is completed, the new product strand and the old template strand can each bind to a free enzyme molecule, giving rise to two more replicas. The growth curves remain parallel as they shift to the right with decreasing template concentration. Even if no template is originally present, an RNA related to Q_{β} fragments is synthesized de novo after a long time lag by interactions of the enzyme and the monomers.



DE NOVO SYNTHESIS is demonstrated by a cloning experiment. A template-free mixture of monomers and Q_{β} replicase enzymes (a) is incubated long enough to amplify any template that might be present as an impurity but not long enough for de novo synthesis. The mixture is then compartmentalized into four test tubes (b-e), in each of which de novo RNA strands multiply at various rates (curves). An optimal template arises by natural selection in each compartment and eventually becomes the sole product; since compartmentation preceded selection, however, the four compartments contain different products. Sometimes a mutation shows up late enough to be seen, as in compartment e. In contrast to the highly reproducible nature of template-instructed replication, large fluctuations are observed in the time required for the appearance of de novo products. The fluctuations reflect the fact that only a very small number of molecules is involved in the rate-determining step: synthesis of the first template.

day biochemistry. Sidney Fox and his colleagues at the University of Miami have shown, for example, that enzymatic functions can be exercised by "proteinoid" polymers made essentially by warming a mixture of amino acids (the constituents of proteins). In addition to such primitive catalysts there were undoubtedly molecules that were receptive to stimulation by sunlight; there were lipids (fats) or lipidlike molecules that could form membranous structures and there were perhaps even polysaccharides, or sugar polymers, that were potential sources of energy. In short, a wealth of functional molecules had been created by nonliving, or "nonorganic," chemical paths.

Such functional molecules may have been important in the chemistry of a prebiotic soup. They could not evolve, however. Their accidental efficiency rested on nonaccidental structural constraints, such as favorable interactions with neighboring molecules or particular spatial foldings. If their efficiency was to improve, and if more functional variants were to be favored over less functional ones, they would have to escape such structural constraints. Only self-replicative, information-conserving molecules could do so. We shall now discuss how the information content of such molecules can improve, how their complexity increases and how they drive out less functional variants.

Self-Replication

The virus Q_{β} , which infects the bacterium *Escherichia coli*, serves as a model system for studies of molecular self-replication. It has as its genome, or total hereditary material, a single-strand RNA molecule about 4,500 nucleotides long. Only part of the molecule constitutes the genetic message; the rest has a variety of functional (rather than informational) roles, such as serving as a target for specific recognition by enzymes. Some years ago Sol Spiegelman, who was then at the University of Illinois, isolated the Q_{β} replicating enzyme and showed that it is able to reproduce the virus RNA in a cell-free laboratory system to yield infectious copies. He also isolated from infected *E. coli* cells a noninfectious "satellite" RNA molecule 220 nucleotides long that is replicated by the Q_{β} replicase with exceedingly high efficiency. The satellite RNA and other similar "minivariants" serve in combination with Q_{β} replicase as convenient model systems for studies of RNA replication.

A typical experiment begins with a solution that includes magnesium ions, a low concentration of highly purified Q_{β} replicase and an energized form of the four RNA substrate monomers. The energized form consists of the nucleoside triphosphates *ATP*, *GTP*, *UTP* and *CTP*, in which the base and sugar are linked to

"Almost anything you can make out of anything, you can make with less energy from plastics!"

"It's no news that plastics are versatile," says Frank Barborek, Plastics Development Manager at Gulf Oil Chemicals Company. "Even more important perhaps is that plastics are remarkably energy-efficient.

6% — 

"Petrochemicals use only about 6% of America's crude oil supply, and it's an energy bargain."



"For instance, it takes less energy to make plastic bags instead of paper bags; plastic bottles instead of glass bottles; plastic cans instead of steel cans; plastic auto hoods and trunk lids instead of aluminum.

"Many plastics are lighter than the materials they replace, so they take less energy to ship and handle. And plastic parts can make cars dramatically more fuel-efficient: there's over 230 pounds of plastic in the average late-model car, and each pound of plastic can replace up to six pounds of steel. And of course plastics don't rust.

"Most plastics are based on chemicals from petroleum; and one reason why Gulf is committed to developing other fuel sources, such as tar sands and oil shale, is to maintain a steady supply of petrochemicals for plastics.

"Only about 6% of America's crude oil supply goes into petrochemicals, and the crude oil we turn into energy-efficient plastics helps make the rest of each barrel of oil go farther. That's more energy for tomorrow."

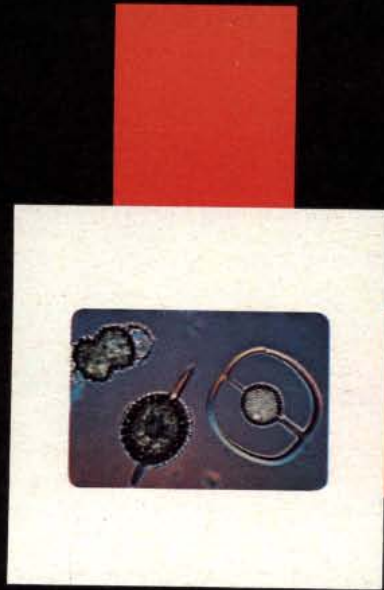


**Gulf people:
energy for tomorrow.**

Gulf Oil Corporation

"From houses to toothbrushes, plastics can do almost anything."

Now turn 35mm slides



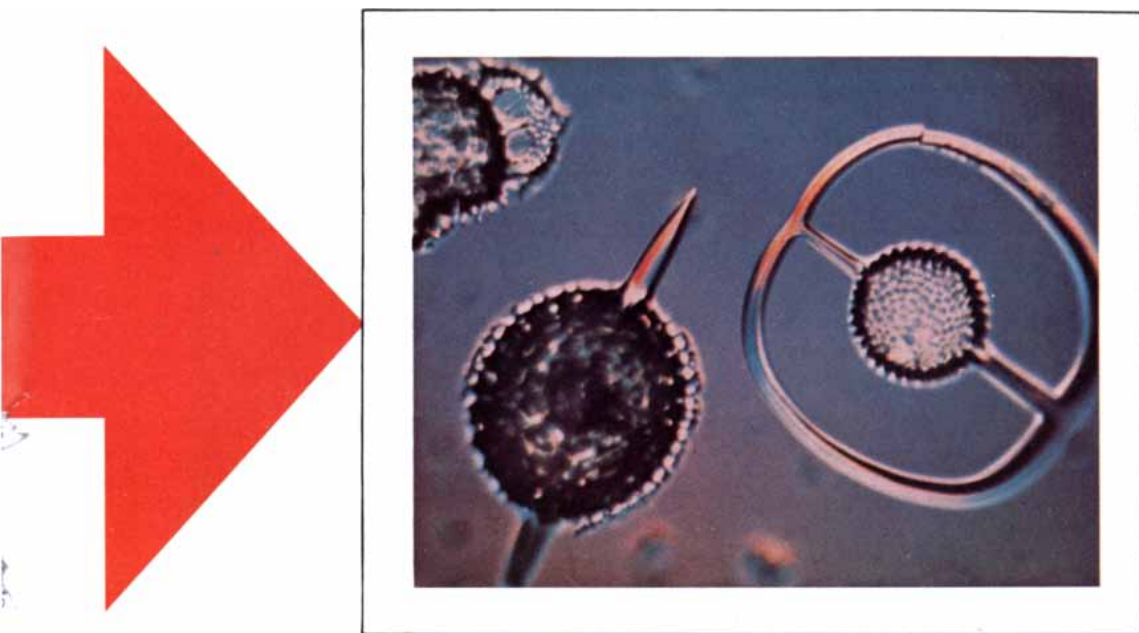
© 1981 Polaroid Corporation
"Polaroid," and "Polacolor" * "Polaprinter"™

Polaroid

Polaprinter Instant Slide Copier

© 1981 SCIENTIFIC AMERICAN, INC

into 3¼ x 4¼" prints. In seconds.



Now you can produce your own prints from your slides. On the spot. In seconds. With Polaroid's remarkable new Polaprinter slide copier, there's no more waiting for days to get your prints back.

How does it work? Frame your 35mm slide on the illuminated preview screen. Then insert it in the copier and push a button. Seconds later you have a beautiful, fully-developed color or black-and-white 3¼ x 4¼" print. Every 10 seconds you can start producing another picture from a slide. It's a no-fuss, push-button job.

The Polaprinter uses our Type 668 Polacolor film to give you sharp, vivid color prints in 60 seconds. With our Type 665 positive/negative film, you get a black-and-white print plus a high-resolution negative in just 30 seconds.

The Polaprinter itself is small and lightweight and takes up less than a square foot of space. It has fully automatic exposure control, a specially designed lens for optimum sharpness and an independent control for adjust-

ing contrast to your liking. A cropping selector is included to adjust for the slight disproportion between the slide and print formats.

You'll find that the Polaprinter is the quickest, easiest way to prepare prints from your horizontal or vertical slides to illustrate reports and articles. Or to keep records of patients or projects. Or to supplement correspondence and proposals. If you have a large collection of slides, you can make yourself a handy print file for quick reference. And, of course, you always have the security of doing your own copy work at your desk or in your lab.

Bring one of your 35mm slides to a Polaroid Polaprinter dealer for an on-the-spot demonstration. See for yourself how easily and quickly you can turn important data or pictures on your slides into superb prints. For the name of your nearest Polaprinter dealer, call us toll-free from the continental U.S.: 800-225-1618. In Mass., call collect: 617-547-5177.



a tail of three phosphate units. One of the four nucleosides (usually *GTP*) is labeled with a radioactive isotope so that the synthesis of new RNA can be followed. To initiate replication a certain amount of template RNA is added and the mixture is incubated.

In replication experiments done in 1974 by Manfred Sumper in our laboratory at the Max Planck Institute for Biophysical Chemistry in Göttingen something quite unexpected happened. When Sumper started incubation with more template than enzyme, he found a linear increase in RNA concentration, leveling off finally at a high concentration. This told us that all the enzyme molecules could simultaneously be occupied with templates being replicated; in spite of the steadily growing concentration of template, the concentration of active enzyme-template complexes was constant. Therefore new RNA continued to be synthesized at a constant rate.

A natural thing to do next was to reduce the amount of RNA template in the initial mixture. This resulted in shifting the linear growth curve parallel to itself to a later time. Successive reductions delayed the growth by an amount proportional to the logarithm of the template concentration. In other words, dilution from 10^6 to 10^4 template molecules per test tube shifted the growth curve by the same amount as dilution from 10^4 molecules to 10^2 . This logarithmic relation clearly indicated that when enzyme is in excess of template, each newly formed RNA molecule can immediately find a free enzyme molecule. The template concentration is therefore amplified exponentially rather

than linearly. This process works even for minute initial amounts of template; indeed it works with just one initial template molecule per test tube. (A method for cloning single molecules is based on this finding.)

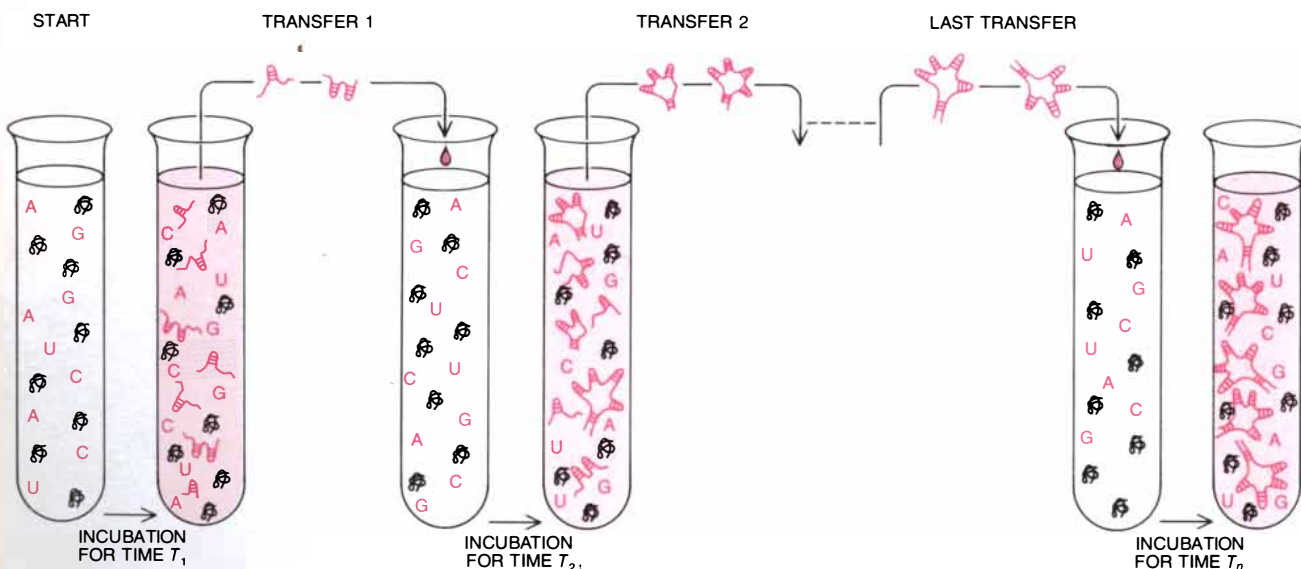
Imagine our surprise when Sumper reported one day that even if not a single RNA template molecule was added initially, RNA was still produced, albeit only after much longer and more variable incubation times. The possibility that one of the enzyme molecules might have carried some RNA impurity into the mixture was eliminated by several procedures. The substrate nucleosides were subjected to conditions under which any polymer would have been totally degraded. The enzymes were purified and analyzed with all possible care. Single impurity molecules of the template were added deliberately in order to demonstrate that an entirely different mode of growth was caused by impurities. Finally we were convinced we had before us RNA molecules that had been synthesized *de novo* by the Q_{β} replicase enzyme. What was most puzzling, the *de novo* product had a uniform composition, which in each trial turned out to be similar to or even identical with Spiegelman's minivariant.

Comparative studies of chemical reaction rates soon showed that the mechanisms of template-induced synthesis and of template-free synthesis are quite different. The complicated template-induced mechanism was resolved into elementary steps in terms of which the results of our rate studies could be quantitatively matched by algebraic expressions. In the template-induced reaction

one enzyme molecule is associated with one template molecule, and one substrate monomer is added at each step in the elongation of the replica; no cooperative interaction among substrate monomers can be observed. Template-free synthesis, on the other hand, does require the cooperation of at least three or four substrate monomers in the rate-limiting step of the reaction. Moreover, at least two enzyme molecules, each one loaded with substrate, participate in this step. One of the enzyme molecules apparently substitutes for the missing template by exposing bound substrate monomers to the polymerizing enzyme.

Spiegelman and Donald R. Mills of the Columbia University College of Physicians and Surgeons had determined the full sequence of the 220-nucleotide minivariant. On analyzing the sequence we saw that it could be represented as being composed (apart from 56 mutations and two insertions) of multiple copies of four tetramers and two trimers. The tetramers are *CCCC* and *UUCG* and their complements, *GGGG* and *CGAA*; the trimers, *CCC* and its complement *GGG*, represent truncated versions of tetramers. The sequence *CCCH* had been identified by Sumper and Bernd-Olaf Küppers as the recognition site that must be present in all RNA's that interact specifically with Q_{β} replicase; *UUCG* is a base sequence that, in a different context (the translation of messenger RNA into protein), binds to one of the proteins that act as subunits of the Q_{β} replicase complex.

Does the discovery of the *de novo* synthesis of RNA violate the central dogma of molecular biology, according



SERIAL TRANSFER, a method devised by Sol Spiegelman of the Columbia University College of Physicians and Surgeons to prolong growth indefinitely, was applied to demonstrate the *de novo* initiation of RNA synthesis and the evolution of optimal templates. A series of test tubes is prepared with Q_{β} replicase, necessary growth factors and monomers of A, U, G and C, but no RNA templates. The mixture in the first tube is incubated by raising its temperature; after

a long time a heterogeneous collection of short templates has been synthesized. A small fraction of this mixture is transferred to the second tube and incubated. The procedure is repeated many times. Eventually a single, optimal template is selected. As an indication of the vast multiplication achieved by this method, if the growth in each tube amplifies the product RNA 10,000 times, 10 transfers are equivalent to a growth that would saturate the world's oceans with RNA.



Obviously, a lot of people like our Skylark Limited because of its luxury.



24 EPA EST. MPG 36 EST. HWY.

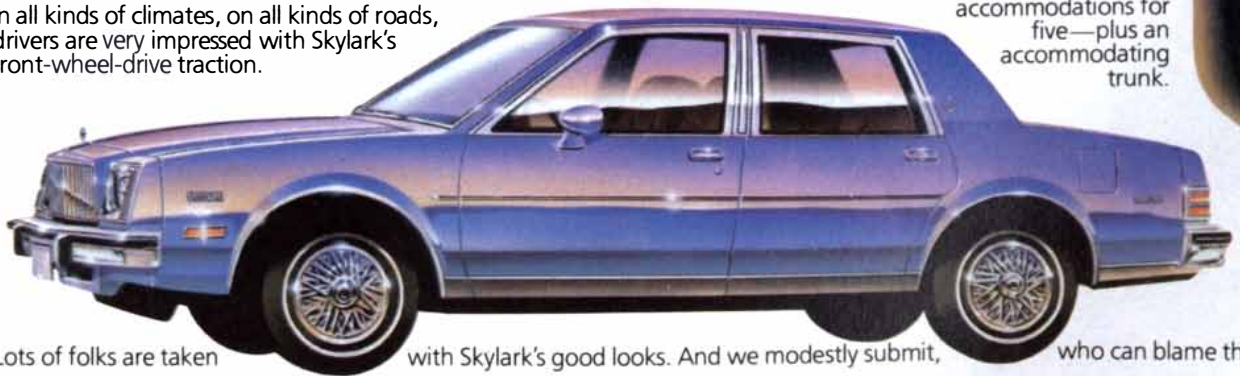
Skylark gets a lot of votes for its stand on economy, thanks to its 2.5 liter four (with available automatic transmission, without air conditioning)*.



In all kinds of climates, on all kinds of roads, drivers are very impressed with Skylark's front-wheel-drive traction.



Others like the fact that Skylark offers roomy accommodations for five—plus an accommodating trunk.



Lots of folks are taken with Skylark's good looks. And we modestly submit, who can blame them?

We couldn't get all the reasons people like Skylark on this page. But we sure got them into the car.



*Use "estimated mpg" for comparison. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower. Estimates lower in California. Buicks are equipped with GM-built engines supplied by various divisions. See your dealer for details.



COURVOISIER VSOP. THE BRANDY OF NAPOLEON

to which information can flow only from nucleic acids to proteins, and not the other way? The selection of these particular tetramers and trimers clearly represented "instruction" on the part of the proteins of the Q_{β} replicase complex. A wealth of possible sequences might have been composed of the tetramers and trimers, however, not just one, and a population of templates in the experiments comprises as many as 10^{12} molecules, only one of which needed finally to be amplified. Was natural selection at work rather than instruction by proteins?

The Role of Selection

Recently Christof Biebricher and Rüdiger Luce in our laboratory did a decisive experiment that answered the question. Their approach was determined by the special kinetics of de novo synthesis. They began by incubating a template-free mixture for a time that was long enough to amplify any templatelike impurity that may have been present but was too short for the formation of de novo RNA. The mixture was then divided into several isolated compartments where optimal conditions for de novo synthesis were maintained. The result was clear: Although each compartment had a uniform population of de novo products, the products differed from compartment to compartment. Further analysis revealed, however, that the different sequences were not completely unrelated.

Large fluctuations in the appearance times of different populations could be observed when compartments were incubated separately. The fluctuations reflect the probabilistic nature of a process in which the synthesis of a single molecule is the first, rate-limiting step. In contrast, template amplification always proceeds deterministically, with well-defined time constants, even if the reaction starts with only one template molecule or a few of them. Fluctuations in the rate of amplification average out during successive replications.

The early products appearing in the different compartments were clearly not yet optimized by any evolutionary process. Some were as short as 60 nucleotides, and still shorter ones had probably prevailed in the early stages of amplification. (An analyzable quantity of RNA consists of at least 10^{12} molecules. This number is approximately equal to 2^{40} , implying that 40 generations of amplification had passed before the products could be evaluated. In those 40 generations the most inefficient templates may well have been improved.)

Serial-transfer experiments, whereby growth can be prolonged through very many stages of amplification, quickly showed what the optimal products were like. They generally had lengths between 150 and 250 nucleotides. There

The Quasispecies Model

Prebiotic RNA chemistry provided an environment for Darwinian evolution: populations of self-replicating species (RNA strands with different sequences) competed for the available supply of "food" (energy-rich monomers). The continuous generation of mutant sequences, some of them having advantageous properties, forced evolutionary reevaluation of the fittest species. A quantitative theory of this molecular Darwinian competition has been developed.

Let the number of nucleotides in any sequence i be N_i and let each nucleotide position be identified by a subscript p that can have any value from 1 to N_i . Let the probability that the nucleotide at position p in sequence i gets copied correctly during self-replication be q_{ip} , then $(1 - q_{ip})$ is the error rate for that position. The symbol q_{ip} therefore describes the quality, or copying fidelity, of replication at position p of sequence i . The probability Q_i that an entirely correct sequence i will result from a replication is the product of all the single-nucleotide copying fidelities:

$$Q_i = q_{i1} \times q_{i2} \times \dots \times q_{iN_i} = \bar{q}_i^{N_i},$$

where \bar{q}_i is the geometric mean of the copying fidelities in sequence i .

Sequence i can survive successive replications only if copying errors do not accumulate. This requires the sequence to be superior in net growth to the average of its competitors by a competitive-advantage factor S_i . Furthermore, i can be selected only if a survival condition called the error threshold is satisfied; the threshold is $Q_i S_i$, and for sequence i to survive it must be greater than 1.

Net growth is governed by the equation that tells how x_i , the fraction of all the sequences present that are exact copies of sequence i , changes with time. Major causes of change in x_i are error-free replication of i and erroneous replication of closely related sequences, collectively designated j , which can give rise to i by mutation. Taking both contributions into account gives the rate of change of x_i :

$$(W_{ii} - \bar{E})x_i + \text{sum of } W_{ij}x_j.$$

In this expression W_{ii} is the rate of correct replication of sequence i and \bar{E} is the average rate of excess production (excess of replicative gain over all losses) of all sequences present; both rates are expressed on a per-copy basis. W_{ij} is the rate of production of sequence i by erroneous copying of sequence j , given per copy of sequence j . In the equation the contributions from all the sequences designated j are summed. Hence the first

term is the rate at which sequence i competes with other sequences and the second term is the rate at which i is produced by mutations of other sequences.

These rates determine how any arbitrary starting set of sequences proceeds to organize itself. The first term can be positive or negative depending on whether W_{ii} is greater than or less than the average excess production \bar{E} . If W_{ii} is greater, x_i grows; if it is less, x_i steadily decreases until sequence i dies out or is produced only by mutation. The decline of sequences for which W_{ii} is less than \bar{E} has the effect, however, of increasing the average excess production rate \bar{E} . This makes it ever more difficult for surviving sequences to satisfy the requirement that W_{ii} be greater than \bar{E} , and hence to grow. The self-replicative competition resembles a high-jump competition in which the bar is raised until only one competitor survives. In the molecular competition, however, a single winner never emerges. Because of the mutation terms $W_{ij}x_j$ the strongest competitor constantly produces mutant sequences with which it must continue to compete. In the steady state that is eventually reached the best competitor, designated the master sequence m , coexists with all mutant sequences derived from it by erroneous copying. We designate this distribution of sequences a quasispecies.

This analysis shows that Darwin's principle of natural selection is not of an axiomatic nature but derives from the set of physical conditions that pertain to self-replication. The end result of selection, the quasispecies, is stable until mutation happens to produce a new sequence with a growth rate higher than that of the existing master sequence (or until a change in environmental conditions has an equivalent effect). When that happens, the new "fittest" sequence proliferates until it (together with its mutants) takes over and the old quasispecies disappears.

Quantitative results describing RNA quasispecies have been derived. For example, the maximum length of a master sequence (the number of nucleotides) is

$$\frac{2.3 \log S_m}{1 - \bar{q}_m}$$

Longer master sequences could not surmount the error threshold, that is, $Q_m S_m$ could not be greater than 1.

This discussion summarizes the main results of mathematical investigations carried out by our group and later by B. L. Jones, R. H. Enns and S. S. Ragnekar of Simon Fraser University in British Columbia and by C. J. Thompson and J. L. McBride of the University of Melbourne.

was a definite, uniform final product for any set of experimental conditions, but there were as many different optimal products as there were different experimental conditions. One of the optimal products appeared to be Spiegelman's minivariant (which under the conditions of Sumper's experiments had shown up reproducibly). Other products of optimization were adapted to conditions that would destroy most RNA's, such as high concentrations of ribonuclease, an enzyme that cleaves RNA into pieces. Apparently the variant that is resistant to this degradation folds in a way that protects the sites at which cleavage would take place. Some variants were so well adapted to odd environments that they had a replication efficiency as much as 1,000 times that of variants adapted to a normal environment.

These observations leave no doubt that Sumper's results demonstrated de

novo synthesis. The uniformity of the de novo products is seen to be a consequence of natural selection and not of faithful sequential instruction by the enzyme. And so the central dogma is saved, at least in its essence.

What is more important here is what the experiments reveal about Darwinian processes. Natural selection and evolution, which are consequences of self-reproduction, operate in the case of molecules as they do in the case of cells or species. What is truly surprising, and an important discovery indeed, is the high efficiency of the process of adaptation in such a simple self-reproduction system.

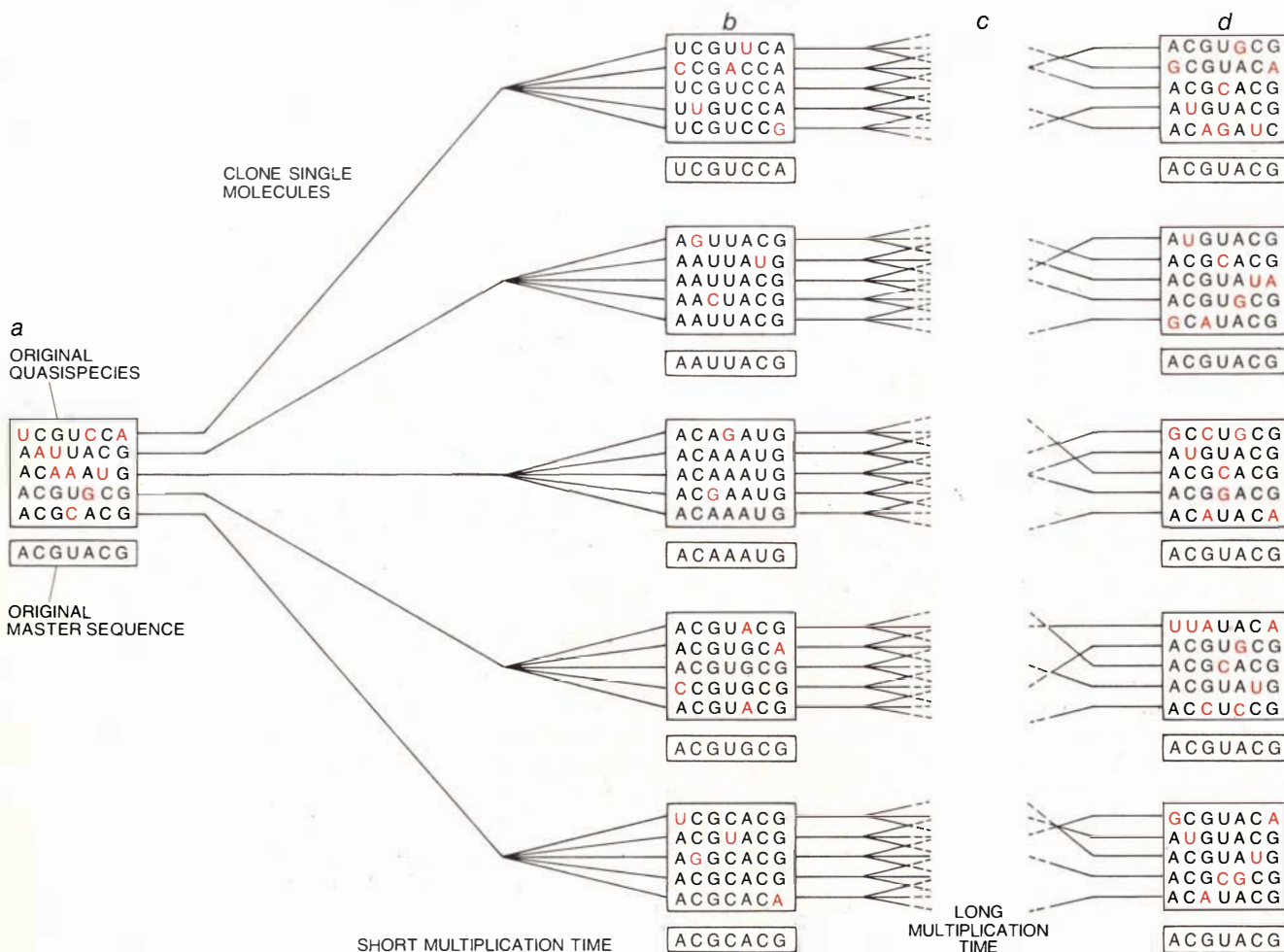
It may be objected that an enzyme such as Q_{β} replicase, a complex biological molecule, should not be present in an experiment designed to represent the prebiotic situation closely, even though the replicase was not a target of reproduction or evolution, but simply a factor

in the environment. Fair enough! And that raises another important question.

Template without Enzyme

If the functioning of RNA as a template always required the participation of something as complex as Q_{β} replicase, prebiotic evolution would require optimization procedures beyond those based on the self-reproduction of RNA. It is therefore important to establish what kinds of self-reproduction and selection can take place in environments simpler than those that include well-adapted replicases. Then we can consider how instructed synthesis of proteins can originate on a Darwinian basis.

This question must be answered by experiment. Important clues can be cited from the recent work of Leslie E. Orgel and his colleagues at the Salk Institute for Biological Studies. Short



UNIQUE MASTER SEQUENCE of a quasispecies is preserved in spite of the continuous redistribution of mutant sequences, as was demonstrated by Charles Weissmann of the University of Zurich. He cloned Q_{β} RNA in bacteria by diluting an original quasispecies of the virus (a) so strongly that every infection was initiated by a single virus particle. Next he analyzed the RNA sequence of each clone (b) by "fingerprinting" (two-dimensional electrophoresis of partially fragmented RNA). Each clone was then allowed to multiply for many generations (c) and was thus subjected to prolonged selection pressure. The result was the establishment of new quasispecies

distributions (d). The principle of the experiment is illustrated here for five original sequences, each of them seven nucleotides long. The master sequence is the one in which each position is occupied by the nucleotide most frequently found at that position. Differences from the master sequence are shown in color. The master sequence is always clearly defined even though it is actually seen very infrequently among the mutated sequences. After cloning, the new master sequences (b) are different from one another, and none of them matches the original master sequence. After very many generations, however, all the master sequences (d) have reverted to the original one.

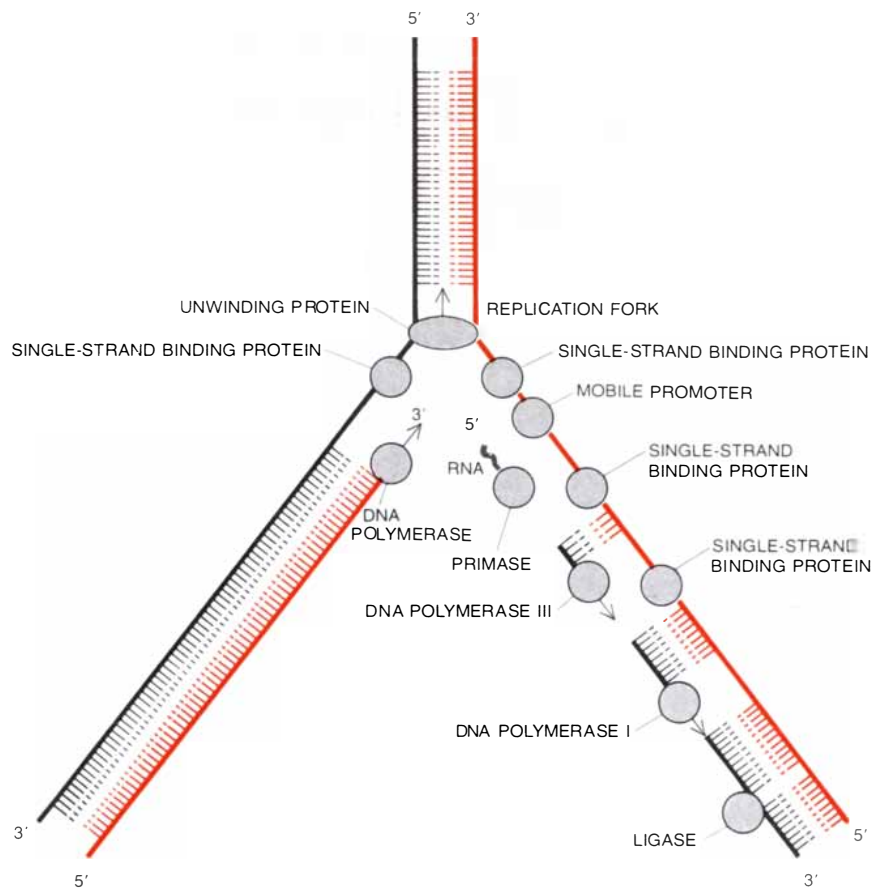
polymers of the adenine nucleotide (oligo-*A*) form spontaneously when *A* monomers are exposed to templates that are long polymers of the complementary nucleotide *U* (poly-*U*), even when no enzyme or other catalyst is present. The length of the oligo-*A* chains averages five nucleotides and can be as much as 10. If lead ions are present as a catalyst, the yield is dramatically increased; in addition the successive monomers are for the most part (75 percent) linked to each other as they are in RNA: by a phosphate group that forms a bridge from the 3' carbon atom of one sugar to the 5' carbon of the next. If poly-*C* is the template in a 50-50 mixture of activated *A* and *G* monomers with lead ions present, the products have an overall 10:1 ratio of *G* to *A*; in other words, more than 90 percent of the base pairing is correct. When zinc ions are present, poly-*C* template and activated *G* monomers yield oligo-*G* chains of up to about 40 bases, and the fidelity is 20 times better than it is with the lead-ion catalyst. Has nature perhaps "remembered" how replication started? Today's RNA polymerases all include zinc ions.

Orgel's data show that polymers rich in *G* and *C* offered special advantages in early evolution. They alone had sufficiently high copying fidelity in the absence of well-adapted replicases; they alone provide enough "stickiness" in their base pairing to enable sizable messenger RNA's to be translated into functional proteins in the absence of ribosomes, the sites of translation in present-day cells. Kinetic and thermodynamic studies by Dietmar Pörschke in our laboratory have put these conclusions on a quantitative basis. *G-C* pairing proves to be about 10 times as strong as *A-U* pairing, so that complementary strands remain in contact with each other much longer when they are rich in *G* and *C*. Furthermore, the binding is strengthened cooperatively by neighboring pairs. From these data we have derived pairing rules for an evolutionary model that makes it possible to identify well-known RNA structures (such as the characteristic cloverleaf of transfer RNA's) as the evolutionary outcome of trial-and-error processes.

The essential conclusion of these enzyme-free studies is that RNA self-replication does indeed take place without the assistance of sophisticated enzymes. One can proceed to consider the evolutionary consequences of RNA self-replication without worrying about whether it really went on in prebiotic times. It did go on.

RNA Quasispecies

Suppose there were an endless supply of activated RNA monomers and the lifetimes of RNA's were infinite. What kind of self-replication would take place? Any RNA formed by nonin-



REPLICATION OF DOUBLE-STRAND DNA is much more "sophisticated" than that of RNA and includes mechanisms for detecting and correcting errors. Twenty or more enzymes are involved. At the replication fork an unwinding protein separates the two parental strands; single-strand-binding protein keeps them apart. Because replication proceeds along both template strands in the 3'-to-5' direction the process is discontinuous for one of the strands (right). A mobile promoter provides a recognition site for a primase that lays down a short RNA primer (which is later replaced with DNA). Polymerase III adds DNA monomers to elongate the strand; polymerase I "proofreads" the sequence, excises incorrect nucleotides and inserts the correct ones. Finally the enzyme ligase fills in the gaps between the daughter-strand fragments. In the absence of proofreading, DNA replication is no more accurate than RNA replication.

structed chemistry would be reproduced by template-instructed chemistry at a rate proportional to the current RNA concentration. The result would be exponential growth. Furthermore, even if only a single template were formed initially by noninstructed synthesis, there would soon be a host of different sequences because errors (point mutations, insertions and deletions) would inevitably be made in the course of replication. Hence in each generation there would be not only a larger number of RNA strands but also a greater variety of RNA sequences. What would happen then? Some of the mutants would be copied more rapidly than others or would be less susceptible to errors in copying, and their concentration would increase more rapidly. Sooner or later these faster-growing mutants would take over.

The same general picture emerges if monomers are supplied only slowly, so that the growing polymers have to compete for them, and if account is taken of the fact that a strand of RNA has a finite

lifetime. Self-replication is a competitive process; the best competitor is the mutant sequence with the most favorable combination of copying fidelity, stability and replication rate. This is the basic line of thought along which one has to proceed in order to understand the self-replication experiments we described above, and on which our theory of self-replication is based.

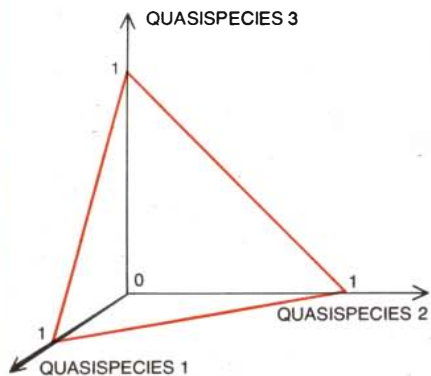
The theory of competition in molecular self-replication is summarized in the box on page 99. The result of the competition is the "survival" of the RNA sequence that is best adapted to the prevailing conditions, which we call the master sequence, together with a "comet tail" of similar sequences derived from the master sequence by mutation. Although the values of the rate coefficients that pertain to primordial chemistry are not known, quantitative conclusions can nonetheless be drawn. One conclusion is that there is a threshold condition for the stable self-replication of a genetic message. Until circumstances allowed this threshold to be

The Hypercycle Model

When RNA-instructed synthesis of proteins became a factor in evolution, there arose a new kind of dynamical interaction among molecules. The characteristic features of the interaction can be understood by means of a "topological" analysis, in which only qualitative and not quantitative conclusions are reached.

Consider a set of several different master sequences with their mutant distributions, each master sequence and distribution being (in the absence of the other ones) a stable quasispecies. The total information content of all the master sequences exceeds the limit established by the error threshold for one master sequence alone. For such a set to be stable and retain this total amount of information three conditions must be met: (1) Each quasispecies itself must remain stable, that is, each master sequence must compete successfully with its mutants, so that errors do not accumulate; (2) the different master sequences, with different selective values, must tolerate one another because of mutual catalytic couplings; (3) the coupled set must remain stable, regulating the populations of its members and competing as a set with alternative sets.

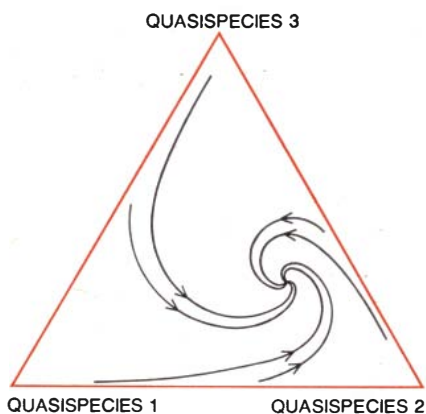
The topological analysis begins with the definition of a composition space in which each coordinate axis represents the fractional population of a quasispecies (the number of RNA strands belonging to the quasispecies divided by the total number of RNA strands present). For three quasispecies a three-dimensional composition space is defined. A state of the system is characterized by three fractional-population values and is represented by a point in the three-dimensional composition space. Because the three fractional populations are all positive and their sum is 1, this point must lie somewhere on the equilateral triangle whose vertices are at 1 on each axis of the composition-space coordinate system.



The equilateral triangle is called the unit simplex. If more than three quasispecies are considered, the unit simplex is a geometric figure of higher dimension. The triangle's corners represent states of the system comprising only one quasispecies, the edges of the triangle represent states with two quasispecies and the interior of the triangle represents states in which all three quasispecies are present.

A temporal sequence of states is described by a "trajectory," a curve on the unit simplex. Methods are available for discovering the qualitative nature of trajectories without solving the dynamical equations, which cannot be solved analytically if more than two quasispecies are present. There are various possible classes of trajectories. In a stationary solution the trajectories converge to a stable point, and thereafter the fractional populations remain constant. Another class of trajectories reflects oscillations of the quasispecies populations and a third class reflects a kind of irregular behavior known as chaos.

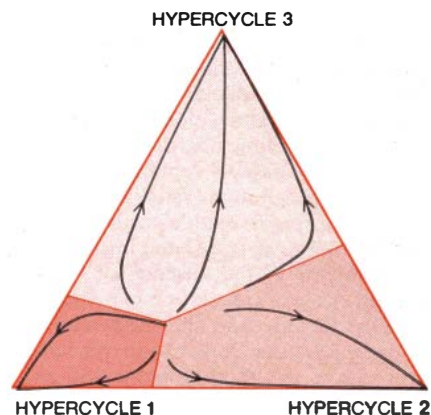
What forms of mutual coupling cause the trajectories to stay inside the unit simplex, and hence imply the coexistence of all quasispecies?



A trajectory leading to an edge or a corner of the unit simplex implies the disappearance of one quasispecies or more. Topological analysis of coupled systems in general has revealed that coexistence, and hence fulfillment of the three conditions stated above, requires a particular kind of coupling we have termed hypercyclic. In a hypercyclic system a closed loop of catalytic couplings connects self-replicative cycles.

Hypercycles as a class have distinctive dynamical characteristics. The growth rate of a hypercycle is proportional not to the current populations of the quasispecies, which would lead to

exponential growth, but to the population raised to a power greater than one. This autocatalytic growth of higher than first order can be called hyperbolic growth. Hypercycles also differ from Darwinian self-replicative systems in that they give rise to "once-and-forever selection." A competition among hypercycles can be analyzed topologically in the same way as a competition among quasispecies. The composition space gives the fractional populations of the competing hypercycles, and the trajectories lie on a hypercycle unit simplex. It turns out that all trajectories lead to the vertex of the area of the unit simplex in which the competition began. In other words, any competition among hypercycles leads to the survival of only one hypercycle.



Once-and-forever selection implies that a hypercycle, once established, cannot be displaced by a competitor appearing in small numbers even if the competitor is more efficient. This reflects the fact that the selective value of a hypercycle depends on its population. That is not the case for Darwinian systems, where a single advantageous mutant molecule can outgrow an established population. Hypercycles can evolve, however, by optimizing their internal linkages as a result of replacements, insertions and deletions in the information-carrying molecules. Because hypercycles provide complete mutual control of the coupled populations they grow coherently.

Hypercycles appeared in evolution with the origin of translation, which introduced a new requirement: it became necessary to feed back an evaluation of the quality of translation products to the genes that encoded those products. Presumably hypercycles still exist today as features of viral infection processes.

crossed no genetic message of any kind could survive.

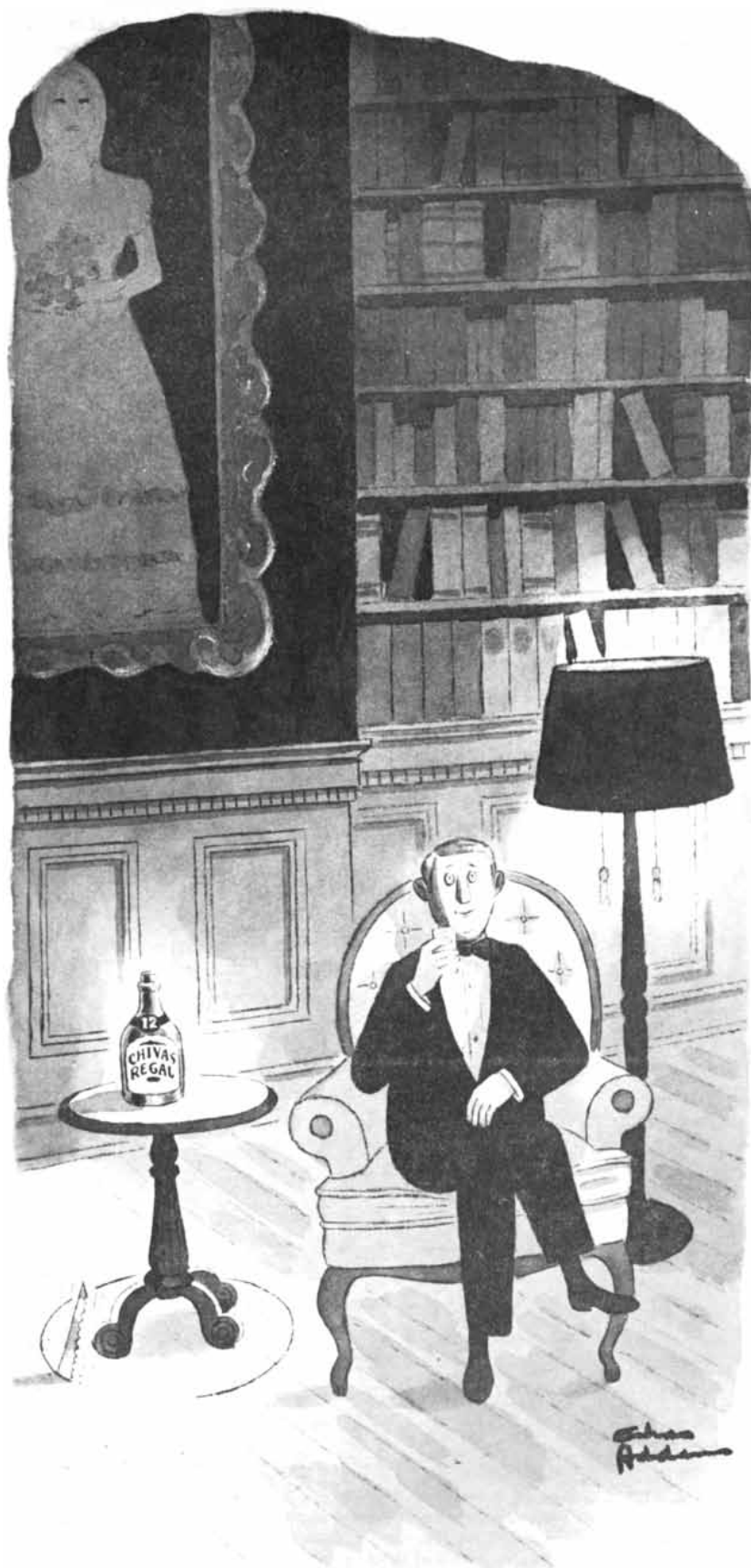
One can estimate the maximum gene lengths available to prebiotic systems by inserting plausible values in the equation that determines the length of a gene in our model of self-replication. The maximum gene lengths range from 50 to 100 nucleotides, which is similar to the size of today's transfer RNA's. This is a reassuring result in that it is long enough for internal folding and stability, but at first it appears to be much too short a genetic message to encode a functional protein.

Before showing how these theoretical results are confirmed in the laboratory, let us consider just what it is that is selected in the RNA self-replication competition. In one sense it is the fittest of all the genes present (that is, the master sequence) because this sequence is present in the highest concentration. The master sequence is likely to constitute only a small fraction of the total gene composition, however. The number of mutants would be extremely large under prebiotic conditions because the chemical kinetics of most mutant sequences could not have been much different from the kinetics of the master sequence itself. Hence the result of the self-replication competition had to be the master sequence together with a huge swarm of mutants derived from it and from which it had no way of escape.

We call this entire mutant distribution a quasispecies. It is the quasispecies mutant distribution that survives the competition among self-replicating RNA's, and not just one master sequence or several equivalent ones that are the fittest genes in the distribution. The essence of selection, then, is the stability of the quasispecies. Violating the error-threshold relation is equivalent to destabilizing the quasispecies. The master sequence is then unable to withstand the accumulation of errors; the distribution starts to drift, and all information is lost.

The theoretical equations describing self-replicative competition have been tested and confirmed with data from cloning experiments on Q_{β} self-replication done by Charles Weissmann and his co-workers at the University of Zurich. They measured short-time and long-time replication rates and studied the competition among mutant clones and wild-type Q_{β} RNA. Quantitative analysis of the data yielded values for copying fidelity and for competitive advantage in agreement with theory. The experiments showed that even with a highly evolved replication apparatus, organisms contend with less-than-perfect fidelity of replication by limiting gene lengths and by surviving with quasispecies distributions rather than with unique genes.

We mentioned above the energy cri-



Chivas Regal • 12 Years Old Worldwide • Blended Scotch Whisky • 86 Proof •
General Wine & Spirits Co., N.Y.

sis that had to be overcome in the first stages of biogenesis. We now discuss an obstacle that played an even larger role in the evolution of life: an information crisis.

Error, Genotype and Phenotype

The first Darwinian molecular systems owed their self-replicating ability to inherent physical forces that brought about the formation of complementary base pairs. The error threshold set a length limit of about 100 nucleotides, attainable only by RNA sequences rich in *G* and *C* nucleotides. This limit was eventually overcome by the development of a capability for the translation of genes into protein, and thus of enzyme machinery that reduced the error rate enough to make possible gene lengths of up to several thousand nucleotides. This new barrier is still reflected in the limited gene lengths of present single-strand RNA viruses, even though the viruses are a much later evolutionary development.

Further extension of gene length was possible only with the appearance of mechanisms for detecting and correcting errors. The distinction of right from

wrong could then be made if the newly formed daughter strand remained associated with its parental template, in which case "wrong" could be identified chemically as "unpaired."

All of this became possible when double-strand DNA appeared on the scene. DNA polymerases are equipped with proofreading and error-suppressing functions so effective that they allow gene lengths to be extended to millions of nucleotides. Lawrence A. Loeb of the University of Washington School of Medicine has shown that if a DNA polymerase is unable to carry out error-correcting functions, it has the same replication fidelity as an RNA replicase, namely a value that lies between .999 and .9999.

The invention of DNA made possible the formation of individualized cells in which the division of the cell is synchronized with the replication of its genetic material. Now, however, a new information crisis had to be overcome: high-fidelity replication narrowed the opportunities to provide variability by means of point mutations. The new barrier was overcome by the development of recombinative processes, including eventually sexual reproduction, which graft-

ed Mendelian genetics onto self-reproduction, the basis of Darwinian systems.

The first information crisis could be overcome only by the organization of a self-optimizing enzymatic replication machinery based on a stable quasispecies. This evolutionary jump required translation of RNA information into a new language: a functional one, namely proteins.

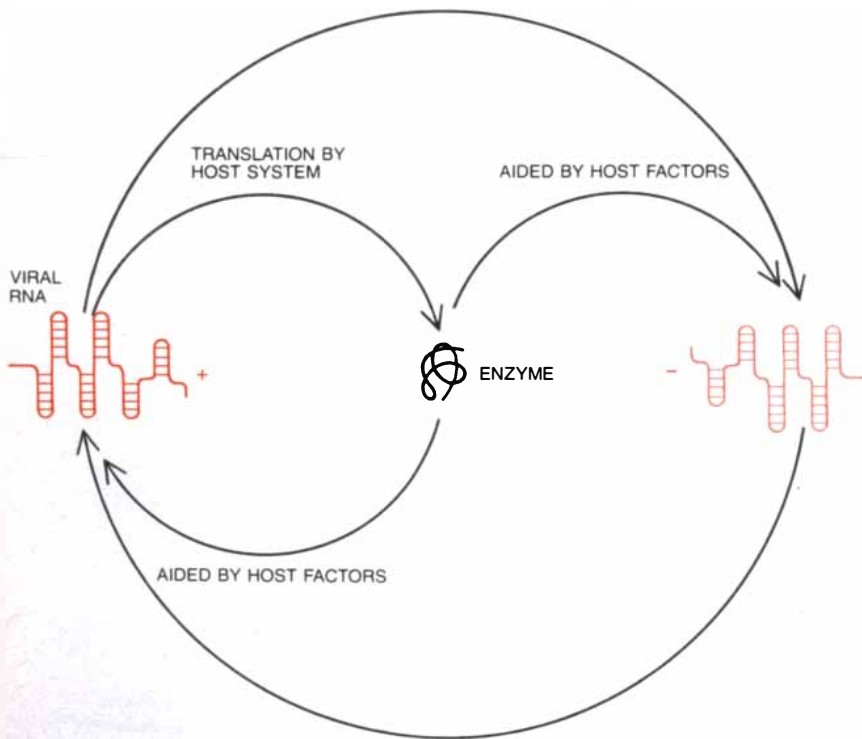
To encode even the most primitive translation apparatus certainly required far more than the 100 or so nucleotides that could be reproducibly stored in one master sequence. However the first protein machinery arose, it required more information than a primitive molecular Darwinian system was able to provide. More information could be stabilized only with the cooperation of differentiated genes, mediated and regulated by their very translation products.

If in the expanded information system the products of translation are also targets of evolution, a new problem arises. Selection has to act on the information content of the nucleotide sequence: on its genotype. Evaluation for selection, however, has to take place at the level of the gene product's function: the phenotype. This genotype-phenotype dichotomy demands that the system feed back information to its own genes, a procedure called second-order autocatalysis. It is second-order because in order to reproduce itself the information carrier needs both the information supplied by the template and the machinery encoded in the template. We have named double-feedback loops of this kind hypercycles. The term includes a large class of higher-order autocatalytic mechanisms. They exhibit a particular kind of temporal behavior different from that of other Darwinian systems.

Hypercycles: Quasispecies Cooperate

Hypercyclic coupling operates today when an RNA virus attacks a cell. If viral RNA were just another template in the replication environment of the host cell, it would not be able to outgrow other host templates. What it does instead is specify information for a replication machine that is highly selective for the viral RNA itself. Most parts of the machinery are provided by the host, but the specific hypercyclic linkage ensures the success of the virus's strategy of attack.

A simple example will explain the basic operating principle of a hypercycle. Suppose RNA sequence 1 codes for an enzyme 1 that helps to catalyze the self-replication of RNA sequence 2. Sequence 2 in turn codes for an enzyme 2 that helps to catalyze the self-replication of RNA sequence 1. What happens? Sequence 1 needs enzyme 2 for its self-replication and sequence 2 needs enzyme



HYPERCYCLES in modern organisms are observed at the simplest level when an RNA virus infects a cell. The virus ensures its own preferential replication by supplying the genetic information for an enzyme that will catalyze the multiplication of that information. The information is supplied as a single plus strand of RNA. It is translated by the host cell's machinery into an enzyme that, with the aid of host factors, replicates the RNA, producing a minus strand; the minus strand is replicated in turn to produce a new plus strand. The double feedback loop, in which both the enzyme encoded by the template and the sequence information on the template contribute to replication of the template, is equivalent to second-order autocatalysis.



This photograph was taken in Saudi Arabia

Bet you didn't think Saudi Arabia looked like this.

Not all of it does, of course.

Some places in Saudi Arabia look like orange groves in Florida. Others look like mountains in Southern California. And *part* of Saudi Arabia looks like what most people think *all* of it looks like.

To us, Saudi Arabia looks like home. We're Aramco, the Arabian American Oil Company. There are 13,000 North Americans there with us. And sailboats aren't the only

things that might surprise you about our lives there.

1. We're doing something important. Aramco produces more oil than any other company. Badly needed oil. Including about 15 percent of the oil the U.S. imports.

2. Aramco is working on some *incredibly* large energy projects. And on huge communications networks, electric utilities, and more.

3. Our people are glad to be in Saudi Arabia with Aramco. They came for excellent pay and professional challenge.

4. After 46 years in Saudi Arabia, Aramco is still growing fast. So is the number of interesting and rewarding jobs we offer.

5. There was a great sailing breeze until just before we took this picture, honest.

ARAMCO
SERVICES COMPANY

1100 Milam Building FS • SA
Houston, Texas 77002
(713) 750-6965

1. Therefore neither sequence can afford to outcompete the other for the available supply of RNA monomers; the two sequences are forced to cooperate. Depending on the rates of the numerous catalytic steps, wide ranges of concentrations may prevail, but as long as there is interdependence only an unusual "fluctuation catastrophe" or a major change in chemical conditions can extinguish an existing hypercycle. Templates and enzymes function together in self-replication, with the protein product of the RNA serving as a replicase, as an activator of a replicase or as some other control device enhancing the speed and accuracy of self-replication.

The kinetic behavior of such coupling into hypercycles has been investigated in detail and it has been shown that hypercycles are the only functional networks that can exceed the error threshold for stable quasispecies. Hypercyclic growth proves to be explosive when compared with first-order autocatalytic growth in a system having comparable rate coefficients. The consequences of

hypercyclic growth in selection are even more strikingly different.

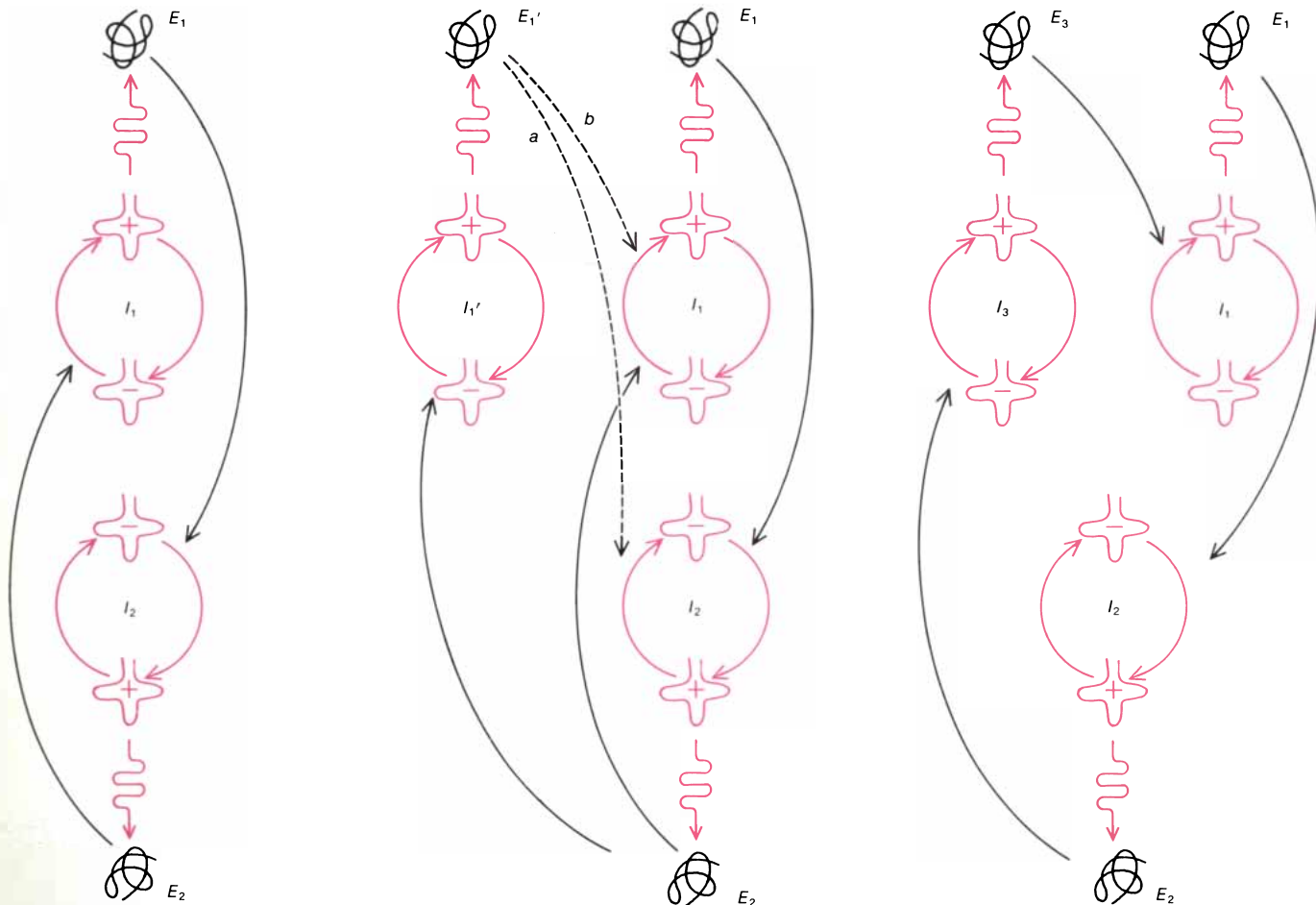
Life could not have originated with such a simple hypercyclic scheme as the one described above. The first catalytic couplings must have been weak and complex, and the number of genetic participants (members of RNA quasispecies) and functional participants (primitive enzymes) very large. The hypercyclic principle itself was nonetheless simple: enforced cooperation among otherwise competing genes allowed their mutual survival and regulated their growth. It also made possible a more refined kind of evolution than that open to quasispecies alone.

In a quasispecies Darwinian competition evaluates the fitness of each mutant RNA according to its rate and accuracy of self-replication and its stability. When quasispecies are locked into a hypercycle, however, new criteria come into play. First, evaluation of the "target function" of each quasispecies becomes crucial: those sequences are fittest that are best able to get themselves replicat-

ed quickly and accurately by the enzyme responsible for their replication. Second, the continual introduction of new mutant RNA sequences means that new catalytic couplings are constantly being tested. The structure of the hypercycle evolves whenever new couplings are discovered to be advantageous.

The hypercycle principle also shares an evolutionary disadvantage with the quasispecies. Both quasispecies competition and hypercyclic cooperation evaluate only phenotypic properties of the RNA's: their stabilities and reaction rates. If a way could be found to evaluate the genotypic properties of the RNA sequences—the genetic messages themselves—the quality of the enzymes resulting from translation of the sequences would be subject to improvement through natural selection. We see only one possible way: by putting the hypercyclically organized quasispecies distributions into compartments that could then evolve through Darwinian competition.

The transition from a single quasispe-



HYPERCYCLIC COUPLING allows cooperation among self-replicating RNA strands that would otherwise compete, with only the "fittest" surviving. In the hypercycle at the left an information-carrying RNA (I_1) codes for a primitive enzyme (E_1) that helps replicate another RNA (I_2), which similarly helps to replicate I_1 through its translation product (E_2). Such a cyclic catalytic coupling mechanism stabilizes the concentrations of coupled information carriers

and enzymes. In the middle hypercycle a mutant sequence (I_1') emerges that can outcompete I_3 ; the consequences depend on the functioning of the translation product (E_1'). If the new enzyme is more helpful to I_2 than the original E_1 is (path *a*) and if it has no effect on I_1 , then I_1' replaces I_1 in the first hypercycle. If instead E_1' is more helpful to I_1 than E_2 is (path *b*), the original hypercycle expands to become the three-member hypercycle shown at the right.

cies distribution to the hypercyclic organization of many distributions probably came about smoothly rather than at a stroke. Primitive translation mechanisms in a quasispecies distribution sooner or later gave rise to proteins that were more helpful to self-replication than the miscellaneous proteins in the soup. At first the help must have been more or less uniform throughout the quasispecies distribution, but as preferences among translation products became more pronounced, cross-assistance between sequences became immensely more probable than self-feedback. The initially complex mesh of interactions became more and more distinct as the advantages derived from more specific catalysis had effect. Finally the differences among the various template-catalyst interactions became so great that each enzyme had a particular catalytic role. At this point the original single quasispecies distribution had diverged to form a set of distinct quasispecies distributions, and the first hypercycle was in operation. Hypercycles

arose as naturally and continuously as quasispecies did; they arose under the force of natural law.

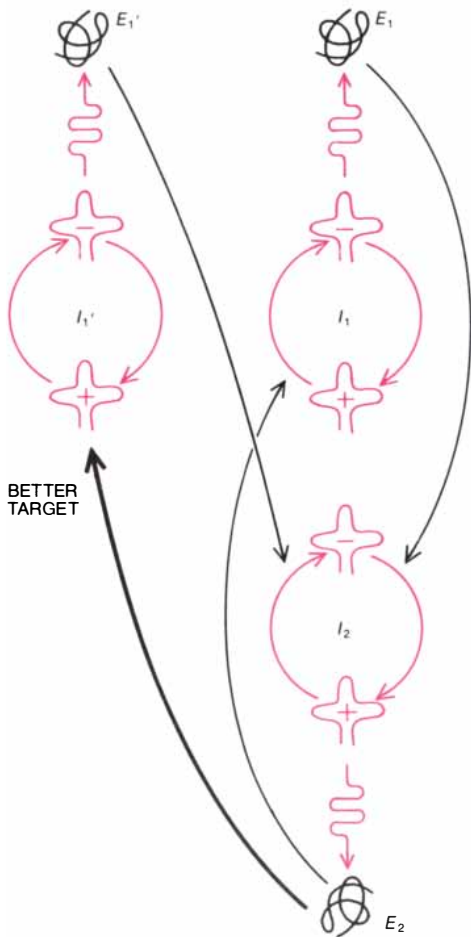
Compartmentation

Life now is everywhere cellular. Why? Among the obvious advantages of cellular organization are protection from fluctuations in the external environment and maintenance of concentration gradients, but such advantages do not explain the origin of the cellular-organization principle. Cellular organization was needed because it was the only way to solve the one problem of information processing in evolution that self-replicative competition and hypercyclic cooperation were unable to address: the evaluation of the information in genetic messages.

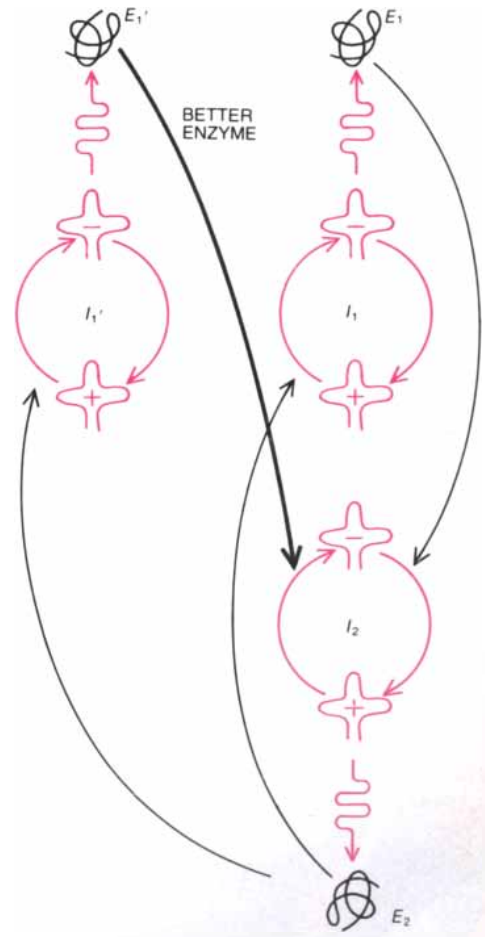
Organization into cells was surely postponed as long as possible. Anything that interposed spatial limits in a homogeneous system would have introduced difficult problems for prebiotic chemistry. Constructing boundaries, transport-

ing things across them and modifying them when necessary are tasks accomplished today by the most refined cellular processes. Achieving analogous results in a prebiotic soup must have required fundamental innovations.

Darwinian competition in a quasispecies distribution was based on selection according to the chemical kinetics of the sequences; what the sequences "meant" played no role. The meaning of the message could not be ignored when hypercyclic organization of enzymes and RNA's came into play, since the meaning governed the strength of the coupling. The one-directional character of the cyclic coupling, however, still excluded any feedback that would allow genetic meaning to be evaluated and so make possible selection of the best information. In a hypercycle as in a quasispecies only the target function of an RNA (its affinity for the proteins that replicate it) is evaluated, not the genetic information the RNA encodes. A hypercycle in solution cannot select for its translation products, whether those



PHENOTYPIC AND GENOTYPIC mutations played different roles in prebiotic evolution. A mutant version of the information carrier I_1 (I_1') has a phenotypic effect (*left*) if it supplies a better target for the enzyme E_2 and hence is more effectively replicated by E_2 than I_1 is. Provided that the new E_1' catalyzes the replication of I_2 , the hypercycle survives; if the new enzyme does not help to replicate I_2 , the mutation is "parasitic": the mutant (I_1') drives out I_1 and the



hypercycle is quenched. On the other hand a mutant version of I_1 has a genotypic effect (*right*) if its own replication by E_2 is not affected, but it gives rise to a product (E_1') whose activity in catalyzing the replication of I_2 is different from the activity of E_1 . Selection pressure favoring the superior version of E_1 could not arise in a homogeneous solution. Before such evolution was possible it was necessary for self-replicating systems to be isolated spatially in compartments.

Overcoming corrosion in an



Lockheed knows how.

The voracious process called corrosion consumes more than metal—it devours masses of money.

The nationwide cost of checking or controlling corrosion is now estimated at more than \$80 billion yearly. Airlines alone have a \$100-million annual corrosion problem.

So here are some of the many ways Lockheed works to blunt, check, or contain the corrosion attack.

Making aircraft metals fight back.

When developing the L-1011 TriStar, Lockheed was determined to lessen that \$100-million airline corrosion problem.

Thousands of Lockheed tests over 18 months, for example, led to the choice of a new wing-skin aluminum temper, 7075-T76. It's dramatically superior in corrosion resistance to widely used alloys and tempers like 2024-T3 and 7075-T6.

The value of that choice is clear. Wing skins using 2024-T3



Three aluminum specimens, treated with different tempers, were exposed to the elements for nine months at Lockheed's corrosion test site. The results: 7075-T6—deep, severe exfoliation, or 'leafing'. 2024-T3—also deep corrosion. 7075-T76—no corrosion damage.

for lower surfaces and 7075-T6 for upper surfaces—as most aircraft do—show severe corrosion in as little as two years. After eight years in service, however, the L-1011 wing skins, which uniquely use 7075-T76 on both surfaces, still show almost no signs of corrosion.

With similar attention given to all parts, smallest to largest, the L-1011 is by far the most corrosion-resistant large aircraft ever built.

Giving steel more sea-life.

Structures that live in the sea face a ravaging confrontation: salt water against steel.



Today, however, Lockheed's advanced marine cathodic systems are effectively protecting offshore oil platforms, loading terminals, and other kinds of marine structures around the world. And these systems are absolute state-of-the-art in such protection.

By feeding electric current to anodes on the underwater structures, the technique inhibits corrosive reaction, foils the seawater assault, and significantly extends structural life.

And now Lockheed engineers are developing advanced computer technology to plot the ideal amount and distribu-

\$80-billion combat zone.



Resembling a devastated landscape, 20-time enlargement of 2024-T3 aluminum sample (see small photo below) reveals the corrosion damage from nine months of exposure to the elements.

tion of protective current for each application. This will bring even greater savings by avoiding any overlapping of protection and by dramatically cutting down on inspection and maintenance.

Taking the sand out of blasting.

Hazardous to lungs and machinery alike, the process of sandblasting metals to remove surface corrosion or old paint may be coming to an end.

That's because Lockheed has developed an advanced system that uses frozen pellets of carbon dioxide—dry ice—to blast and clean metal surfaces.

The key to the technique is a Lockheed-patented system that makes the tiny, solid pellets in the right size, then delivers them at the right density and pressure for the surface being blast-cleaned. After striking the target surface, the volatile pellets quickly vaporize, and the vapor harmlessly dissipates. No infiltrating sand dust or other blasting agent debris remains. No cleanup of residual abrasives is required. And no blasting or toxic stripping agent harms the environment or human lungs.



Operator needs no protective mask in clean CO₂ blasting operations.

Best of all, the process (1) is comparatively low-cost, (2) can clean complex machinery without the disassembly needed for sand-blasting, and (3) can be used on anything from ships and bridges to delicate electronic assemblies.

Keeping alloys fresh and safe.

Beneficial spin-offs from aerospace technology are not uncommon. One of them, a protective sealing process called Duralox, evolved directly from Lockheed's anticorrosion work on the L-1011 TriStar's aluminum alloys.

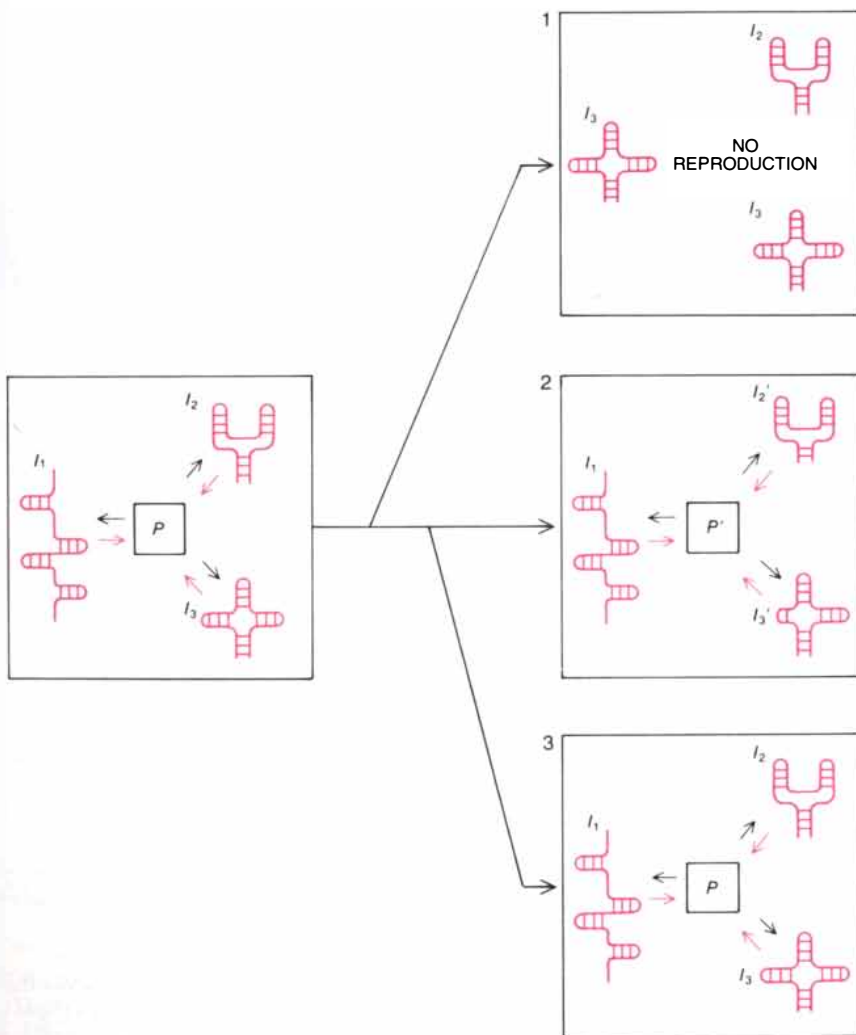
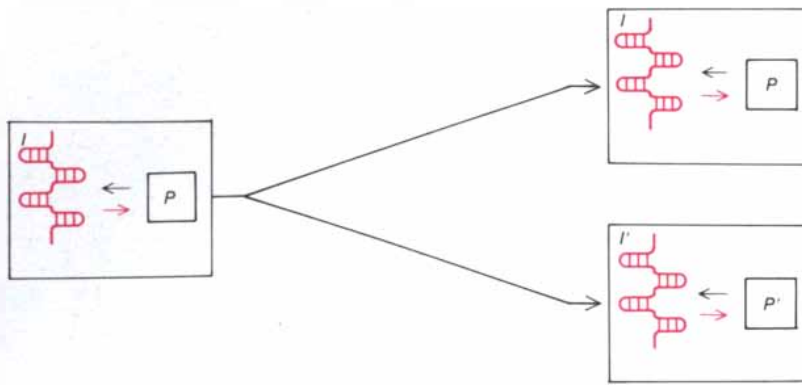
A two-bath chemical process, Duralox has proved remarkably effective on various anodic-coated aluminum alloys and thin anodic coatings exposed to even the toughest corrosive environments.

Usable anywhere corrosion is a problem, the process shows impressive beauty-keeping performance in new buildings that display large amounts of exterior aluminum sheet and extrusions.

One proof: Military and aircraft company specifications are aiming at 500 to 1,000 hours of surface survival in abusive salt-spray chamber tests. On many alloys, with very thinly coated surfaces, Duralox consistently betters 8,000 hours of such onslaught.

These are but a few examples of Lockheed's dedication to the fight against corrosion. When it comes to applying advanced technology to defeat or restrain corrosion—on land, in the air, in space, or in water—Lockheed knows how.





COMPARTMENTATION puts selection pressure on genotypes. Suppose information carrier I encodes machinery (P) for translation and replication (*top*). After mutation of I to I' and translation of I' to P' , the system may be recompartimentalized (*right*) into I/P and I'/P' daughter compartments. Such compartmentation allows selection of the gene encoding the more efficient machinery. For selection to occur, the superiority of the machinery must compensate for errors made in each round of replication. Compartmentation by itself, however, does not suffice to select for superior information divided among several information carriers (*bottom*). Daughter compartments with incomplete sets of components are not viable (1). Compartments with sets that are complete but include mutations (2) allow evolutionary change. Even if the original set is the optimal one (3), the rate of proliferation required to maintain stability is much higher than it would be in a homogeneous solution without compartmentation. The net effect, in other words, is a severe sharpening of the error-threshold condition.

products are advantageous or disadvantageous.

We see only one answer to the problem of evaluating the quality of the information in early genes. The answer was to provide a means of breaking up the homogeneity of the soup; it was to compartmentalize the evolutionary process. Once the mutational events in one compartment were made independent of contemporary events in other compartments, a way had been found to improve genetic information; the way, of course, was Darwinian evolution. Compartments that were fitter than others could be selected on the basis of their total performance, including the possession of better genetic information. As long as there was a way to pass on genetic information from one generation of compartments to the next, the evolution of the total information content was assured.

Now the flow of logic governing prebiotic evolution is complete. To attain any information stability at all required the self-replication of short RNA sequences. Darwinian competition among the mutant sequences led to a single quasispecies distribution as the potential product of evolution. Then hypercyclic organization arose among the mutant sequences and allowed many quasispecies distributions to coexist in the same soup. This expanded the amount and the variety of information present far beyond what had been allowed (by limited copying fidelity) in a single primitive gene, but it did not provide any opportunity for the information to be evaluated on the basis of its function in a competition that would lead to evolutionary improvement. That opportunity was provided by compartmentation and subsequent intercompartmental competition.

Compartmentation by itself could not have provided all that is logically necessary for the origin of life. In a compartment the problems of limited copying fidelity and of competition among self-replicating genes still had to be confronted. The only way so far discovered for maintaining enough genetic information to code for a minimal starting amount of enzymatic function is hypercyclic organization. Hypercycles and compartments addressed two independent problems of prebiotic evolution. Hypercycles provided for the stable coexistence of a variety of self-replicating genes and thereby solved the first information crisis. Compartments provided a way to evaluate and thus improve the genes' information content. In other words, compartments dealt with the genotype-phenotype dichotomy.

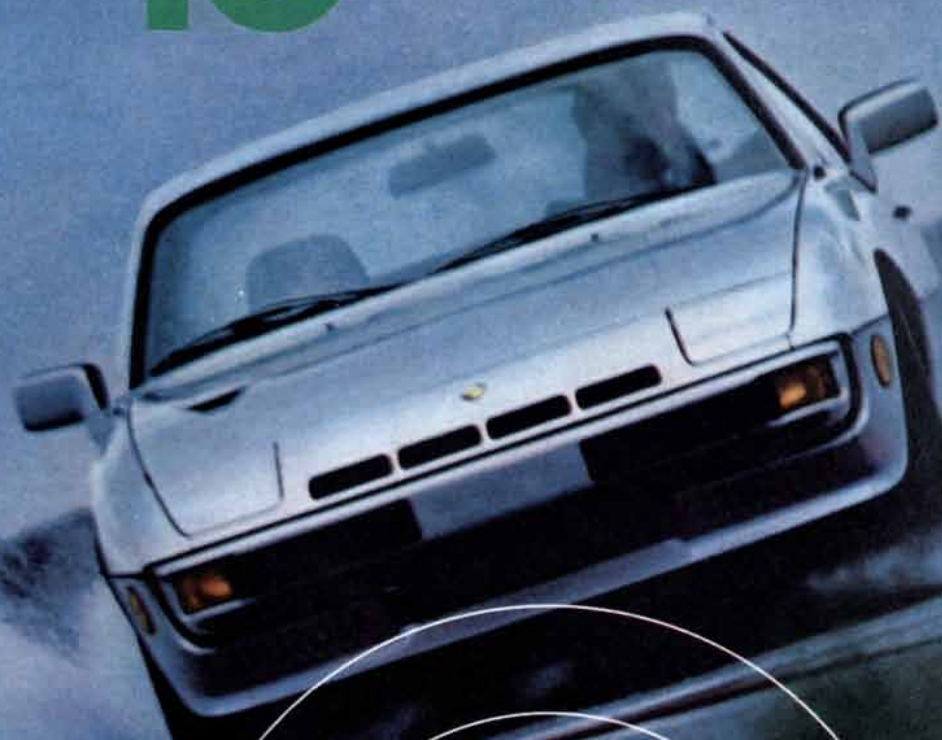
Life in the Test Tube?

If one can really deduce the natural laws that operated to create life on the

Consumer Orientation
No. 10 in a Series
Subject:
The Dynamic Response
of Tires

10

Porsche 924 Turbo



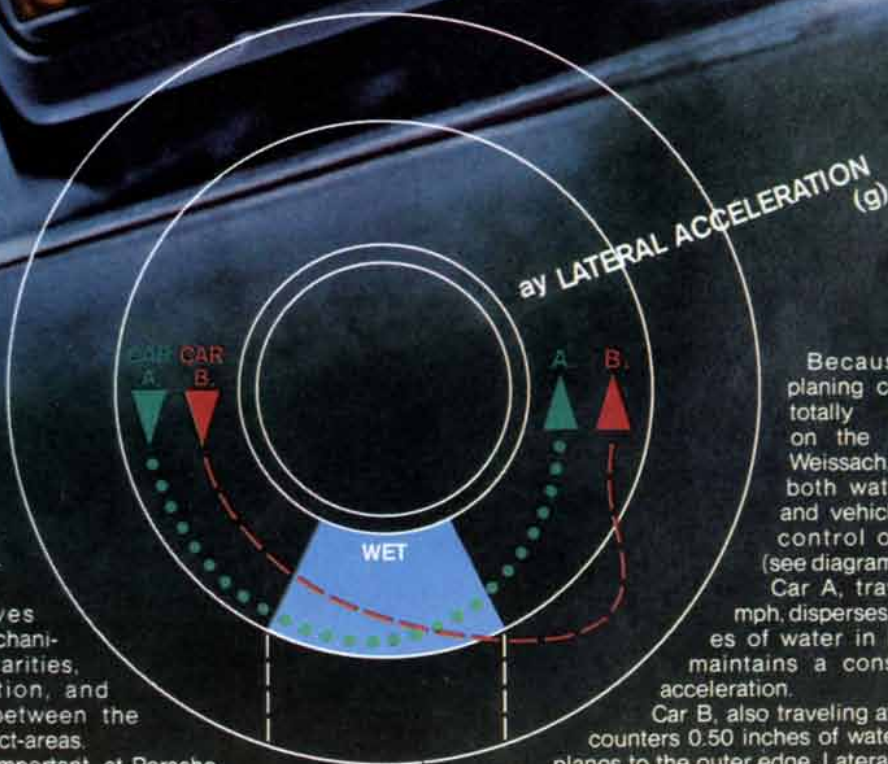
An automobile propels itself and changes direction through reaction to the ground.

Most of the external forces that act upon it are generated through and limited by the friction of its tires.

The tires themselves generate friction by mechanical gripping of irregularities, high-speed deformation, and molecular adhesion between the rubber and ground contact-areas.

Because tires are so important, at Porsche we do more than merely specify their type and size. We work with tire manufacturers in their design and testing.

Consider our hydroplaning tests. Hydroplaning occurs when a vehicle's tires cannot displace water from the tire-ground contact area fast enough to allow them to make contact with the ground. Instead, the vehicle's tires actually ride up on a film of water. And, as a result, traction, directional stability, and control are greatly reduced.



Because hydroplaning can never be totally eliminated—on the skid pad at Weissach, we test for both water dispersal and vehicle directional control on recovery (see diagram at left).

Car A, traveling at 45 mph, disperses the 0.25 inches of water in its path and maintains a constant lateral acceleration.

Car B, also traveling at 45 mph, encounters 0.50 inches of water and hydroplanes to the outer edge. Lateral acceleration diminishes during hydroplaning, but returns on recovery.

What we learn from our hydroplaning tests, we incorporate into the tires, the suspension, the total car. Because in over 30 years of competition driving, we've learned it's the sum of an infinite number of details that separates winning from all else.

Test drive the Porsche 924 Turbo today. For your nearest dealer, call toll-free: (800) 447-4700. In Illinois, (800) 322-4400.

PORSCHE + AUDI
NOTHING EVEN COMES CLOSE

Our widely accepted *UNIX*[™] operating system is simplifying the use of computers. Designed to handle a variety of applications, it is being used to manage and maintain the telecommunications network, control experiments, develop software, process text, prepare documents, and teach computer science.

The *UNIX* system allows users to take small programs and assemble them like building blocks to perform complex tasks. In text processing, for example, the command "Spell Bell Labs Ad" tells a computer to proofread this ad against a dictionary filed in its memory. The program that performs the task was created by simply combining several smaller *UNIX* programs.

Another useful feature of our *UNIX* operating system is its ease in typesetting mathematical expressions. To typeset an equation like

$$\sum_{i=1}^{\infty} \frac{1}{x_i} = \pi$$

someone need only type: "Sum from i = 1 to infinity 1 over x sub i = pi." The computer does everything else.

Since 1969 the Bell System has installed more than 1100 *UNIX* systems. Along with other Bell Labs innovations in computing technology, these systems are enhancing the way the nation's telecommunications network is planned, designed, built, and operated. Through licensing agreements with Western Electric, universities have installed over 800 *UNIX* systems, and government and

industrial facilities are using over 400.

The *UNIX* operating system can be used with computers of different manufacturers because it is small, cleanly designed, and written in a general-purpose programming language. Such portability in a computer operating system saves time and money.

Building on the past

The *UNIX* system is just one of many Bell Labs advances in computing science and technology over the years. Among our innovations:

- Application of telephone switching concepts and technology to early computers
- First demonstration of remote computer terminal and data link
- Conception of electronic analog computer
- First design of AND and OR gates for diode circuitry
- Formulation of Information Theory
- Invention of error-detecting and error-correcting codes
- Demonstration of first general-purpose transistorized digital computer
- Development of computer operating systems
- Design of computer languages, including ALTRAN, SNOBOL, L6, and C
- Creation of computer graphics techniques for storing,

manipulating, and presenting information

- Development of Fast Fourier Transform
- Design of central processors for switching systems having virtually no downtime

Looking ahead

Computing technology is having a major impact on the telecommunications business. It's increasing the Bell System's ability to provide new services and handle existing ones more efficiently. Today, for example, the nationwide telecommunications network links thousands of software-controlled electronic systems, making it by far the world's largest distributed processing network. We and our partners—Western Electric and the telephone companies of the Bell System—are putting technology to work so that the network will continue to evolve and expand its information-handling capabilities. The object, of course, is to meet the fast-growing and changing needs of Bell System customers.

For information about employment, write:

*Bell Laboratories, Room 3C-303,
600 Mountain Avenue,
Murray Hill, N.J. 07974*

Please include a resume.

An equal opportunity employer.



Bell Laboratories

From Science: Service

earth, why not just assemble the necessary materials and re-create life in a test tube? Anyone attempting such an experiment would be seriously underestimating the complexity of prebiotic molecular evolution. Investigators know only how to play simple melodies on one or two instruments out of the huge orchestra that plays the symphony of evolution. The investigator substitutes for a single instrument much as an amateur musician does with a "music minus one" record.

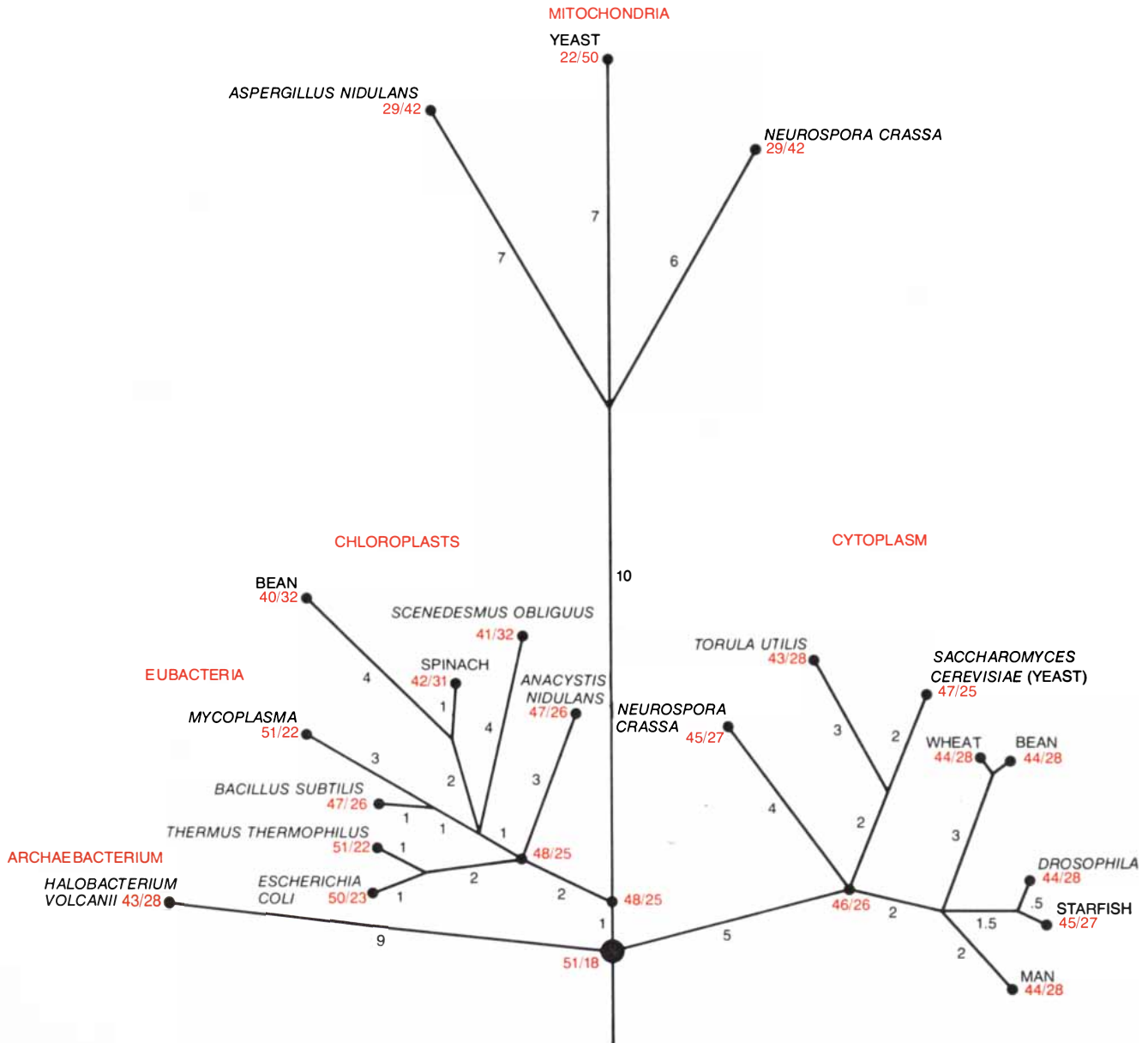
In this article, for example, we have described some test-tube experiments on molecular evolution that showed

how RNA templates with quite sophisticated phenotypic properties originate when the necessary enzymatic machinery is present as an environmental factor. The next step is to find out how such machinery evolved in the first place. We would like to make all its components and test them experimentally.

What amino acids were components of the first proteins synthesized by RNA instruction? In the present-day genetic code each of the 64 (4^3) possible triplets of the four RNA nucleotides is a "codon" that tells the translation machinery to add a particular one of the 20 amino acids to the protein chain (or to start or

stop the translation process). That is surely much too complex a system to have been initiated in one step by natural causes. Was there once a more primitive precursor code? What was its structure? These questions have to be addressed before experiments on the self-organization of translation can be designed.

Primitive proteins must have been made from fewer amino acids, and so an initial one- or two-character code would have been sufficient to assign them all. There is no chemically simple way, however, to change from a singlet or doublet codon to a triplet one because in



PHYLOGENETIC TREE for the transfer RNA adapted to the "start" codon reveals only small changes (black numbers) in its nucleotide sequence over billions of years. The sequence is almost the same in all mammals studied to date; there are only a few differences between man and the fruit fly *Drosophila*. The numbers in color give the GC/AU ratio: the ratio of guanine and cytosine to adenine and

uracil. The ratio is highest close to the earliest nodes of the tree and is smallest at the ends of long branches. In the sequences of transfer RNA's in the cellular organelles called mitochondria the GC/AU ratio is about 1:2, compared with 2:1 near the early nodes. This shows that a high ratio is not required today; it may have been required originally because GC pairing is more stable than AU pairing.

**Every week,
we send
Nature to
the world's
most
important
science
addresses.**

Scientists who are leaders in their field and who get to reap the rewards of their achievements are among *Nature's* most eager readers.

And they are renewing their subscriptions to *Nature* in record numbers.

If, like them, you keep searching for ways to speed up your research, or if you're looking for advancement where you work, or are thinking of moving on to something more stimulating, how can you afford to be without your own personal subscription to *Nature*?

It's the most certain and consistent way to get the research results, news, opinion, and analysis you need, while they're still fresh and can be applied most usefully to your own work.

With your own *Nature*, you'll be well ahead of the monthly journals whose reports often lag a year or more behind ours. And there isn't a science weekly anywhere that covers research developments in as many parts of the world, as fully and authoritatively, as *Nature* does.

The good news is that we've cut *Nature's* personal subscription rate in half and, to top it off, if you subscribe for a full year, we'll send you, free, a copy of *Nature's* new, annual *Directory of Biologicals*. Planned for fall 1981, the *Directory* will be the first truly worldwide buyers' guide to bio-



**Now: *Nature*
at ½ price – plus a
bonus book, free**

logical products. It will be available only to yearly subscribers (or, for \$45, to buyers of the hardcover version, in bookstores).

A personal subscription to *Nature* has never been more attractive. We urge you to clip and mail the coupon today. For credit-card orders only, **call toll-free now: (800) 824-7888.**

Ask for operator 130.

In California: (800) 852-7777
Ask for operator 130.

To: **Nature** 15 East 26th Street,
New York, N.Y. 10010

- Please send me a year of *Nature* at ½ price, \$86.50, plus my free *Directory of Biologicals*
- I prefer to try *Nature* for 6 months: \$50. 3 months: \$30
- I enclose my *personal* check
- Please bill my credit card account

Account No. _____

Visa Master Amer. Ex. Exp. Date _____

Name _____

Address _____

City _____ State _____ Zip _____

Orders *must* include personal check or credit card data. Offer good in the U.S. and Canada and is subject to change without notice. In the U.K., send personal check for £37.50, or credit card data, to: *Nature*, 4 Little Essex St., London WC2R 3LF.

such a transition all existing messages would be nonsense until they were completely recoded. The genetic code must therefore have been based on a triplet frame from the start. What was that frame?

Prebiotic translation imposed on the code some tasks that today's translation mechanism solves by refined measures not connected with the code itself. At first the code itself had to establish the direction of readoff and the punctuation of the message by defining a "reading frame." In 1976 F. H. C. Crick, Sydney Brenner, Aaron Klug and George Piecznik of the Medical Research Council Laboratory of Molecular Biology in Cambridge proposed that directionality and framing were initially fixed by translating only triplets that had the se-

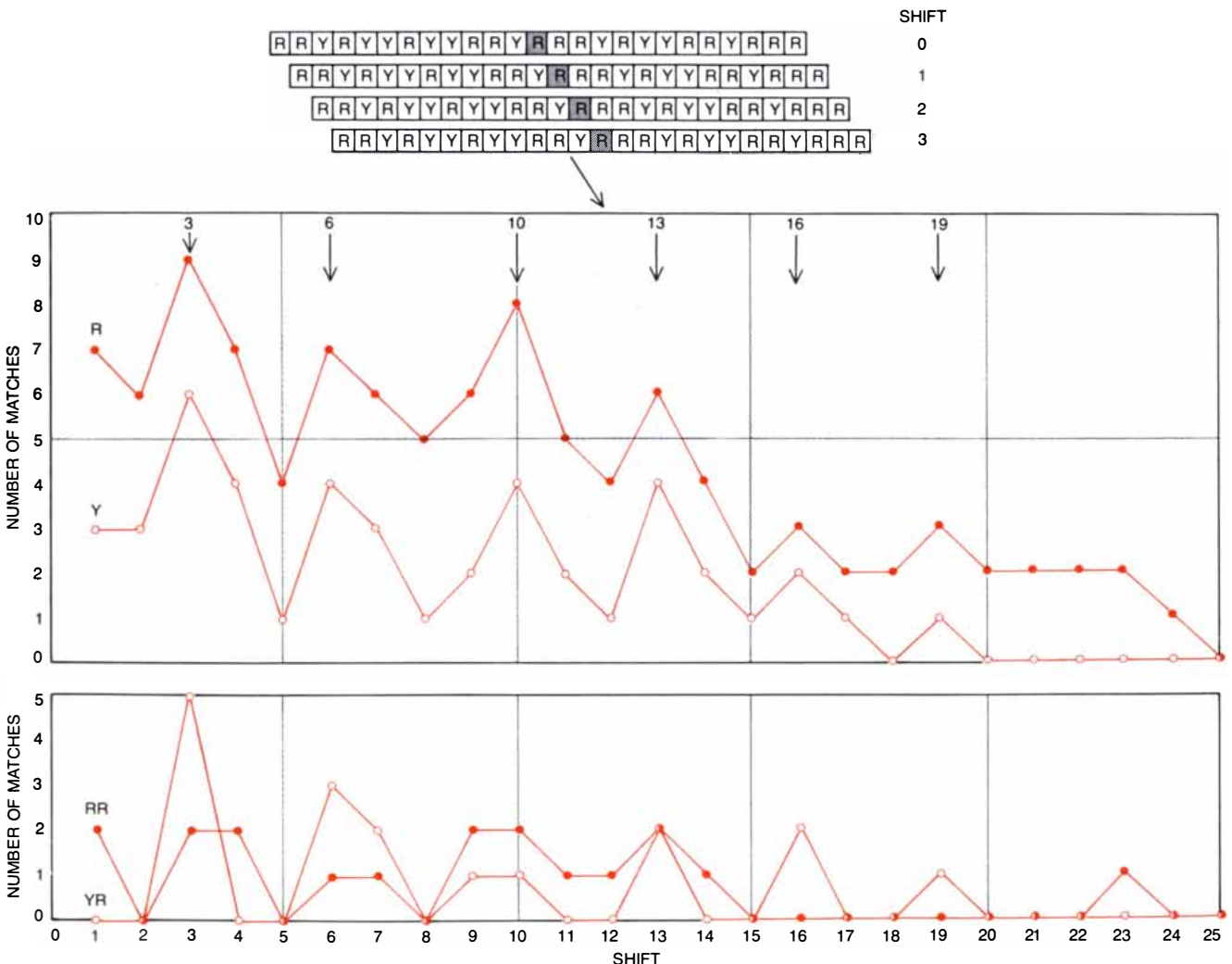
quence *RRY*; *R* signifies a purine nucleotide (*G* or *A*) and *Y* a pyrimidine (*C* or *U*). They also noted that *RNY* triplets would serve the same purpose, with *N* signifying any nucleotide. Directionality and punctuation are established by *RNY* just as well as by *RRY*, and with *RNY* the same frame pattern would apply to both the template strand and its complementary daughter strand.

Evolution of the Code

Are there any clues in the present code and machinery as to whether it arose from frames such as *RRY* or *RNY*? The creation of a computer archive of sequence comparisons by Margaret Oakley Dayhoff and her co-workers at the National Biomedical Research

Foundation has made it possible to undertake large-scale searches for genetic relations among biological polymers, and in particular to construct phylogenetic trees showing relations among homologous proteins or nucleic acids in different species. Transfer RNA's are particularly appropriate for such analysis with respect to the question of origins. Their function is to adapt each amino acid to its codon. Given such a key role, their structure might still reflect how the assignments of codons to amino acids came about.

In searching for primitive sequences by analyzing present ones it was not enough simply to process a lot of sequence data with a computer program designed to find the optimal phylogenetic tree. Analytical criteria to determine



FREQUENCY ANALYSIS OF GENES employs a method devised by John Shepherd of the University of Basel. Such an analysis is illustrated for a sequence of *RNY* codons with one extra base (inserted at position 13) and replacement of a *Y* by an *R* at position 25. *R* stands for a purine (either adenine or guanine), *N* for any of the four bases and *Y* for a pyrimidine (either cytosine or uracil). A computer program shifts the sequence to the right one base at a time (top) and at each step counts the number of matches (vertical agreements) of single bases and of *RR* and *YR* pairs. Peaks in the number

of matches reflect repetitions within the sequence. The single-base peaks are seen to come at intervals of three (with a perturbation caused by the insertion), as do the peaks for *YR*, which is implicit in an *RNY* code. Much longer sequences of biological RNA and DNA have been examined for such correlations and have been carefully checked for extraneous effects. The presence of some correlations and the absence of others can be interpreted as reverberations of the ancient codon *RNY*, supporting theoretical arguments for that codon and also supporting results of the transfer-RNA study.



1. USE IT



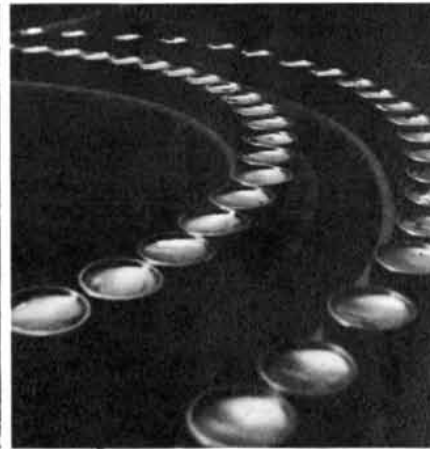
2. SAVE IT



3. RETURN IT



4. REMELT IT



5. REMAKE IT



6. REUSE IT

The aluminum can. Recycling it conserves energy and natural resources.

It's light, quick chilling and convenient. And it's fast becoming America's most popular beverage container. But that's only half the story. Because by returning used aluminum cans for recycling, Americans are really helping to save energy.

Every time an aluminum can is recycled, we save 95 percent of the energy needed to make new metal from bauxite. Since 1972 the aluminum industry has received over 2.2 billion pounds of used aluminum beverage cans for recycling.

By putting those cans back to work, enough energy was conserved last year alone to meet the residential needs of a city like Pittsburgh,

Pennsylvania for over a year. And recycling also helps Alcoa® reduce its consumption of coal and bauxite, both valuable natural resources.

Recycling is growing fast. In the last eight years, thousands of civic groups and individual collectors have earned over \$400 million through recycling. And recycling has added an estimated 20,000 new jobs to the economy — jobs that didn't previously exist.

For additional information about this exciting new industry that saves more and more energy as it grows, write Aluminum Company of America, 523-D Alcoa Building, Pittsburgh, PA 15219.

We can't wait for tomorrow.



the degree of "treelikeness" in a set of sequence data had first to be formulated and confirmed, and this was done through topological analyses in cooperation with Andreas Dress of the University of Bielefeld.

When these criteria and programs like Dayhoff's were applied to, the analysis of all currently known transfer-RNA sequences (about 200), two fascinating conclusions emerged. First, the sequences of a given transfer RNA (for example the one that mediates initiation of translation) in all species studied do appear to be related to one another in a treelike pattern that reflects only small evolutionary divergence compared with the changes in other biological polymers. Apparently this particular early information has survived later stages of evolution quite well. Second, the sequences of different transfer RNA's in a single organism reflect divergence from a common ancestor, but they seem not

to be related to one another in a treelike manner, at least not in the two organisms (*E. coli* and yeast) for which we have enough sequence data to make the analysis statistically meaningful.

The sequences seem instead to represent a mutant distribution similar to that of a quasispecies. In a kind of back-extrapolation through the quasispecies' history to the time of the origin of translation, the analysis identified what appeared to be ancestors of modern transfer RNA's and led to two important inferences about them. They were much richer in *G* and *C* than in *A* and *U*, and in their master sequences (which were determined by assigning the most prevalent base at each position) they showed a clear reverberation of a triplet pattern of the form *RNY*.

One can look for ancient genetic information in other places: wherever there is evidence that selection pressure and genetic drift have not yet reduced

the "remembrance" of ancestral sequences to below the noise level. John Shepherd of the University of Basel recently applied a new method of computer-aided sequence analysis suitable for long genetic messages. His method measures the distance between repetitions of characters or groups of characters along a sequence. Ancestral information can be distinguished from later modifications. His first conclusions, drawn from studies of several DNA viruses and genes of bacteria and higher organisms, are again that there is in these modern genes still a memory of ancient sequences and that the triplet *RNY* dominated those sequences.

The stability of the *G-C* pairing strongly suggests that the initial *RNY* code must have been limited to the four *GNC* codons. The present assignments of those codons are *GGC* = glycine, *GCC* = alanine, *GAC* = aspartic acid and *GUC* = valine. Simulations of primordial chemistry done by Stanley L. Miller of the University of California at San Diego suggest that these amino acids were among the most prevalent ones in the primordial soup. If that is a coincidence, it is certainly a suggestive one.

We have crossed the threshold of being able to have confidence in our ability to reconstruct ancestral RNA and protein sequences. On the basis of this information we are beginning to reconstruct and resynthesize ancestral sequences, both of proteins and of RNA's, and to test their interplay in a continuous-flow reactor, or what amounts to an evolution machine.

If the first proteins were indeed made of the four amino acids we mentioned, they had a negative electric charge. In general such amino acids would not readily associate with a negatively charged RNA species unless specific forces stabilize a particular interaction. Claude Hélène of the University of Orléans has shown that there is a strong specific interaction between the carboxylate (COO^-) groups of such amino acids as aspartic acid and the *G* nucleotides of RNA. Hence particular sequences can indeed build up patterns for specific contacts, which then may be stabilized with the help of metal ions. The first specific catalysts for replication and translation were probably just such structures, mediating specific contacts and thereby supporting weak chemical functions.

All such functions must have been recruited from the information of an initial quasispecies, the mutants of which finally differentiated when they organized themselves into hypercyclic functional linkages. The principles guiding the evolution of such an organization have been formulated and experimentally verified. Now what remains to be discovered is just what the favorable molecular structures were.

	FIRST POSITION	SECOND POSITION				THIRD POSITION	
		G	C	A	U		
<i>N</i> CODE IN <i>GNC</i> , <i>GNV</i> FRAMES		GLY	ALA	ASP	VAL		
<i>RN</i> CODE IN <i>RNY</i> FRAME	<i>G</i>	GLY	ALA	ASP	VAL		
	<i>A</i>	SER	THR	ASN?	ILE		
<i>RNN</i> CODE	<i>G</i>	GLY	ALA	ASP	VAL	<i>Y</i>	
		GLY	ALA	GLU	VAL	<i>R</i>	
	<i>A</i>	SER	THR	ASN	ILE	<i>Y</i>	
		ARG	THR	LYS	ILE/MET	<i>R</i>	
	<i>NNN</i> CODE	<i>G</i>	GLY	ALA	ASP	VAL	<i>C</i>
			GLY	ALA	ASP	VAL	<i>U</i>
GLY			ALA	GLU	VAL	<i>G</i>	
GLY			ALA	GLU	VAL	<i>A</i>	
<i>A</i>		SER	THR	ASN	ILE	<i>C</i>	
		SER	THR	ASN	ILE	<i>U</i>	
		ARG	THR	LYS	MET ("START")	<i>G</i>	
		ARG	THR	LYS	ILE	<i>A</i>	
<i>C</i>		ARG	PRO	HIS	LEU	<i>C</i>	
		ARG	PRO	HIS	LEU	<i>U</i>	
		ARG	PRO	GLN	LEU	<i>G</i>	
		ARG	PRO	GLN	LEU	<i>A</i>	
<i>U</i>	CYS	SER	TYR	PHE	<i>C</i>		
	CYS	SER	TYR	PHE	<i>U</i>		
	TRP	SER	"STOP"	LEU	<i>G</i>		
	"STOP"	SER	"STOP"	LEU	<i>A</i>		

EVOLUTION OF THE GENETIC CODE may have begun with the assignment of the four bases *G*, *C*, *A* and *U* in the middle position of a triplet frame to the four most abundant amino acids then present. The triplet had the form *GNC*, where *N* is any base. *U* later became an alternative to *C* because *G* can sometimes pair with *U*, giving a *GNV* frame. (*Y* is either *C* or *U*.) The presence of *U* in the third position made its complement, *A*, another possibility for the first position, leading to an *RN* code in an *RNY* frame. (*R* is either *A* or *G*.) The penetration of *R* into the third position to give an *RNV* code allowed *Y* to penetrate the first position of complementary strands, expanding the set to the present-day code of 64 triplet codons *NNN*.

Professional achievements require more than the technical knowledge that suggests their possibility. Achievements also demand that professional people act on their special knowledge with confidence and creativity. To do this effectively, they need to be free of the fears their liability as professionals can create. Among these is the fear that even a responsible action might backfire and threaten their professional reputation.

For more than a quarter of a century, the CNA Insurance Companies have been committed to easing fears and protecting reputations with professional liability insurance. CNA's years of experience assure professionals the kind of protection they can depend on. Our profes-

sional liability specialists can custom-tailor programs to meet the specific needs of various professional groups. They are ready to work with professional groups to develop loss control programs which improve professional performance, reduce accidents and help control insurance costs.

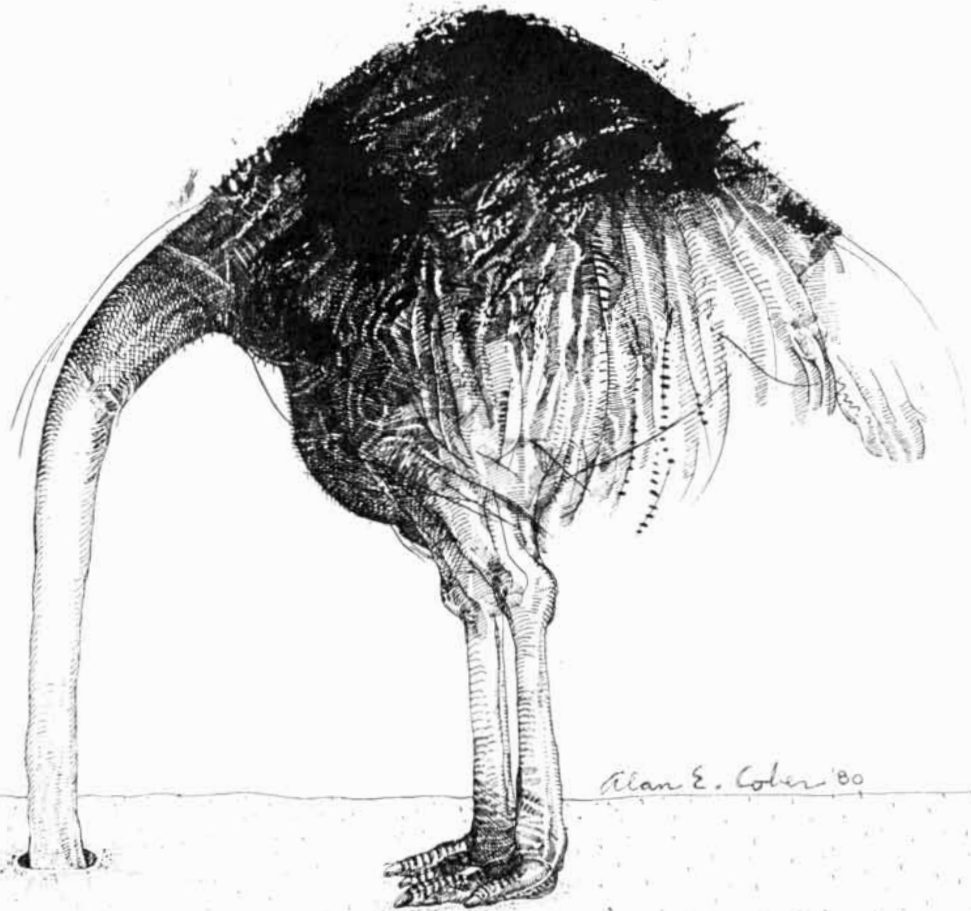
Learn how CNA's professional liability insurance can help you and the members of your

professional group meet your responsibilities with confidence and creativity. Ask your independent insurance agent or your association executive to

contact Vice President, Professional Liability, CNA, CNA Plaza, Chicago, IL 60685.

"HE WHO FEARS BEING CONQUERED IS SURE OF DEFEAT."

Napoleon Bonaparte



CNA Professional Liability Insurance protects professional group members including: □ Doctors □ Attorneys □ Architects □ Engineers □ Accountants □ Corporate Directors and Officers □ School Boards □ Hospitals □ Medical Clinics □ Publishers □ Broadcasters

INSURANCE FROM
CNA

The Shells of Novas

A nova (as distinct from a supernova) is a white-dwarf star that blows off a shell when a companion spills fresh nuclear fuel on it. The shell's spectrum is a clue to what happened

by Robert E. Williams

Most stars expend their supply of nuclear fuel at a remarkably constant rate for hundreds of millions or billions of years. Roughly once every 10 years, however, one of the stars in our region of the Milky Way galaxy suddenly increases in brightness by a factor of between 10,000 and a million, and for a brief period it rivals the brightest stars in the sky: it becomes a nova. The most recent bright nova appeared in the constellation Cygnus in the late summer of 1975.

The overall incidence of novas in our galaxy, consisting of more than 100 billion stars spread over a disk some 80,000 light-years in diameter, is difficult to ascertain primarily because of the presence of interstellar dust. In the absence of dust the brightest of novas should theoretically be visible to the unaided eye out to a distance of 25,000 light-years. A typical nova in a region heavily obscured by dust, however, might flare up as close as 1,000 light-years away and never be noticed. The best estimate of the incidence of novas has been obtained by monitoring their occurrence in nearby spiral galaxies. Such surveys have also demonstrated that two qualitatively different types of stellar outburst can be distinguished: novas and supernovas. In a typical galaxy there are about 25 novas every year but only about two or three supernovas per century. No supernova has been seen in our galaxy since 1604. Supernovas are tens of thousands of times brighter and more energetic than novas.

The two phenomena are completely different and are not related to each other. A supernova represents the brief final event in the evolution of a massive star. At sufficiently high temperatures and pressures the nuclei of heavy elements in the core of the star participate in a series of reactions that withdraw energy from the star, causing it to contract rapidly and then explosively bounce back outward. In the process the star is disrupted except for a high-density central core, which is left behind as a remnant in the form of a neutron star. Novas, on the other hand, are not part of

the evolution of normal stars. They are thermonuclear episodes that take place, probably periodically, on the surface of a white-dwarf star that is very closely bound gravitationally to a much larger, expanding cool star.

In spite of their basic differences novas and supernovas do share one trait: both eject gaseous material into space. Remnants of supernova explosions in the form of filamentary nebulas may persist for centuries. The Crab Nebula in Taurus is the remnant of a supernova that exploded in A.D. 1054. The shells ejected by novas are smaller and less massive than supernova remnants, and they can usually be seen only around the nearer and brighter novas. Recent spectroscopic examination of nova shells has provided clues to the types of thermonuclear reactions that underlie the nova's explosive release of energy. These studies have shown that nova shells are richer than normal stars in heavier elements such as carbon, nitrogen and oxygen.

The time required for a nova to reach maximum luminosity can vary considerably: it can be as brief as two days or as long as several months. A graph of the visual brightness of a nova over a period of time is called the light curve of the nova. The light curves distinguish two general classes of novas: fast and slow. Fast novas usually increase in brightness by a factor of more than 100,000 in only a few days. Their peak brightness is maintained for less than a week and then declines steadily. Initially the decline of a fast nova is fairly rapid: a decrease of as much as one stellar magnitude (a factor of 2.512) every two days. Slow novas reach their maximum brightness more gradually and more erratically than fast novas and decline much more slowly. Moreover, their increase in brightness is less. The total energy released in an outburst, however, is roughly the same for both kinds of nova. As best as can be determined novas eventually return to the same brightness they had before the outburst.

Novas are initially named after the

constellation in which they are observed and the year they erupt. Thus the 1975 nova in Cygnus, which for several nights was as bright as the nearby first-magnitude star Deneb, was called Nova Cygni 1975. Later the nova is given an official name: the Latin genitive for the constellation is prefixed either by letters or by a V (for variable) and a number signifying the order of discovery of the variable stars in that constellation. Hence Nova Cygni 1975 is now officially known as V1500 Cygni, and another well-known bright nova, Nova Herculis 1934, is known as DQ Herculis.

Our current understanding of nova outbursts can be traced to the discovery in 1954 by Merle F. Walker of the Mount Wilson and Palomar Observatories that DQ Herculis is an eclipsing binary, or double-star, system. Walker was monitoring the light output of DQ Herculis in an effort to determine the nature of the flickering that many old novas exhibit when he observed that the star's brightness decreased significantly for almost an hour and then returned to its original intensity. Further observations established that the same pattern was repeated every four hours 39 minutes; the star was unmistakably being eclipsed by an unseen orbiting companion star. The discovery that DQ Herculis was in fact a double star, with the shortest period of revolution of any binary system known at the time, made it possible to determine some fundamental properties of the two stars.

A well-known law in astronomy, which can be derived from Newton's laws of motion and gravity, states that the separation between any two gravitationally bound objects in orbit around each other can be calculated from their masses and their orbital period. The masses of most stars fall within a fairly narrow range, from about a tenth the mass of the sun to 10 times the mass of the sun, so that the distance between the two bodies in the DQ Herculis system could be estimated rather closely even without knowing the specific masses of the two stars. Because of the very short period of the system it was evident the

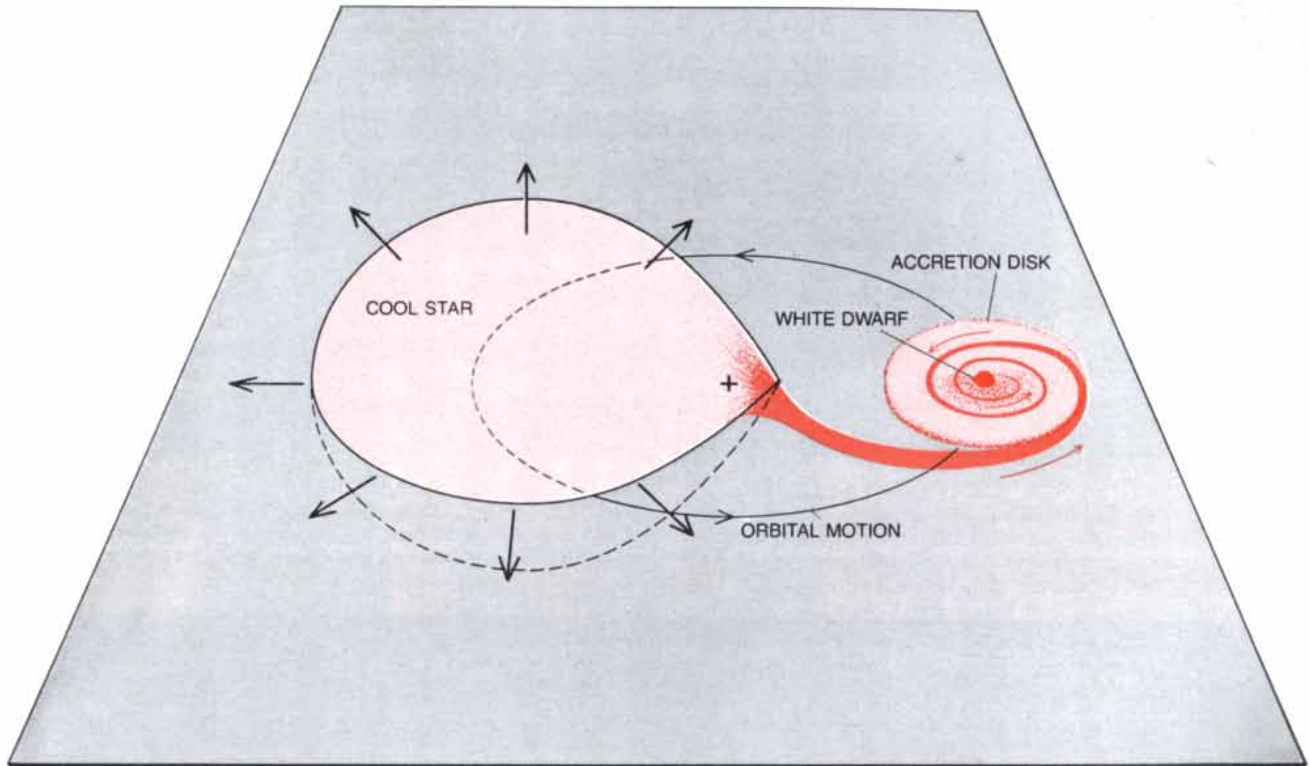
two stars must be very close to each other, separated by little more than the diameter of a typical star.

Astronomers found this result exciting. Was it just a coincidence that the nova of 1934 was a member of a very close binary system? Or was the outburst somehow related to the close proximity of the two stars? A search was begun to detect possible companion stars

in other known nova systems. All known novae are too far away for any close double-star systems to be directly resolved into the individual stars, and so indirect detection methods had to serve. The existence of binary systems can be deduced in two ways: by one star's eclipsing the other, as in the case of DQ Herculis, or by the observation of a Doppler shift in the spectral lines of a

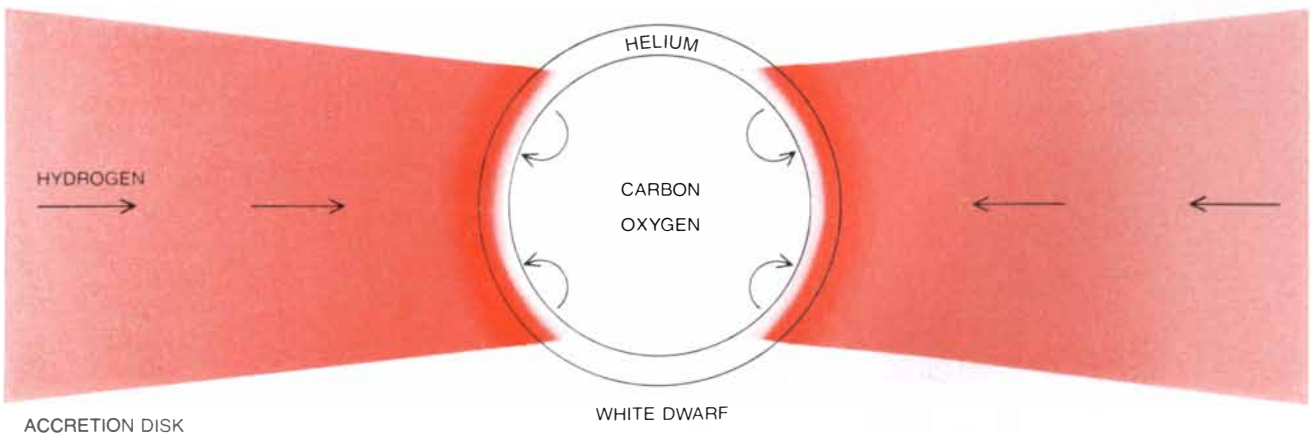
star in such a system resulting from the star's motion around its companion.

Robert P. Kraft of the Mount Wilson and Palomar Observatories studied 10 old novae and found that seven showed definite evidence of the presence of a previously undetected companion star. In most of the instances where the orbital periods of the binary systems could be determined they were less than a day.



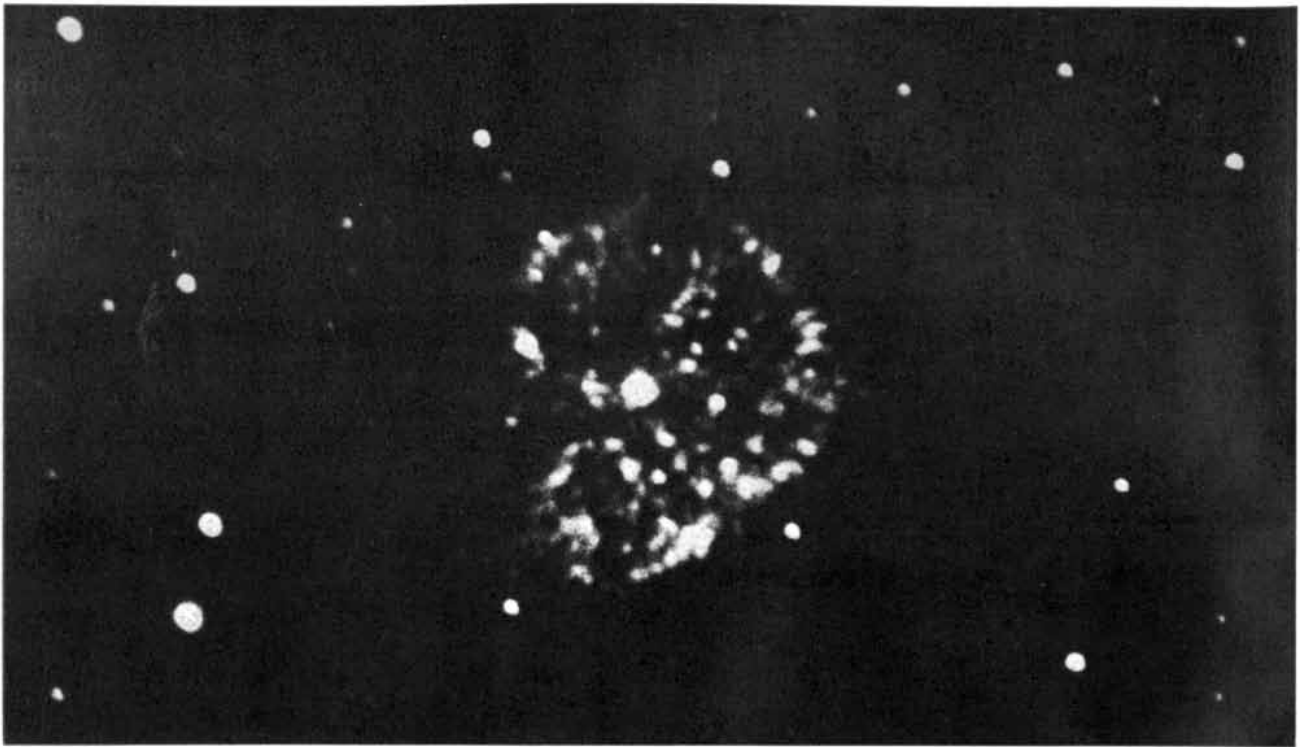
NOVA BINARY SYSTEM is represented schematically from above the plane of the orbits. The cross is the center of rotation. The white-dwarf member of the system is a star that has converted virtually all its initial supply of hydrogen and helium into heavier elements. Lacking an energy source, it has contracted to a diameter roughly that of the earth. Its companion is a star of normal size that has exhausted most of the hydrogen in its core and has begun to expand as thermo-

nuclear reactions have moved outward from the center toward the surface. As the star expands on its way to becoming a red giant its shape is distorted by the gravitational attraction of the white dwarf. Hydrogen gas streaming off the expanding star is swept into an accretion disk that swirls around the smaller star. Ultimately most of the gas spirals into the white dwarf at high velocity, raising the temperature of the dwarf's surface and setting the stage for the outburst.



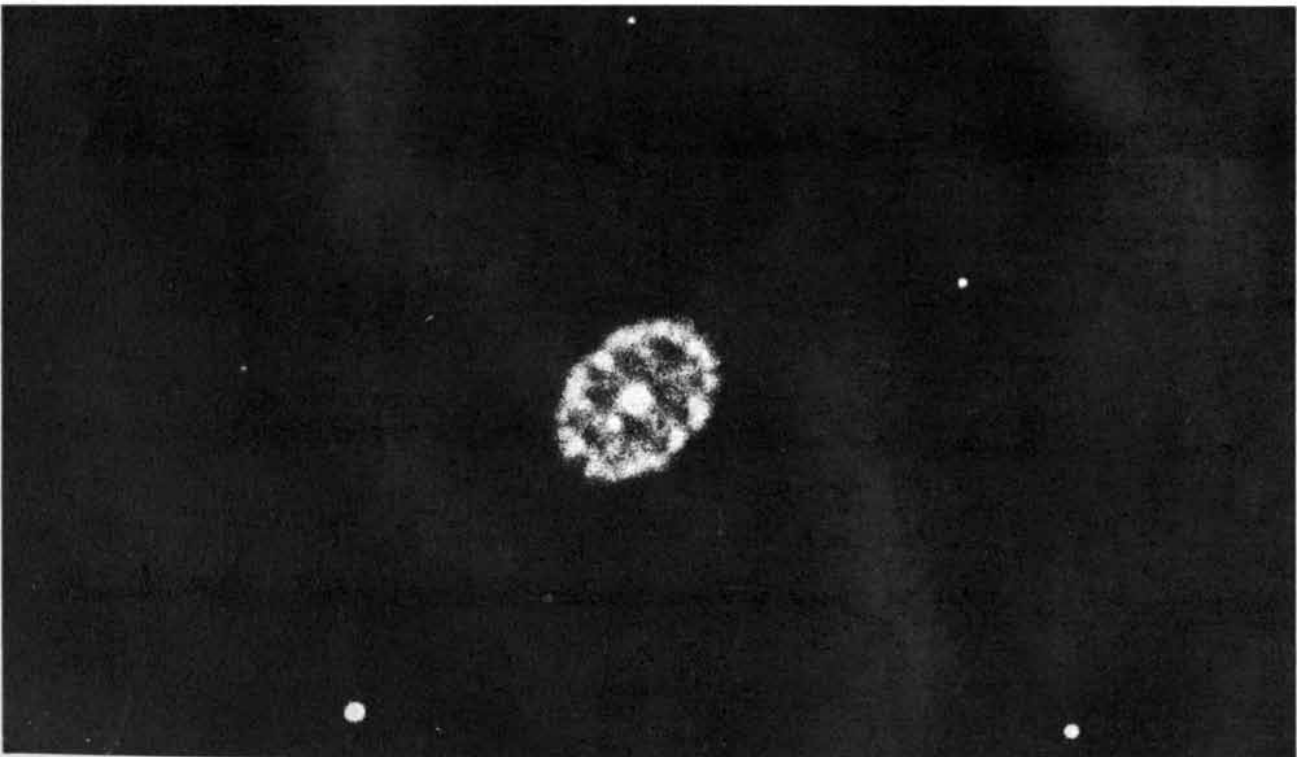
IMPACT OF HIGH-VELOCITY GAS from the accretion disk, seen here edge on, gives rise to a hot boundary layer on the surface of the white dwarf. The highly compressed matter in the white dwarf is in a degenerate state: it is more like a solid than a gas. It therefore does not expand on being heated. The hydrogen from the accretion disk

mixes with heavier elements such as carbon and oxygen, which may be transported to the surface of the white dwarf from the interior by currents generated by the falling gas. At a critical temperature, above 20 million degrees Kelvin, the fusion of hydrogen and carbon initiates the runaway chain reaction that climaxes in the nova outburst.



EXPANDING SHELL OF GAS is visible around the 1901 nova GK Persei. The nova reached its peak rapidly, briefly outshining Procyon, the eighth-brightest star in the sky. The shell is not only the largest ever observed but also unusual in being inhomogeneous and very

hot. Its temperature of 30,000 degrees K. is much higher than the temperature a nova is likely to maintain. Evidently its patchiness and high temperature arise from the collision of the shell, expanding at 3,000 kilometers per second, and the gas in the interstellar medium.



SHELL AROUND DQ HERCULIS was ejected by a slow nova that took some weeks to reach its maximum intensity in 1934, when it rivaled the brightness of Deneb, a star in the nearby constellation Cygnus and the 19th-brightest in the sky. The photograph, recorded in light emitted by hydrogen, was made with the 2.3-meter telescope of the Steward Observatory of the University of Arizona. Unlike the

shell around GK Persei, the shell around DQ Herculis is symmetrical. Although the gas in the shell is ionized, as if it were hot, it is actually quite cold: 500 degrees K. The author's spectral analysis of the shell shows that it has an unusually high abundance of the elements carbon, nitrogen and oxygen. Presumably these elements were present in the outer layers of the white-dwarf star just before the outburst.

Kraft's observations could not reveal a binary system if the plane of the system's orbits was aligned nearly perpendicular to the observer's line of sight, but it seemed reasonable to suppose that all the old novae Kraft studied were members of binary systems and that the distances separating the two stars were generally comparable to the diameter of normal stars. In any case his observations were strong circumstantial evidence that the nova phenomenon is directly related to the existence of very close double-star systems.

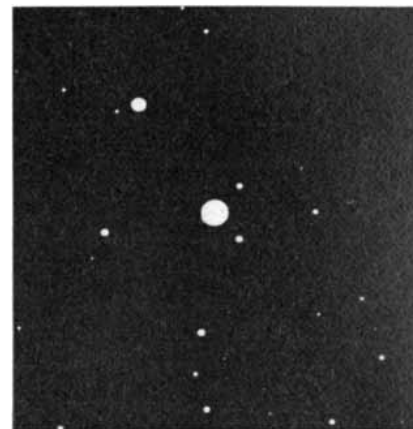
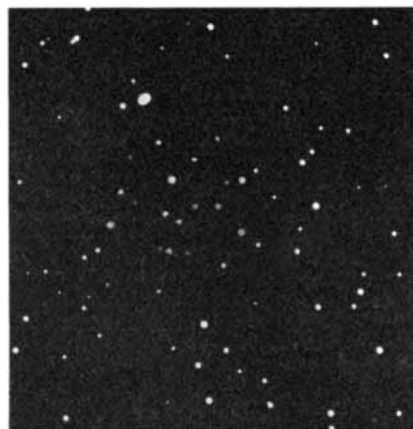
In those instances where observations of old novae revealed an eclipsing system additional information could be deduced about the nature of the stars in it. If the orbital velocities of the stars are known from the periodic Doppler shifts in their spectral lines, the sizes of the stars can be calculated from the duration of the eclipses. Furthermore, the relative brightnesses of the two stars can be found by comparing the intensity of the system when the stars are eclipsing each other with the intensity when they are not eclipsing. This kind of analysis of old nova systems led to another interesting and unexpected result: It is common for one star in each nova binary system to be moderately hot and very small. Indeed, the smaller companions were smaller than any normal star could possibly be. They could only be very dense white dwarfs: stars typically about two-thirds as massive as the sun and roughly the size of the earth.

A white dwarf is a star at the very end of its evolutionary cycle. It has previously converted virtually all its available fuel of hydrogen and helium into carbon and oxygen by nuclear fusion and can no longer produce energy. Lacking an energy source to support itself against the mutual gravitational attraction of its constituent atoms, it contracts to an extremely dense, compact state where matter cannot be further compressed. The very dense matter in such a star, more like a solid than a normal gas, is described as being degenerate. The incompressibility of a degenerate gas prevents the star from contracting further, bringing stellar evolution to an end. Normally a white dwarf cools slowly over billions of years and ultimately fades into darkness.

The finding that most novae appear to be members of close binary systems in which one member is a white dwarf suggested a plausible explanation for nova outbursts. Drawing on the related ideas of Evry Schatzman of the Institute of Astrophysics in Paris and Leon Mestel of the University of Leeds, Kraft proposed that nova outbursts are caused by the transfer of matter from the normal star of a binary pair onto the degenerate white-dwarf companion. Because of its high density the white dwarf has a very



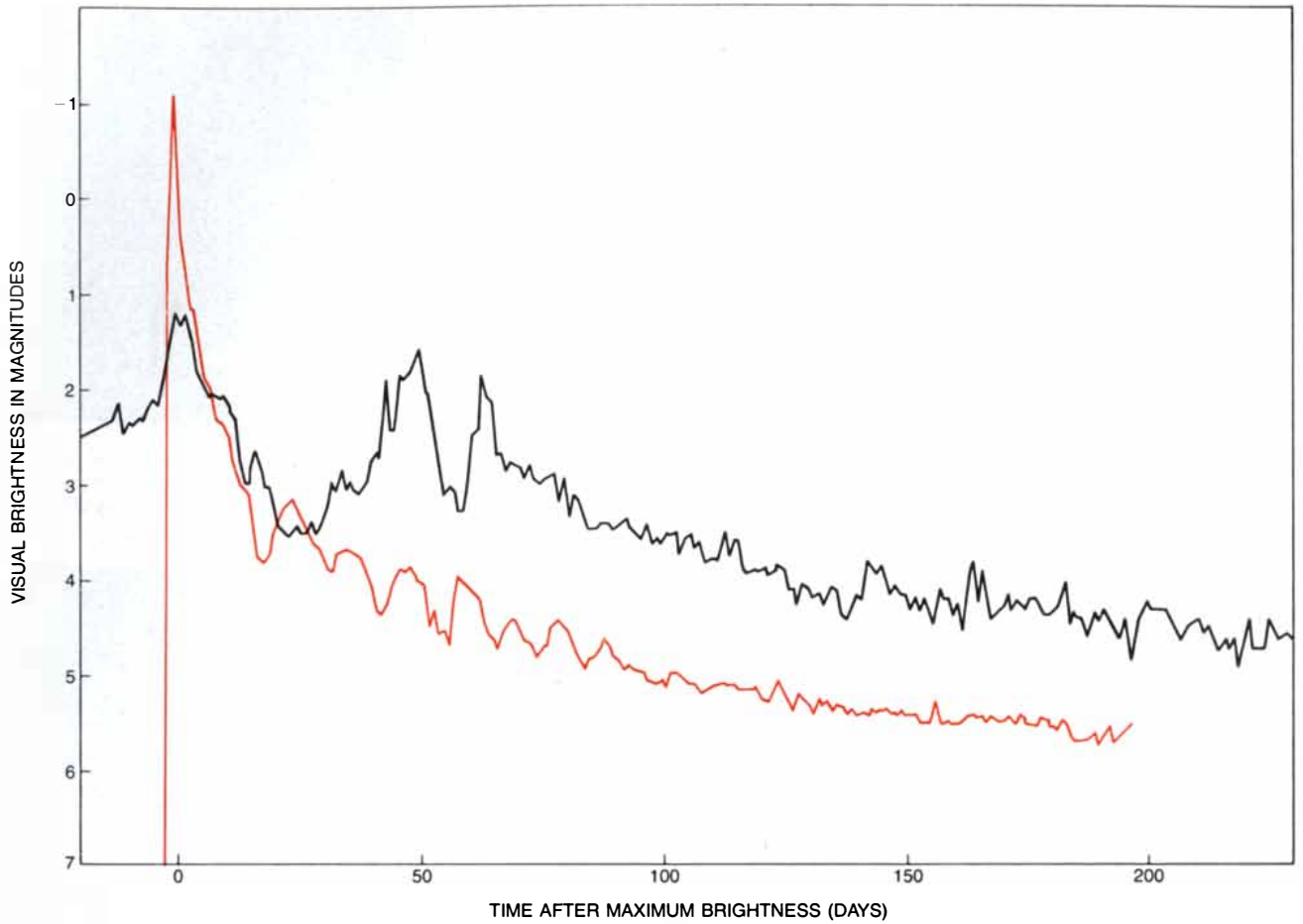
NOVA DQ HERCULIS (*right*) was photographed in 1934 at the Yerkes Observatory of the University of Chicago when its visual magnitude reached 1.4. Before the outburst (*center of picture at left*) its magnitude was 14. Therefore the star increased 100,000 times in brightness.



NOVA AQUILAE OF 1918, the brightest nova to have been observed in the past 100 years, reached a visual magnitude of -1.1 , making it for a few days the second-brightest star in the sky. In the photograph at the left, which was made before the outburst, the future nova is the faint star of magnitude 10.6 in the middle of the picture. The photograph at the right shows Nova Aquilae near its peak brilliance. During its outburst the nova increased in brightness by nearly 12 magnitudes, or a factor of 60,000. Photographs were made at Yerkes Observatory.

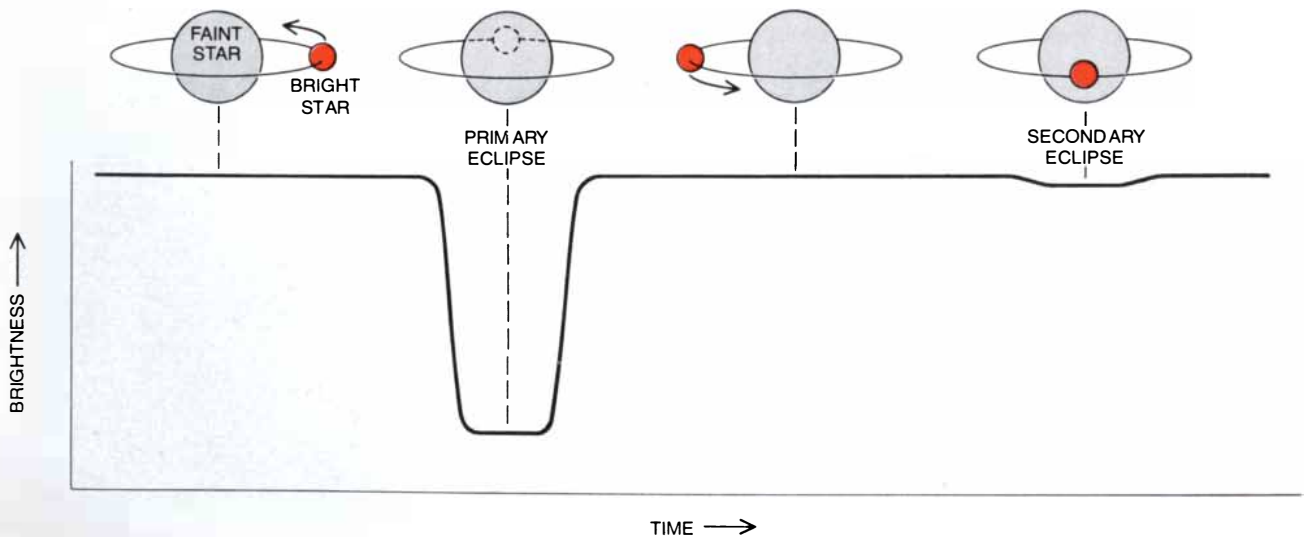


NOVA CYGNI OF 1975, the most recent bright nova, was photographed (*left*) at the Lick Observatory on August 31, 1975, when it had reached its peak magnitude of 1.8. In the photograph at the right, made three months later, the nova had faded to magnitude 11, a drop of a factor of 4,800. The brightness of the star before the outburst is uncertain because the star was too faint to have been recorded in any photograph. By the same token, however, its brightness must have been less than that of stars of magnitude 20. Accordingly its increase in brightness must have amounted to a factor of at least 19 million, a record for known nova outbursts.



DIFFERENCE BETWEEN FAST AND SLOW NOVAS can be seen in the light curves of Nova Aquilae 1918 (*color*) and Nova Pictoris 1925 (*black*). Both novae briefly matched the brilliance of the brightest stars. (A change of five magnitudes represents a 100-fold change in brightness.) The very rapid brightening of Nova Aquilae

was followed by a steep decline during which it exhibited periodic fluctuations. Nova Pictoris, in contrast, took several months to reach its peak brightness, after which it faded more erratically and gradually than Nova Aquilae. Nova Pictoris was discovered by a night watchman in South Africa who noticed it as he was walking home.



PERIODIC DIMMING OF A BRIGHT STAR is evidence that the star is a member of an eclipsing-binary pair. Such evidence has established that nova systems are binaries in which one member is a white-dwarf star. For dimming to be observed the stars must be oriented so that the plane of the system's orbit is seen nearly edge on. The two stars will then be observed to eclipse each other at regular intervals.

The relative brightness of the system during the two eclipses is directly related to the surface brightnesses of the two stars. Furthermore, the length of time it takes for each eclipse to become total, compared with the length of the total eclipse, can be used to determine the relative sizes of the two stars. No nova binary system is close enough to the solar system for the two member stars to be observed separately.

strong gravitational field, and gas accreting onto its surface is accelerated to very high velocities. The material colliding with the surface of the white dwarf is therefore heated to extremely high temperatures. Eventually the temperature becomes high enough to ignite nuclear reactions on the surface of the dwarf, leading to a violent release of energy.

This sudden release of energy is made possible by the fact that the nuclear reactions on the white-dwarf surface proceed in a degenerate gas. In a normal gas the explosive liberation of energy cannot occur because with increasing temperature the gas expands. With expansion the temperature of the gas falls, and the nuclear reaction rates, which are a function of temperature, fall along with it. This thermostatic effect enables all normal stars to radiate energy at a stable level for billions of years.

Degenerate matter, because of its incompressibility, does not expand when it is heated. Therefore when nuclear reactions occur, the material steadily increases in temperature, causing the nuclear reactions to proceed at a higher rate. This heats the gas still further, so that the process goes unchecked, giving rise to a runaway, or unstable, series of nuclear reactions. Energy production rapidly increases to tremendous levels, culminating in an explosive outburst.

The transfer of gas from a normal star onto a nearby degenerate compact companion star is not unique to novae. A similar process gives rise to X-ray stars: binary systems in which the compact object is not a white dwarf but is thought to be either a neutron star or a black hole. The matter in a neutron star is far more condensed (and in a black hole infinitely more condensed) than that in a white dwarf. As a result the gas spiraling into the compact object in an accretion disk accelerates to extremely high velocity. Because of the increased kinetic energy acquired by the gas the radiation associated with the disk is emitted primarily in the X-ray region of the spectrum rather than in the optical region.

Much of the theoretical work on novae in the past decade has been devoted to testing the idea that a nova eruption is a runaway thermonuclear process on the surface of a white dwarf in a close binary system. The leading proponents of the hypothesis have been Sumner G. Starrfield of Arizona State University, James W. Truran, Jr., of the University of Illinois at Urbana-Champaign and Warren Sparks of the Goddard Space Flight Center. Their extensive calculations of the nova phenomenon are in basic agreement with a number of the important characteristics of the outbursts. The picture that has emerged from the work they and others have done can be summarized as follows.

The accretion of gas onto the white

4,000,000 People A Year Visit Scotland. It Probably Isn't For The Sunshine.



We have all kinds of weather in Scotland. But not much of it is very good. So we know that people aren't coming here for a suntan.

A good number do come to play golf, climb castles, or smell heather.

But that still leaves a lot of unexplained visitors.

It's whispered that they come to try our national drink. Scotch.

We suggest you test that theory. And we recommend you try J&B Rare Scotch because of its lightness and warmth. J&B is a very carefully culled selection of the finest Scotch whisky. It's blended for smoothness and subtlety. The result is a taste that we say whispers.

By the way, we don't have a lot of sunshine in Scotland, because we don't need a lot of sunshine.

J&B. It whispers.



86 Proof Blended Scotch Whisky, © 1981 The Paddington Corp., NY

dwarf is continuous, driven by the steady expansion of the outer layers of the companion star as it evolves toward becoming a red giant. The gas, which forms into a disk in the plane of the stars' orbits, consists largely of hydrogen. When it spirals at high velocity onto the surface of the white dwarf, it mixes with the material making up the outer layers of the dwarf: mostly helium, perhaps with an admixture of carbon and oxygen that have originated in the dwarf's interior.

White dwarfs normally lack hydrogen because it has been converted into the heavier elements helium, carbon and oxygen by nuclear reactions earlier in the evolution of the star. The buildup of a fresh supply of hydrogen on the surface of the dwarf is significant because hydrogen, by virtue of its simple nucleus (one proton), is the element that reacts most readily with other nuclei to release energy. It is the addition of fresh hydrogen to the heavier elements already present in the dwarf that makes possible a new series of nuclear reactions. The initiation of the fusion reactions requires a temperature of about 20 million degrees Kelvin. At that temperature individual nuclei are accelerated to velocities sufficiently high to overcome the electrostatic repulsion exerted by their positive charges, enabling the nuclei to collide, fuse and release energy.

The necessary temperature is gener-

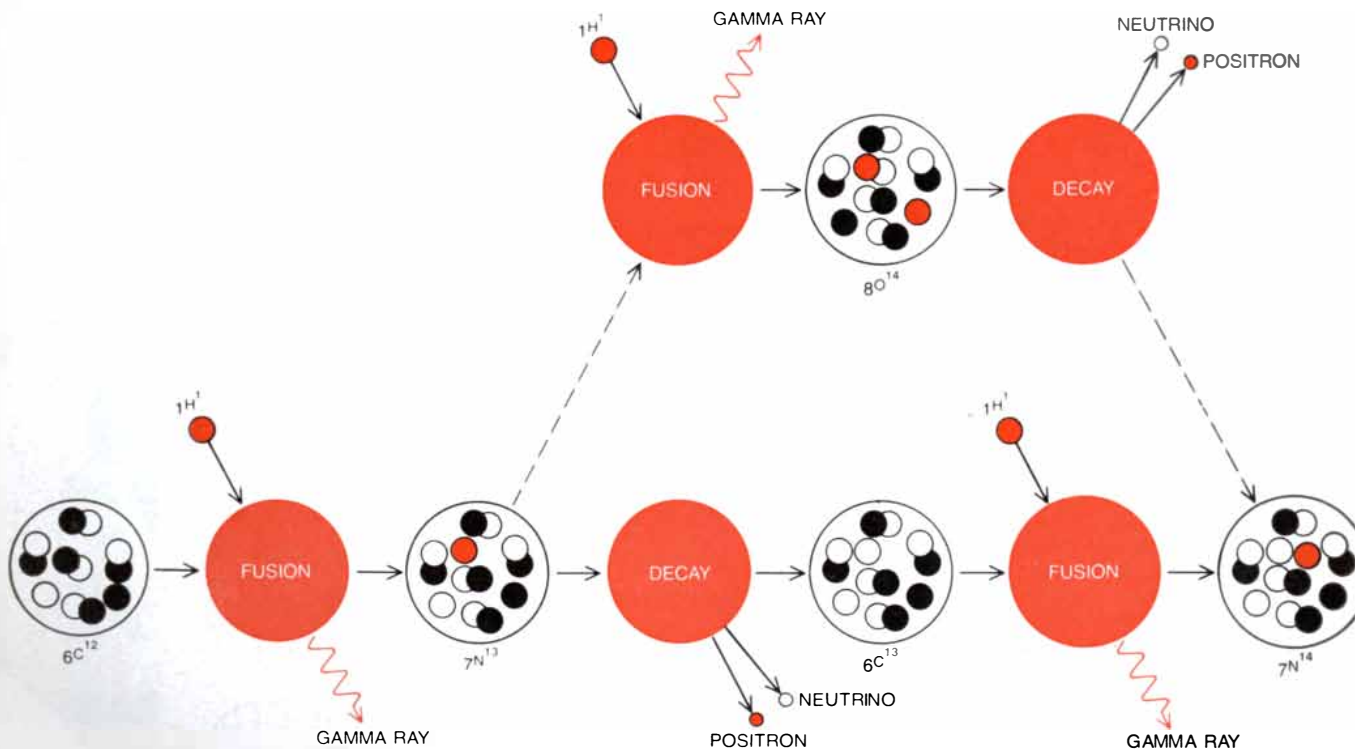
ated by the impact of gas spiraling at high velocity onto the surface of the white dwarf. The time required for the temperature to reach the level necessary for fusion depends on several factors, including the rate at which the expanding companion star feeds gas into the accretion disk, but it is typically on the order of tens of thousands of years. Eventually a temperature is reached at which the degenerate material on the surface of the white dwarf, seeded with fresh hydrogen, begins to support fusion reactions.

In the fusion reactions protons are captured by the nuclei of the heavier elements. The newly fused nuclei subsequently decay at rates fixed by their characteristic radioactive half-lives. In the reactions postulated for the surface of white dwarfs the decay usually results in the release of a positive electron (positron) and a neutrino. The chain of reactions is exactly the same as the one that converts hydrogen into helium in the core of normal stars more massive than the sun. The sequence of collisions, which results in the net conversion of four hydrogen nuclei into one helium nucleus, involves the nuclei of the elements carbon, nitrogen and oxygen as catalysts and is therefore known as the CNO cycle [see illustration below].

The calculations of Starrfield, Truran and Sparks have demonstrated that the

thermonuclear reactions of the CNO cycle will start on the surface of a white dwarf that accretes enough hydrogen over a period of time. The calculations show that given the conditions expected to prevail in close binary systems, the onset of nuclear reactions does lead to a fast thermonuclear runaway in the degenerate surface material. Certain characteristics of these theoretical models are in good agreement with the observed features of novae, such as the shape of the light curves and the total energy generated. The calculations also indicate that one of the most important variables in determining the characteristics of the nova outburst is the abundance of carbon, nitrogen and oxygen with respect to hydrogen. A degenerate material rich in these heavier elements generates energy much faster and in greater amounts than a material significantly less rich. Indeed, the models suggest that one of the major differences between fast and slow novae is the chemical composition of the material on the surface of the white dwarf.

According to the theoretical studies, if carbon, nitrogen and oxygen are about 100 times more abundant with respect to hydrogen than they are in normal sunlike stars (that is, if the number of carbon, nitrogen and oxygen atoms together equals 1 percent of the number of hydrogen atoms), the nuclear reactions in the degenerate gas proceed so fast that the surface temperature quick-



CHAIN OF NUCLEAR REACTIONS thought to be responsible for nova outbursts converts hydrogen into helium by the same process that takes place in the core of normal stars. Four hydrogen nuclei, or protons (${}^1\text{H}^1$), fuse successively with nuclei of carbon, nitrogen and oxygen in the sequence of reactions known as the CNO cycle. The

first three fusion reactions in the cycle are accompanied by the release of energetic gamma rays. The last fusion reaction, in which a proton reacts with a nitrogen-15 nucleus, is followed by the ejection of a helium nucleus, or alpha particle (${}^2\text{He}^4$), and the regeneration of a nucleus of carbon 12 to start the cycle over again. Some of the

ly reaches 100 million degrees K., and within a period of minutes the energy is released explosively. The outburst, in the form of both radiation and particles, encompasses the surface of the white dwarf. Within hours a substantial fraction of the accreted outer layer of the dwarf is ejected into space, and the star brightens rapidly much as a fast nova does. Alternatively the models show that if the degenerate surface layers of the white dwarf have concentrations of carbon, nitrogen and oxygen no higher than those in normal stars, the thermonuclear runaway is spread over a longer time span and the nova is a slow one.

The prediction that differences in chemical composition may be largely responsible for the different types of nova outbursts offers an opportunity for checking the validity of the models and the basic picture of the outburst. Perhaps the most straightforward method of establishing the composition of the gas on the surface of the white dwarf is to study the shells of gas that are ejected in an outburst. Once the shells are ejected they are not disturbed by any later activity of the nova system; they should therefore reflect the conditions that led to their formation.

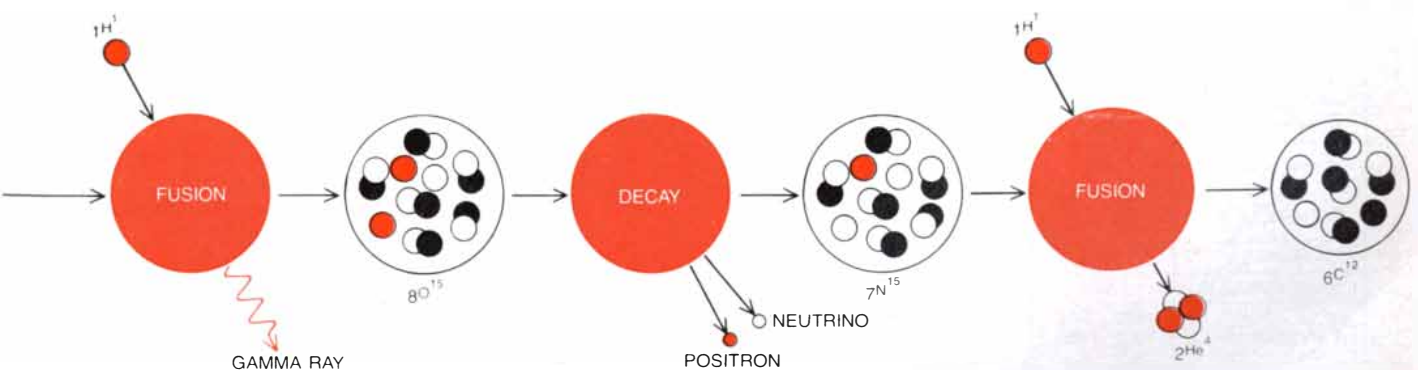
The spectra of novae during an outburst always show evidence that material has been rapidly ejected from the nova system. The ejected shells consist

of an appreciable fraction of the material that was accreted onto the white dwarf from its larger companion, amounting to about .01 percent of the total mass of the dwarf. The ejected material, blasted into space at velocities of some 1,000 kilometers per second, forms an expanding envelope around the binary system. Such shells rapidly dissipate as they expand, generally fading from sight before they become large enough to be observed. The shells blown off by some of the novae closer to us have, however, been detected.

At the Steward Observatory of the University of Arizona my colleagues and I have undertaken a program to search for the shells of old novae and to examine them. The models have suggested that the ejecta of fast novae should be rich in carbon, nitrogen and oxygen and the shells around slow novae should not be. The composition of a shell and the physical conditions in it can be determined by examining its spectrum. A rarefied gas such as a nova shell radiates a bright-line, or emission, spectrum. The gas radiates only at certain definite wavelengths that are determined by the composition, density and temperature of the gas. By measuring the relative intensities of the emission lines at various wavelengths one can obtain information about the abundances of the elements and the temperature of the shell.

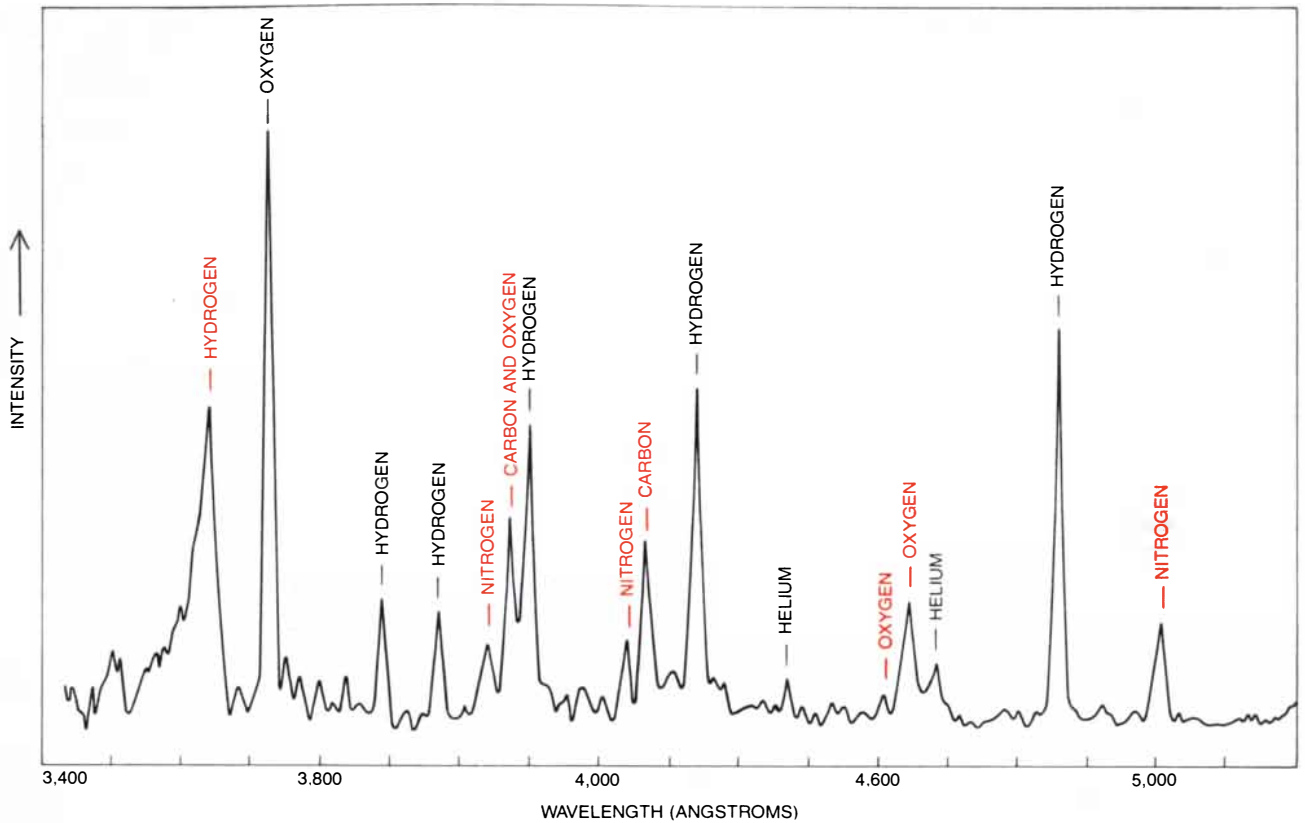
The first shell we analyzed spectroscopically was the symmetrical envelope around DQ Herculis, the slow nova that appeared in 1934. We found that the spectrum of the shell, obtained with the 2.3-meter telescope of the Steward Observatory, was unlike any other known up to that time. As a rule the spectra of radiating gas clouds in space, such as the Great Nebula in Orion, are all quite similar to one another. Nebulae shine by virtue of the energy they receive from hot stars with which they are associated. Hot stars radiate primarily in the ultraviolet region of the spectrum, and when this energetic radiation shines on gas around the star, it ionizes, or strips electrons off, the atoms in the gas. The gas is thereby heated to some 10,000 degrees K., and collisions between ions and the liberated electrons give rise to the radiation of the nebula's characteristic emission spectrum.

The spectrum of the DQ Herculis shell was quite peculiar: many of the stronger lines in it had no counterpart in the emission spectra of nebulae. The mystery of the origin of the emission from the shell was solved only when we realized that one of the features in the spectrum originating with hydrogen required that the gas be far cooler than the gas in emission nebulae: some 500 degrees K. This explained why many of the familiar nebular spectral lines did not appear in the shell spectrum. When the



proton-capture reactions are followed by decay reactions that also yield energy in the form of an energetic positron (positive electron) and a neutrino. (The emission of a positron transforms a nuclear proton into a neutron.) The fusion-reaction rate depends sensitively on the temperature of the gas. Once the reactions begin in the degenerate

gas on the surface of the white dwarf they raise the temperature of the gas, thereby causing the reactions to go faster. The eventual result is the thermonuclear runaway that gives rise to the outburst. Computer models suggest that whether the outburst is slow or fast depends on the concentration of carbon, nitrogen and oxygen in degenerate gas.



SPECTRUM OF SHELL OF DQ HERCULIS, recently obtained by the author and his colleagues with the 2.3-meter telescope of the Steward Observatory, is unlike that of any previously known nebula. The spectrum of the shell of this slow nova consists of faint continuous radiation on which are superposed lines emitted by elements in various states of excitation. The radiation from elements labeled in black is usually prominent in the spectra of normal nebulas,

whereas the emission lines labeled in color are generally extremely weak or nonexistent. The strength of the emission from carbon, nitrogen and oxygen shows that the shell of DQ Herculis is unusually rich in these elements. The strong hydrogen emission at 3,645 angstrom units requires that the gas be very cold, about 500 degrees K. The oxygen emission at 3,727 angstroms, on the other hand, is an anomaly; it is usually associated with gas whose temperature is 10,000 degrees K.

emitting atoms were finally identified, they consisted of ions of hydrogen, helium, carbon, nitrogen and oxygen. This was again puzzling, because high energies are needed for the ionization of atoms, and high energy is usually associated with high temperature.

The paradoxical ionization of a cold gas can be explained if the nova shell is not in equilibrium but is still recovering from the effects of the outburst. It is known from observations near the time of a nova outburst that the ejected shell is highly ionized and hot, with tempera-

tures exceeding 15,000 degrees K. With the passage of time the surface of the white dwarf cools off, no longer emitting the intense ultraviolet radiation that originally heated and ionized the shell. Sometimes alternative sources of energy are available to the expanding shell. If

NOVA	YEAR OF OUTBURST	MAXIMUM MAGNITUDE	SPEED CLASS	ANGULAR SIZE OF SHELL (ARC SECONDS)	SHELL TEMPERATURE (DEGREES KELVIN)	ENRICHED ELEMENTS IN SHELL
T AURIGAE	1891	4.2	SLOW	20	≤3,000	HELIUM, NITROGEN AND OXYGEN
GK PERSEI	1901	0.2	FAST	75	>25,000	NITROGEN
V476 CYGNI	1920	2.0	FAST	10		
RR PICTORIS	1925	1.2	SLOW	25	15,000	HELIUM AND NITROGEN. POSSIBLY NEON
DQ HERCULIS	1934	1.4	SLOW	20	500	CARBON, NITROGEN AND OXYGEN
CP PUPPIS	1942	0.2	FAST	15	≤1,000	NITROGEN
V533 HERCULIS	1963	3.0	FAST	5		
T PYXIDIS	1890 1944 1902 1966 1920	6.6	SLOW	10	~10,000	NONE

EIGHT NOVAS WITH OBSERVABLE SHELLS are under investigation by the author and by others. Although there is evidence that all novae eject gas during their outburst, relatively few produce shells that are detectable. Most of the gaseous ejecta are too faint to show up in photographs because the shells tend to fade away before they

reach an observable size. Shells form around both fast and slow novae and exhibit a wide temperature range. With the exception of the recurrent nova T Pyxidis, the novae appear to create shells with one characteristic in common: the shells are enriched in one or more of the heavier elements compared with the composition of normal stars.



Renault Presents The Remarkable 18i Sportswagon.

Renault, maker of Europe's best selling cars, presents a remarkable new wagon for America...the Renault 18i Sportswagon.

Remarkable, because it is, at once, a wagon that provides comfortable seating for 5, *and* muscle to move you crisply from 0 to 50...

...a wagon with up to 65.5 cubic feet of load space, *and* the good manners to go where you aim it through precise, quick rack and pinion steering (lock to lock in just 3 turns)...

...a wagon that takes care of the whole family, *and* carefully meters out fuel for outstanding efficiency*...

...a wagon that bristles with innovative technology of the 80's.

38 HWY EST* **25** MPG EST

Bosch L-Jetronic Fuel Injection

This is the Electric Multi-Point Bosch system also used by Porsche 928 and Jaguar XJ-S. It precisely measures out the fuel required for optimum performance and efficiency from the 18i's enthusiastic 1.6 liter (1647 cc, aluminum block and head) 4-cylinder engine.

20 Years Experience in Front-Wheel Drive Design.

The 18i front-wheel drive system is a study in compact, lightweight design. Front drive transaxles, engine and 4-speed manual transmission (5-speed, avail.) are deftly mounted to provide a 60/40 weight ratio, front to rear. It also allows quick access for service at any of the more than 1300 Renault and American Motors Dealers.



Slip Stream Aerodynamics

The 18i Sportswagon silhouette is notably slippery. A wide front modesty panel—more discreet than add-on air dams—encourages turbulence to slide beneath the 18i. Slip stream styling effects promote excellent fuel efficiency* as the wagon carves the air.

Road Adhesion

The 18i Sportswagon is masterfully tuned for the road. With 155SR x 13 Michelin steel belted radials, box section "live" rear axle (it flexes in the turns), front and rear sway bars, and beefy helical coil springs (variable-flex in the rear) surrounding long-travel shock struts.

Inner Space Geometry

The 18i's elegantly tailored seats are bio-formed with special support for the lumbar region and upper thighs.

Controls and gauges are strategically angled and positioned for driver access.


A wide rear bench seat folds forward, providing a lushly carpeted flat load bed more than 5½ feet long.

The Renault 18i Sportswagon. Remarkable? We think so. It provides what wagon users ask for...generous space and great mileage. And something many have been missing...the sheer joy of commanding a responsive, nimble, sensitive, exciting road machine.

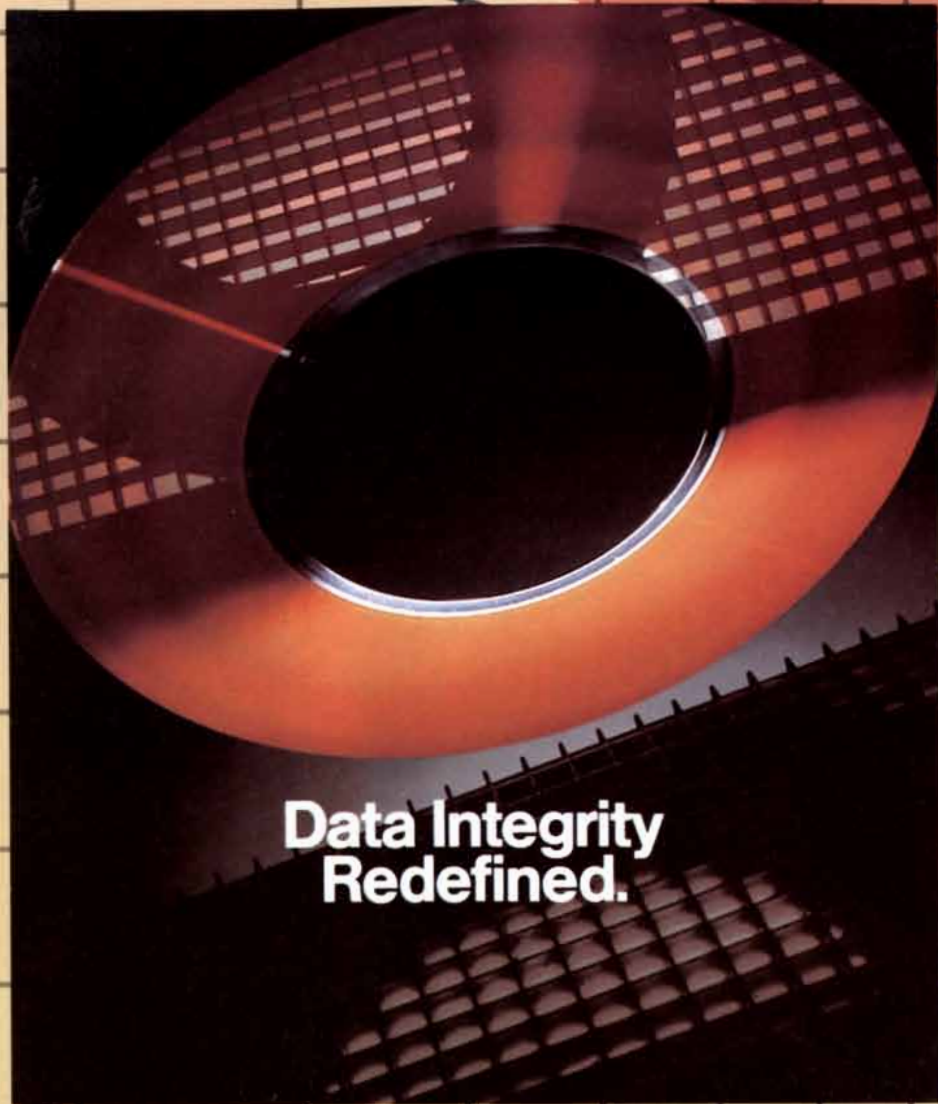
*EPA estimated at 25 mpg, 38 mpg highway est. Remember. Compare this estimate with estimated mpg for other cars. Your mileage may differ depending on speed, trip length and weather. Your highway mileage will probably be lower.

Renault 18i

More than just economy

At  Renault and

 American Motors dealers.



SIGNAL TO NOISE (DB)

OPTIMUM RANGE

MEMOREX

LEADING COMPETITION

Data Integrity Redefined.

FLUX CHANGES PER INCH (thousands)

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

The new Memorex® M Formula magnetic coating is redefining performance standards for all new generation rigid disc drives. Containing our exclusive Hi/Oxide™ formulation, M Formula offers the highest signal to noise ratio—>30 dB at 10,000 flux changes per inch—of any rigid media in the industry. With higher performance margins and protection against random errors, you're assured greater data integrity. And that means more flexibility for the advanced drive designer, too.

The unique M Formula coating has created a new benchmark for surface smoothness. CLA has been reduced to $< .5\mu$," and the deviation has been reduced to within 5μ ." This represents a 100% improvement over existing industry averages. The smoother surface significantly reduces noise and improves head flying uniformity.

Our revolutionary Hi/Oxide™ coating forms a hard surface that resists wear and has set new standards for durability. M Formula's extraordinary

thermoset properties also allow it to withstand a wide range of temperature changes.

It all adds up to trouble-free data handling. But most importantly, the new M Formula ensures far greater data integrity than you'll find in any other rigid media available.

To learn more about how Memorex is redefining data integrity, contact Memorex Rigid Media and Components Division, 1500 Memorex Dr., Santa Clara, CA 95052 or call N. "Vis" Viswanathan at (408) 987-1928.

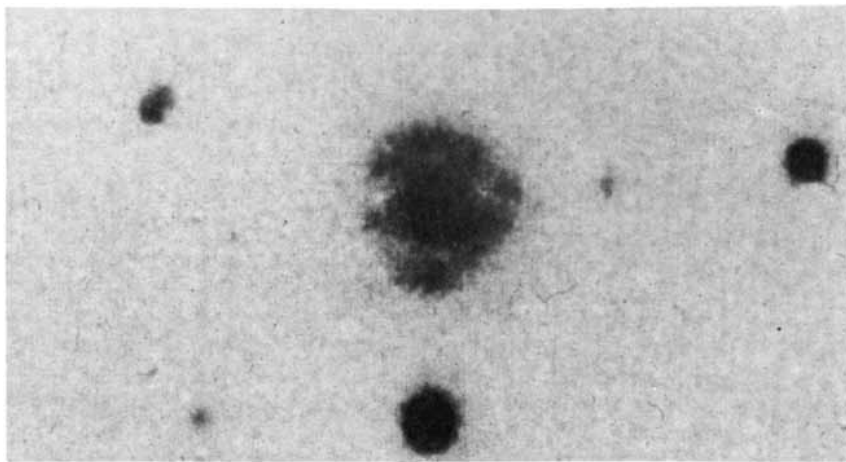
MEMOREX

the shell is expanding in a region where the tenuous gas of interstellar space is relatively dense, the collision between it and the interstellar gas can be a highly energetic one. Such a collision may well explain the unusual appearance of the shell of GK Persei, a brilliant nova of 1901, which left one of the largest and hottest shells known. The shell of DQ Herculis of 1934, on the other hand, is evidently expanding without interference from interstellar gas; it lies off the central plane of our galaxy, where most of the interstellar gas is. Free to expand without any additional input of energy, the shell of DQ Herculis has been cooling steadily. The electrons that were stripped from the atoms are now being recaptured by the ions. The time required for the electrons to be completely recaptured, however, is substantially longer than the time it takes for the gas temperature to drop. The result is that the shell has now cooled to far below its original temperature, and yet most of the electrons have not been recaptured; hence the seeming paradox of a cool gas that is still highly ionized.

When we completed our analysis of the shell spectrum, we found that the combined abundances of carbon, nitrogen and oxygen with respect to hydrogen are a factor of 100 higher than they are in normal stars. According to the conventional nova models, such abundances would be expected to correspond to those of a fast nova. Since DQ Herculis was a slow nova, however, the spectral analysis and the models are in apparent conflict.

Care must be taken in generalizing the results from a single object to other novae, and therefore John S. Gallagher of the University of Illinois at Urbana-Champaign and I are now examining other nova systems with extended shells. We have completed work on two additional objects, the envelopes that surround the slow novae RR Pictoris, which flared up in 1925, and T Aurigae, which did the same in 1891. The spectra show that the shell of RR Pictoris is substantially enriched in helium and nitrogen; the shell of T Aurigae is enriched in helium, nitrogen and oxygen. Gary J. Ferland and Gregory A. Shields of the University of Texas at Austin have conducted a similar study of V1500 Cygni, the fast nova of 1975, and have found that its shell, still too small to be resolved in photographs, is much richer than normal stars in many of the heavier elements, including carbon, nitrogen and oxygen.

Several other nova shells, some in the sky of the Southern Hemisphere, are accessible for analysis, and their composition is being studied. On the basis of the results available so far it seems that most novae, fast and slow, show enrichments in various of the heavier elements. The source of the enrichments is



RECURRENT NOVA T PYXIDIS has undergone five outbursts since 1890, the most recent one in 1966. Recurrent novae brighten by a factor of only about 1,000 during an outburst, an increase considerably smaller than that of a normal nova. It has been discovered only recently that recurrent novae also throw off observable shells (as is demonstrated in this negative print of a picture of T Pyxidis, obtained electronically with an image tube on the four-meter telescope of Kitt Peak National Observatory). The shell was discovered in 1978 by Harvey R. Butcher of Kitt Peak and D. A. Kopriva and the author in a search for shells around novae.

not known, but it is quite possibly the interior of white dwarfs, which are known to be rich in carbon and oxygen.

The presence of the high abundances of heavy elements in fast novae is not at all surprising. The discovery that the shells around the slow novae DQ Herculis, RR Pictoris and T Aurigae are enriched in heavy elements, however, is at variance with the theoretical calculations. In particular the shell of DQ Herculis contains amounts of carbon, nitrogen and oxygen that, if they had been present before the outburst, should have generated enough energy to make DQ Herculis a fast nova.

The difficulty in reconciling the observed enrichments in the shells of slow novae with the predictions of the CNO runaway models has prompted some theorists to speculate that nova eruptions may not always involve the CNO cycle of reactions. The enrichments could be explained if nova outbursts were generated by nuclear reactions capable of synthesizing carbon, nitrogen, oxygen and other elements directly out of hydrogen or helium, which the CNO cycle does not do. There are, for example, thermonuclear reactions in which three helium nuclei (alpha particles) fuse to yield carbon 12, from which other heavier elements can be formed by the subsequent capture of protons. This "triple alpha" process is known to be an important source of energy in red-giant stars. There is probably ample helium on the degenerate surface of white dwarfs for this reaction to occur, and so it seems possible that some nova flare-ups could be due to the triple-alpha process rather than to the CNO cycle based on hydrogen.

Additional evidence for the fusion of

helium in some nova outbursts may be indicated by the abundance of helium in accretion disks before the outburst compared with its abundance in nova shells after the outburst. Recent studies of old novae at the Steward Observatory have demonstrated that the amount of helium in the ejected shells is less than that commonly found in the gas accreting onto the white dwarfs. A logical explanation for this apparent decrease in the helium content of the gas is that triple-alpha reactions have occurred during the outburst and have converted helium into carbon and other elements.

In spite of uncertainties in some of the details, there is now general agreement that nova eruptions are probably caused by thermonuclear reactions on the surface of white-dwarf stars in close binary systems. The process is initiated by the transfer of material from an expanding companion star. The infalling material strikes the degenerate surface of the white dwarf at such high velocity that the surface is heated to the 20-million-degree temperature needed to trigger runaway nuclear reactions. Nova models based on the reactions of the CNO cycle have successfully accounted for many of the observed characteristics of novae. Recently, however, new data on the composition of old nova shells are requiring modification of some of the former ideas about the nature of the outbursts. Further study of old nova shells should prove stimulating not only because of the information it must inevitably yield about novae but also because the shells have already introduced astronomers to unusual environments quite unlike any heretofore encountered in our galaxy.

Filter-feeding Insects

Insects of three orders hatch underwater and gather food with nets, brushes and other fine-mesh filters. They play a role in opposing the tendency of ecological systems to lose organic matter downhill

by Richard W. Merritt and J. Bruce Wallace

Many insects, notably the moth larva we call a silkworm, spin threads to build a cocoon. Less familiar are the insects that spin threads to trap food. They not only weave fine-mesh nets in order to collect the organic matter they ingest but also conduct the enterprise entirely underwater. The net spinners belong to the group of filter-feeding insects that hatch in streams, lakes and other aquatic environments and pass their immature lives entirely submerged. Ten of the 27 orders of insects have aquatic representatives, but only three—the true flies (the Order Diptera), the caddisflies (the Order Trichoptera) and the mayflies (the Order Ephemeroptera)—include species that are known to be filter feeders.

The true flies and the caddisflies are endopterygotes: they are wingless until the pupal stage, when the wing structures develop as the larva assumes its adult form. The mayflies are exopterygotes: wing structures are present from the time the organism hatches until it reaches adulthood, the immature insect's appearance in general presages its adult morphology and it does not pupate at all. Endopterygote larvae molt several times during their approach to the pupal stage. Exopterygote young, generally called nymphs, also increase in size through a series of molts; the adult insect emerges after the final immature stage.

Whatever their life cycle, various species of true flies, caddisflies and mayflies occupy aquatic habitats as diverse as swift alpine streams, meandering rivers, quiet lake bottoms and tidal estuaries, and their habitats frequently overlap. Whatever their order, they are generally divided into two groups: species that live where active water currents allow a passive mode of food collection and species that live where currents are minimal and the insect itself must make the water move.

What nourishes a filter-feeding insect? Numerous analyses of their gut contents show that they frequently do not distin-

guish between inorganic and organic material in the water and therefore ingest particles of silt and small grains of sand along with such plant foods as bacteria and algae and such animal foods as protozoans and small invertebrates. By far the largest part of their ration, however, consists of organic particles, often of unidentifiable origin, known collectively as fine detritus. Among the sources of the organic particles are, first, the feces of scavenging aquatic insects, the "shredders" that feed on decaying vegetation; second, the feces of other aquatic animals that prey on smaller animals or eat living plant tissue; third, organic matter transported from land to water by runoff and, fourth, aggregations of organic matter that has come out of solution. Each detritus particle may also support a frosting, so to speak, that consists of the flora of decay: bacteria, fungi and other microorganisms.

The fine detritus, usually the most abundant food available to filter feeders, is the least rewarding in terms of assimilation efficiency, that is, the percentage of ingested food the feeder absorbs. The assimilation-efficiency rating for the detritus ranges between 2 and 20 percent, compared with 30 percent for algae and better than 70 percent for animal tissue.

This account of the feeding strategies of filter-feeding insects will begin with representatives of those filter feeders whose habitats are in fast currents. One of the least elaborate filtering mechanisms is that of the nymph of the mayfly genus *Isonychia*. The insect's forelegs have a dense fringe of setae, long bristlelike structures. Each bristle bears two rows of fine hairs. The hairs of one row are moderately long and the hairs of the other short and hooked. When the hooked hairs of one bristle are engaged with the long hairs of the next, the filter formed by the interlocked hairs can trap particles with a diameter smaller than one micrometer (one thousandth of a millimeter). To gather its food the

mayfly nymph attaches itself to some convenient surface, faces into the current and raises its forelegs with their arrays of setae interlocked. After a time the insect brings its forelegs within reach of its mouth parts, sweeps off the captured particles and ingests them.

The larvae of two caddisfly genera, *Brachycentrus* and *Oligoplectrum*, have a system almost as simple. They build portable oblong shelters, working with inorganic materials and plant debris held together and anchored with their silklike secretions. The open end of the shelter faces upstream. The larva takes refuge inside the shelter, and when it is filtering, it extends all six of its legs outside the opening to form a fanlike array. Its hind legs and middle legs bear rows of bristles. As food particles are captured by these four filters, the insect works with its forelegs to comb the bristles clean and form the collected particles into a pellet suitable for ingestion. *Brachycentrus* larvae do not rely exclusively on filter feeding; they also graze on microscopic plants.

The larvae of the black-fly family Simuliidae have evolved a feeding system that is both structurally and behaviorally well adapted to a fast-current habitat. They are legless but have a circlet of hooks at the end of their abdomen. With the aid of silklike secretions from their salivary glands they attach the hooks to rocks or submerged plants and then twist their body so that the lower surface of their head, with its mouth parts, faces into the current. This preferential positioning constitutes the behavioral aspect of the insects' adaptation to their habitat.

One structural aspect of the black-fly larva's adaptation consists of a pair of unusual mouth parts called head fans: retractable organs between the antennae and the mandibles. When the larva has assumed its position with respect to the current, the head fans are extended to form a filtering apparatus considerably larger in area than the head itself. The head fans are then re-

tracted one after the other, and the larva cleans off and ingests the captured particles with its mandibles.

That is not the black-fly larva's only structural adaptation. Analysis of the contents of the larvae's gut shows that they capture algae, detritus particles, their associated bacteria and bits of sand and silt in a range of sizes from as large as 350 micrometers to as small as .01 micrometer, which is considerably smaller than the mesh of the head-fan filter. A possible mechanism by which such a filter could capture the smaller particles has recently been pointed out by Douglas H. Ross of the University of Georgia and Douglas A. M. Craig of the University of Alberta. It appears that the black-fly larvae secrete mucus from glands at the front of their head and that the movement of their mandibles spreads this sticky material over the surface of the head fans. When the small particles strike the mucus in passing through the fan, they adhere to it. A number of filter-feeding marine inverte-

brates are known to secrete mucus, but until now such secretions were unknown among aquatic insects.

This brings us to the filter feeders mentioned at the beginning of this article: those that spin nets. A good place to start is with the caddisfly family Hydropsychidae. The larvae of certain members of this group first build a shelter of organic and inorganic debris that they bind together with their own silk. The open end of the shelter faces upstream. Next the larva builds a hooplike oval frame that will support a food-catching net at the open end of the shelter. Then the net itself is spun, starting near the base of the frame, with the larva swinging its head in a series of motions following the path of a figure eight.

The first strand of sticky silk is drawn from one side of the frame diagonally to the base. The second strand is drawn in the same manner but from the other side. The alternation continues, with the strands on each side running parallel to

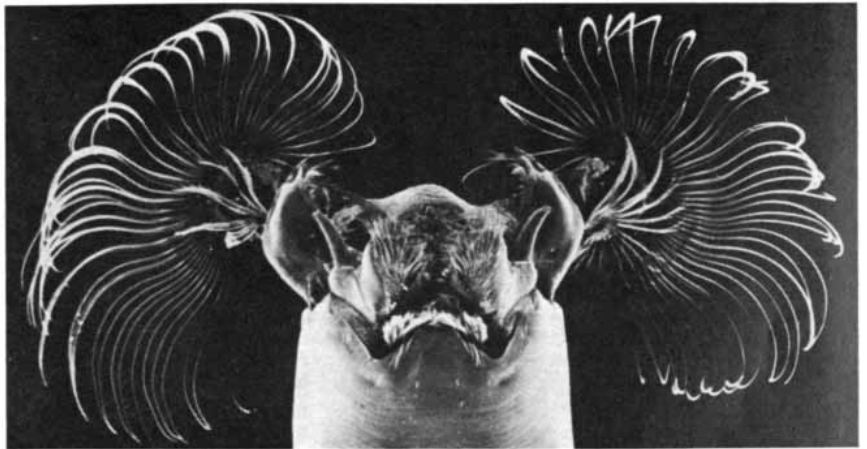
one another. The end result is a net with a rectangular mesh and a central seam that divides one half from the other. The German entomologist Werner Sattler, who had studied the caddisfly genus *Hydropsyche*, observed that the larva needs from seven to eight minutes to weave its net. If the net is torn, the larva will patch it in a random manner; if the net is badly damaged, the larva will weave a new one.

The caddisfly larva molts several times before it reaches the adult stage. The catch nets constructed after each molt are successively larger and coarser in mesh and their threads are heavier. Although the nets with coarser meshes are less efficient in capturing small particles, they not only are larger but also are often set up in places where the current is faster. Then they filter a greater volume of water than a fine-mesh net in a slower current does.

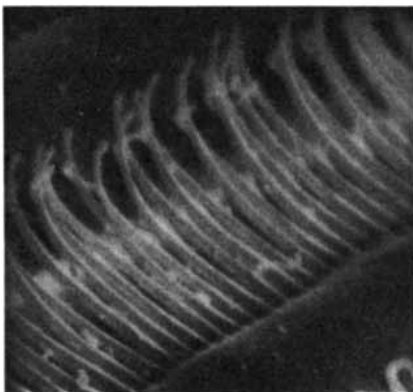
Theodore J. Georgian, Jr., of the University of Georgia and one of us (Wallace) have developed a model of particle



TWIN FANS on the head of a larval black fly can be seen in these scanning electron micrographs. At the left they are in their retracted position between the antennae and the mandibles. At the right the fans are extended with their rays spread so that they filter food parti-



cles out of the water. The particles adhere to a covering of mucus on the rays, and the larva feeds by bringing the rays to its mouth and removing the particles with its mandibles. Both of the micrographs were made by Douglas A. M. Craig of the University of Alberta.



FANLIKE ARRAY of bristling setae on the foreleg of a nymph of the mayfly genus *Isonychia* appears in the scanning electron micrograph at the right. Visible at a greater enlargement at the left are the



tiny hairs on each seta; half of them are short and hooked and the other half are longer and curved. The hooked hairs on one seta engage the long hairs on the next to form a filter that can trap fine food particles.

capture by net-building caddisflies in a stream in southern Appalachia. The model suggests that the larvae of all caddisfly species, regardless of their stage of maturity and the mesh size of their nets, manage to filter out more food than they need. Most of their catch, however, is detritus, a low-quality foodstuff. Analysis of the gut contents of various caddisfly species indicates that those with large-mesh nets feed primarily on relatively scarce and relatively large particles of animal tissue, whereas those with fine-mesh nets feed primarily on detritus of smaller particle size, which is far more plentiful. Hence the advantage of a large volume of water passing through a coarse net, compared with a small volume passing through a fine net, is that the coarse net may be more selective for

animal foods. The differences observed in mesh size in the nets of caddisfly larvae of different species and at different stages of growth appear to be related more to the selection of different kinds of food than to the selection of particles of any given size.

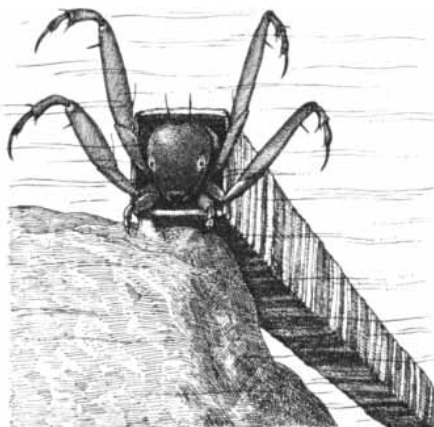
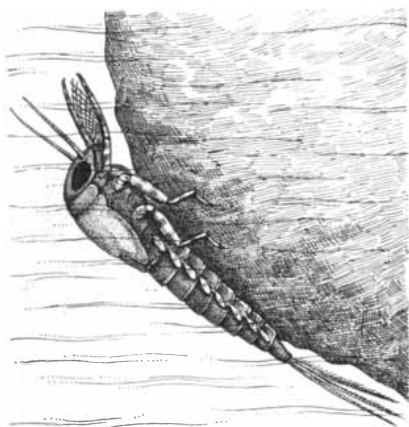
Having captured food particles, large or small, in nets of coarse or fine mesh, how do caddisfly larvae ingest them? Like the filter feeders that do not spin nets, the net spinners have developed diverse behavioral and structural adaptations. For example, one of the coarsest meshes woven by any member of the hydropsychid family is that of the genus *Arctopsyche*, which preferentially inhabits waters with fast currents. The larvae of this genus often capture living prey in their nets, and their spiny forelegs are

useful for that purpose. The larvae of the genus *Macronema* inhabit quieter waters and spin nets with a very fine mesh through all their successive stages of growth. Their forelegs and their mouth parts are equipped with dense arrays of bristles; with these "brushes" the larvae collect and ingest the food particles that accumulate in the net. The larvae of two other caddisfly genera, *Phylocentropus* and *Protodipseudopsis*, also collect food particles from their nets with brush-bearing forelegs.

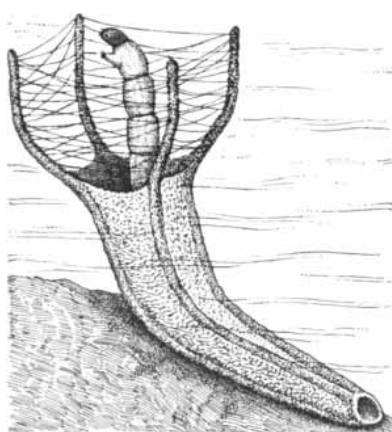
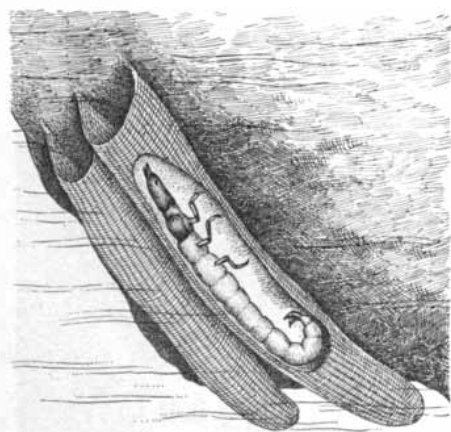
Some net-weaving caddisfly larvae collect food by other methods. Those of the philopotamid family build a sac-like tube that serves simultaneously as a shelter and a catch net. These sacs are remarkable in two respects. First, they are large: as much as five centimeters long and three centimeters in diameter. Second, their mesh is extremely fine. For example, the millions of individual rectangular mesh openings in the sac that larvae of the genus *Dolophilodes* construct in their final larval stage measure .5 micrometer by 5.5 micrometers. The mesh openings of the genus *Wormaldia*, formed by superposed layers of rectangular mesh, measure only .4 by .4 micrometer. The fine-mesh shelters have a large opening at the upstream end and a small opening at the downstream one. Inside the shelter the philopotamid larva periodically sweeps the particles of fine detritus from the mesh into its mouth, working with an array of bristles on its upper lip.

The larva of the small midge *Rheotanytarsus*, a filter feeder in the order of true flies, builds a tubelike shelter on the surface of submerged stones or plant debris. The structure is made out of silt particles bound together by the larva's silklike saliva; the upstream opening is large and the downstream one small. To this structure the larva adds two to five slender arms projecting upward from the large end, and between the arms it strings several threads to snare passing detritus. From time to time the larva emerges halfway from its shelter, eats the threads along with whatever food particles have adhered to them and attaches a new set of threads.

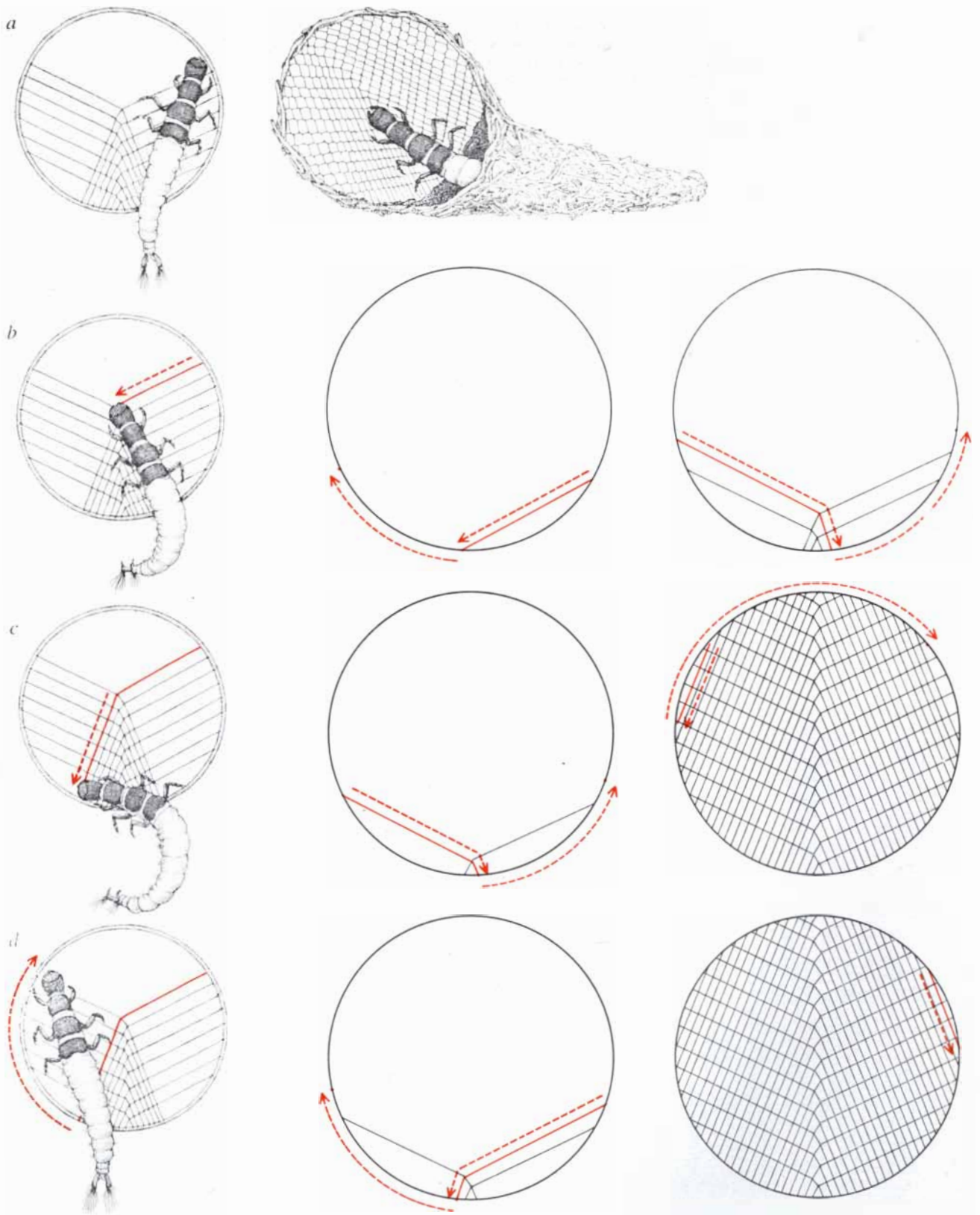
Various filter-feeding caddisflies, mayflies and true flies spend their immature stages in slow-moving or still waters. The best-known of these are the true flies of the family Culicidae: the mosquitoes. Mosquito larvae, called wrigglers because of their active body movements, feed on suspended organic matter in the water with a pair of modified mouth brushes. The brushes are similar to the head fans of the black-fly larva in structure, musculature and function. The water the mosquito larvae occupy need not be clean or of any great extent. Wrigglers are found in tree hol-



MODIFIED LEGS are used by the nymph of the mayfly genus *Isonychia*, at the left, and the larva of the caddisfly genus *Brachycentrus*, at the right, to capture food particles. The mayfly nymph faces the current and uses the brushes of its mouth parts to comb its foreleg setae clean of trapped food particles. The caddisfly larva builds a portable shelter that it occupies after anchoring it to face upstream. It then pushes all six of its legs out into the current. Its filtering arrays of short setae are on the hind legs and middle legs; these it sweeps clean with its forelegs.



SIMPLE NETS are built by a larva of the caddisfly philopotamid family, at the left, and a larva of the fly genus *Rheotanytarsus*, at the right. The caddisfly larva constructs a long, saclike net with a very fine mesh, anchored so that the larger opening faces upstream. The larva occupies the net and periodically uses the brush on its upper lip to remove the food particles trapped on the net's inner surface. The fly larva builds a tubular shelter out of particles of silt bound together with silklike saliva and adds projecting arms to its upstream end. It then strings sticky threads between the arms and usually eats the threads and the food particles adhering to them.



COMPLEX NET is built by larvae of the caddisfly genus *Hydropsyche*. It is shown at the top with its hooplike frame in place at the entrance to the larva's underwater shelter. The sequence at the left shows the method of construction, described by Werner Sattler. The larva grasps the frame and netting and attaches a strand of silk to the right-hand edge of the frame. It draws out the strand (*color*) first to

the center of the net and then down to the edge of the frame on the other side. A swing up to the left and a mirror-image repetition of the first motion completes the larva's figure-eight movement and adds another strand to the net. The idealized diagrams in the middle and at the right outline the attachment of the first four strands and last two. The larva is able to construct its net in seven to eight minutes.

lows, in snow pools and even in the water droplets that collect at the base of a leaf; they are also found in brackish water and latrine pits. They can thrive in such stationary habitats because the motion of their mouth brushes generates small currents that bring food particles within reach.

Some groups in the mosquito family are not filter feeders but have evolved mouth parts modified for browsing. Where the filter-feeding wigglers tend to sweep up particles just below the surface of the water, the browsers generally feed on the bottom. Their modified mouth parts make it possible for them to scrape food particles off organic bottom debris. Most larvae near the surface capture and ingest particles less than 50 micrometers in diameter.

Studies show that the survival of mosquito larvae is regulated not only by such major environmental factors as day length and the temperature, salinity and oxygen content of the water they inhabit but also by subtle chemical factors. Rex H. Dadd of the University of California at Berkeley has demonstrated that some of these chemical regulatory factors hasten the growth of the larvae by raising the rate of food intake and increasing the time the larvae spend feeding. The ultimate effect of this chemical stimulation would be to

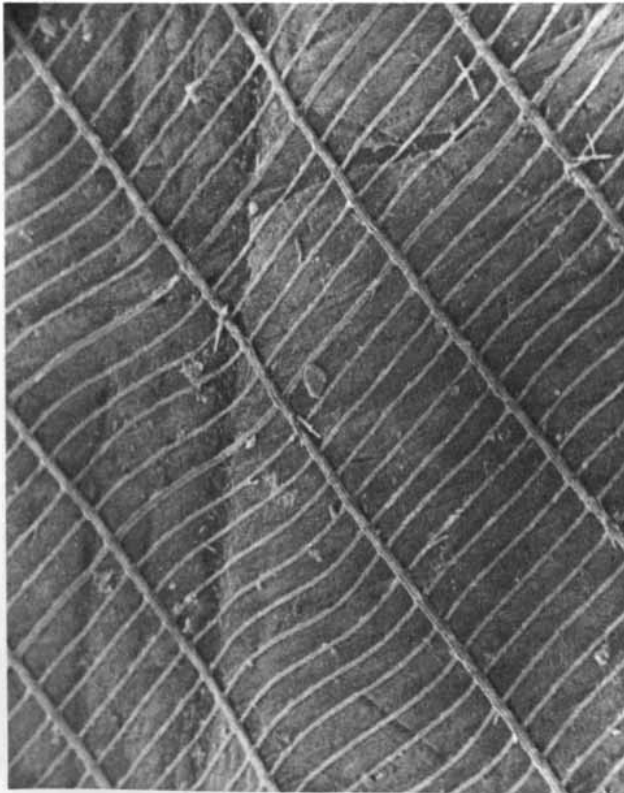
accelerate the buildup of dense larval populations if it were not for the fact that when the maturer larvae grow under crowded conditions, they produce substances that are highly toxic to the less mature ones.

Like mosquito larvae, the silk-spinning midge larvae of the chironomid family favor lakes and other still or slow-moving waters. They may dig a burrow in the bottom sediment or attach their shelter to the surface of a submerged log or the stem or a leaf of an aquatic plant. The midge larvae's spinning abilities have enabled them to adapt to a wide range of habitats. They are one of the most important primary consumers in aquatic food chains and have been known to reach population densities of more than 50,000 per square meter of bottom.

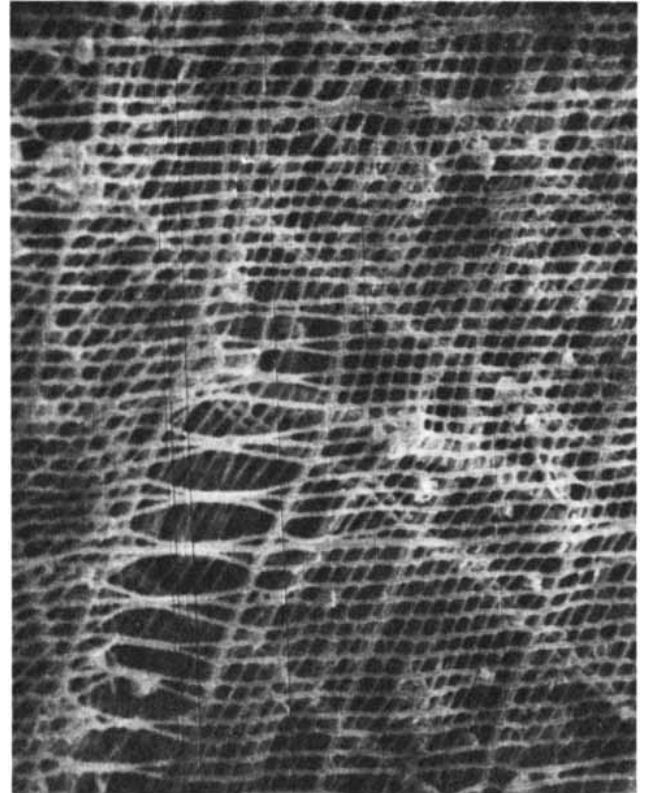
The midge larva may simply dig a burrow in the soft lake sediment or may work with its silk to build a shelter out of available particulate matter. The larva spins a thin conical net across the mouth of the shelter, a task that takes 30 seconds or so, and then, by undulating its body, pumps water through both the net and the shelter. If an accumulation of detritus plugs up the net, the larva reverses its undulations, generating a strong countercurrent that usually clears the obstruction.

In order to feed, the midge larva attaches itself firmly to the silk lining of its burrow with hooked claws and then devours both the net and the food particles adhering to it. In the interval between eating the old net and spinning a new one the larva defecates. The total elapsed time in the cycle, including the 30-second net-spinning interval, is between three and four minutes. Except for the fact that the midge larva creates its own water flow and eats its net rather than cleaning away the adhering food particles, it plays a role in the food chain of slow-moving depositional waters equivalent to that of the net-spinning caddisfly larva in fast-moving erosional waters.

The nymphs of some mayfly genera inhabit the silt and mud of near-shore lake bottoms and slow-moving streams; they may also live in submerged wood, such as tree trunks and dock pilings. Two of the common bottom dwellers are the nymphs of the genera *Hexagenia* and *Ephemera*; both use legs that have been modified for digging to construct a U-shaped burrow in the bottom sediment. Once the nymph is in the burrow it begins to undulate its respiratory gills. The current generated by the movement carries a supply of both oxygenated water and food particles through the burrow. A number of studies suggest that



MESHES OF TWO SIZES, both woven by caddisfly larvae, appear in these micrographs. At the left is the oblong mesh of the hydropsychid genus *Macronema*; each pore is about five micrometers wide



and 40 micrometers long. At the right is the two-layer mesh of the philopotamid genus *Wormaldia*. Overlap of two rectangular meshes produces the smallest-known caddisfly net pore: .4 by .4 micrometer.

Buy the Olympus OM-2, the camera that defines the state of the art.



An incredible camera. The Olympus OM-2 was created with a list of firsts that has made it second to none.

It was the first 35mm SLR to measure light during exposure. Directly off the film. For fool-proof exposures under any light conditions.

The OM-2 was the first to offer complete exposure control in every mode of camera operation. Even with multiple-flash. Even at the motor drive's blazing five frames-per-second.

It was the first to combine total automation with a light, compact, rugged design. With controls that are designed to be used. Not struggled with.

Add to this compatibility with over 250 interchangeable components of the OM System. Result: the camera that defines the state of the art.

And we'll turn your favorite OM-2 slide into a work of art. A 16x20 LaserColor® print.

Offer Expires May 31, 1981



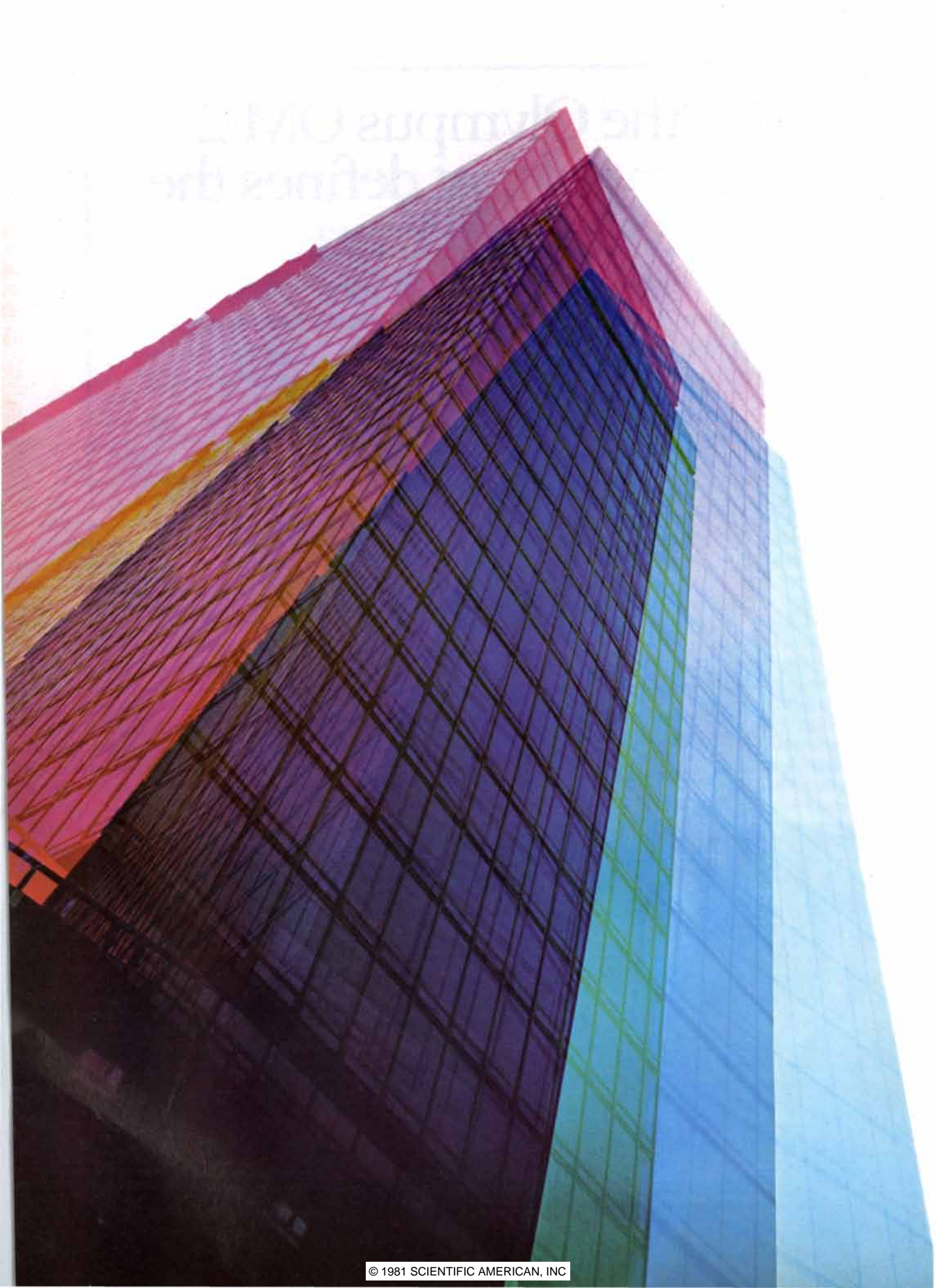
Photographer: David Deahl. Camera: OM-2. Print by LaserColor Laboratories

An incredible offer. And now, when you buy an OM-2, we'll create a spectacular work of art from one of your slides. A LaserColor print worth over \$40. Combining laser light and computer electronics, the LaserColor process makes prints of unsurpassed quality from slides. Compare them to ordinary prints from slides. For the first time, true color, detail and subtle shades come through.

Let your Olympus dealer show you what no ad could ever do justice to. The brilliant OM-2 and an amazing LaserColor original print.

You'll see why the camera that's had so many photographic firsts is first among knowledgeable photographers. For information, write Olympus, Woodbury, N.Y. 11797. In Canada, W. Carsen Co., Ltd., Toronto.

OLYMPUS®



The Olympus 612
defines the

Energy & Communications: the single-system solution.

The Bell System offers you a single solution to two critical problems: how to reduce energy consumption and how to improve communications capabilities.

Energy Communications with Dimension® PBX is an innovative service that makes use of any building's telephone lines, so you also save considerably on the cost of installation. It is easy to operate and doesn't require a special technical staff.

At Chicago's 650-room McCormick Inn, the system paid for itself in eight months. In a single month, it saved \$9,000 in electricity costs alone.

The system also provides businesses with new communications flexibility, improving overall productivity by providing greater control over internal and external communications.

Your Bell Account Executive, an industry specialist, is ready now to discuss the application that meets your unique needs.

Put our knowledge to work for your business.

The knowledge business



GM



FOR A DAY ON THE ROAD OR A NIGHT ON THE TOWN.

If you could buy only one Sport Coupe in the whole world, which would you select? Possibly there is only one that provides such a seductive response to that question. Camaro.

Here is a car conceived and engineered to the road. Designed and styled to the city. A calculated blend of line and stride.

Here, too, is a car that's

making performance affordable. For with all that roadability and class, Camaro also provides responsible fuel usage:

28 Est. Hwy. 20 EPA Est. MPG

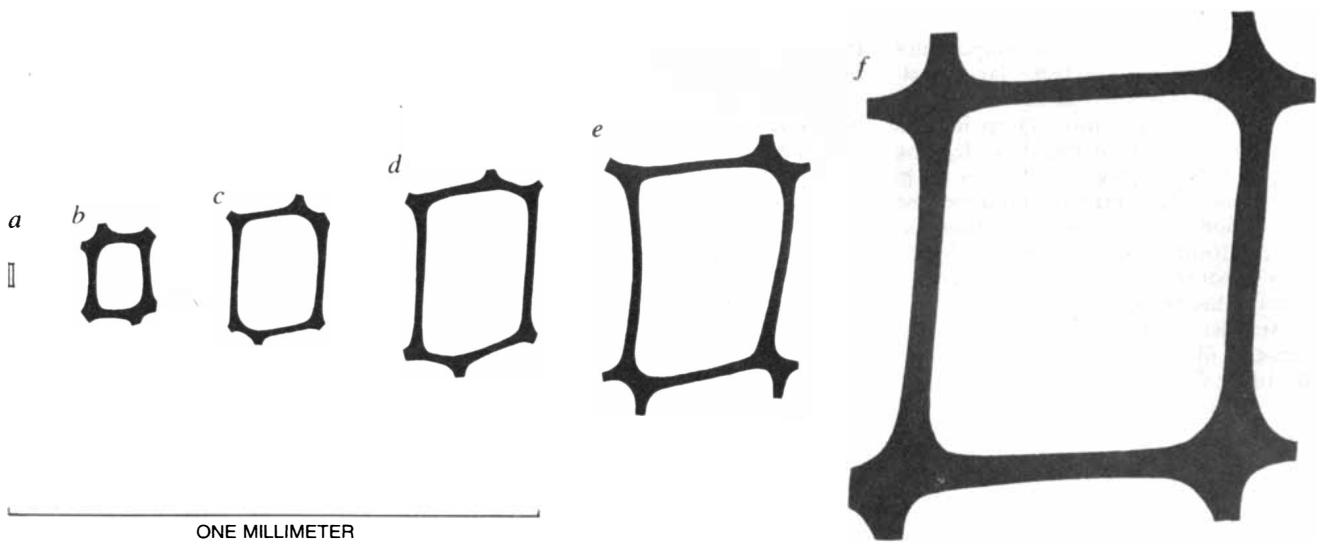
**1981 CHEVY CAMARO
THE HUGGER**



Use estimated MPG for comparisons. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower. Calif. highway est. higher.

See your Chevy dealer today about buying or leasing a new Camaro Sport Coupe.

And see how much one car can bring to all 24 hours of your life-style.



MESH SIZES among different genera of hydropsychid caddisflies vary greatly. The individual pores shown in outline here range from the oblong of *Macronema* (a) to the rectangle of *Arctopsyche* (f). All

six meshes are representative of the nets that the caddisfly larvae of the Savannah River basin construct during their final larval stage. The *Arctopsyche* larvae preferentially inhabit fast-flowing streams.

the nymphs can feed in their burrows by filtering particles out of the passing water. They are also known, however, to emerge and browse on the bottom. In many areas of the U.S. Middle West hordes of these mayflies once blanketed the ground during their brief summer emergence. Their numbers have now been greatly reduced by the pollution of the kind of waters they inhabit.

The nymphs of two tropical mayfly genera, *Povilla* and *Astenopus*, are among those that attack submerged wood. They excavate a U-shaped burrow in the wood with their sturdy mouth parts and take shelter in it after lining it with a silky material. Then they wave their long abdominal gills to generate a flow of water that carries oxygen and food particles through the burrow. They filter the food out of the passing water with dense arrays of bristles on their forelegs, head and mouth parts.

The caddisfly larvae that similarly inhabit lakes and sluggish streams include the larva of one genus, *Neureclipsis*, that weaves a characteristically cornucopia-shaped net. The German entomologist Caroline Brickenstein, studying these larvae, found they took three days to construct their net. The biggest nets may be more than 20 centimeters long and have an opening 13 centimeters in diameter. The silk loses much of its elasticity and strength after a few days, and the larva spends additional time between feeding periods replacing worn-out strands on the interior wall.

The *Neureclipsis* larva feeds primarily on small aquatic invertebrates, and so its nets are often found in great abundance in the outflow streams of lakes where these tiny prey exist in large numbers. Although the preferred habitat of the larvae is slow-moving water, they sometimes build their nets in water flowing at

velocities of up to 30 centimeters per second and even higher. The occasional nets seen in such fast-moving streams are noticeably smaller than those found in slower-moving water, suggesting that the higher velocities impose limitations on net size.

The larva of the caddisfly genus, *Phylocentropus*, is one of the few caddisfly net spinners that live in stream areas where deposition predominates over erosion. It builds a long, Y-shaped tube that is buried several centimeters deep in the stream bottom. One arm of the Y is elongated and extends several centimeters upward into the water; the other arm, which has a bulge in it, is shorter and barely protrudes above the bottom. The larva normally occupies the longer of the two arms. By undulating its body it causes water to flow into the longer arm and out through the shorter one, where the bulge holds a catch net. Between intervals of undulation the larva enters the shorter arm of the burrow and feeds on the very fine detritus adhering to the net and the inner walls of the tube. Gut analyses indicate that most of the particles the *Phylocentropus* larvae ingest have a diameter of less than 10 micrometers.

The habitats available in a particular aquatic ecosystem are limited in number. How do filter-feeding insects manage to share them? The answer is clear: the various genera have evolved many different adaptive mechanisms, both behavioral and structural. A further mechanism, involving habitat selection, is common among certain filter feeders, particularly black-fly larvae and the net-spinning larvae of caddisflies of the hydropsychid family.

The water of a lake or a reservoir holds products of decomposition (derived from the bottom sediments) and large populations of microscopic plants

and animals. Accordingly these larvae aggregate in great numbers near lake outlets and dam spillways. The time of year when the nutrient-rich water spills downstream in abundance is often spring, and so immature filter feeders near an outlet at that season probably enjoy a selective advantage over insects maturing at other seasons of the year or at habitats downstream. It has been found, as might be expected, that some insect species occupying such nutrient-rich spill habitats grow faster and reach the adult stage sooner than representatives of the same species living elsewhere. The result is an abbreviated life cycle best suited to the exploitation of a seasonal abundance of food.

The tendency of black-fly larvae to aggregate in these spill habitats is presenting a serious health problem in Africa. The bite of the adult female black fly transmits the filarial worm that in man causes onchocerciasis, or "river blindness." The construction of dams and other impoundments in the developing nations of Africa has led to an increase in the number of black-fly breeding sites and a spread of the disease.

The length of a filter feeder's life cycle is affected by factors other than the abundance of food. The water temperature, or more precisely the accumulation of heat over a period of time, is one such influence. Working in Michigan with Ross, one of us (Merritt) found that black-fly larvae took longer to develop when they hatched at near-freezing winter water temperatures than they did when they hatched and developed in the rising water temperatures of spring. The larvae that developed in winter also went through more molts and were larger when they entered the pupal stage than the larvae that developed more rapidly in the spring. In Georgia one of us (Wallace) has observed a somewhat

similar response to water temperature among maturing caddisfly larvae, although the lower temperatures did not result in additional molts. Other investigations have shown that filter-feeding insects living at higher altitudes or in cooler latitudes normally produce one generation per year whereas those at lower altitudes or in warmer latitudes may produce two or even three generations in the same length of time.

Another adaptive factor in the coexistence of filter feeders is a staggering of the life cycle, as when the insects occupy

the same habitat but do not go through the same phases of development at the same time. Such temporal variations in life cycle can serve a number of adaptive purposes. For example, at any one time one population may be consuming foods that are different from those of the other. Or the period of maximum growth of the individuals in one population may be different from that of the other, and so the times of maximum food demand are staggered.

There are also changes in diet with larval development. A number of inves-

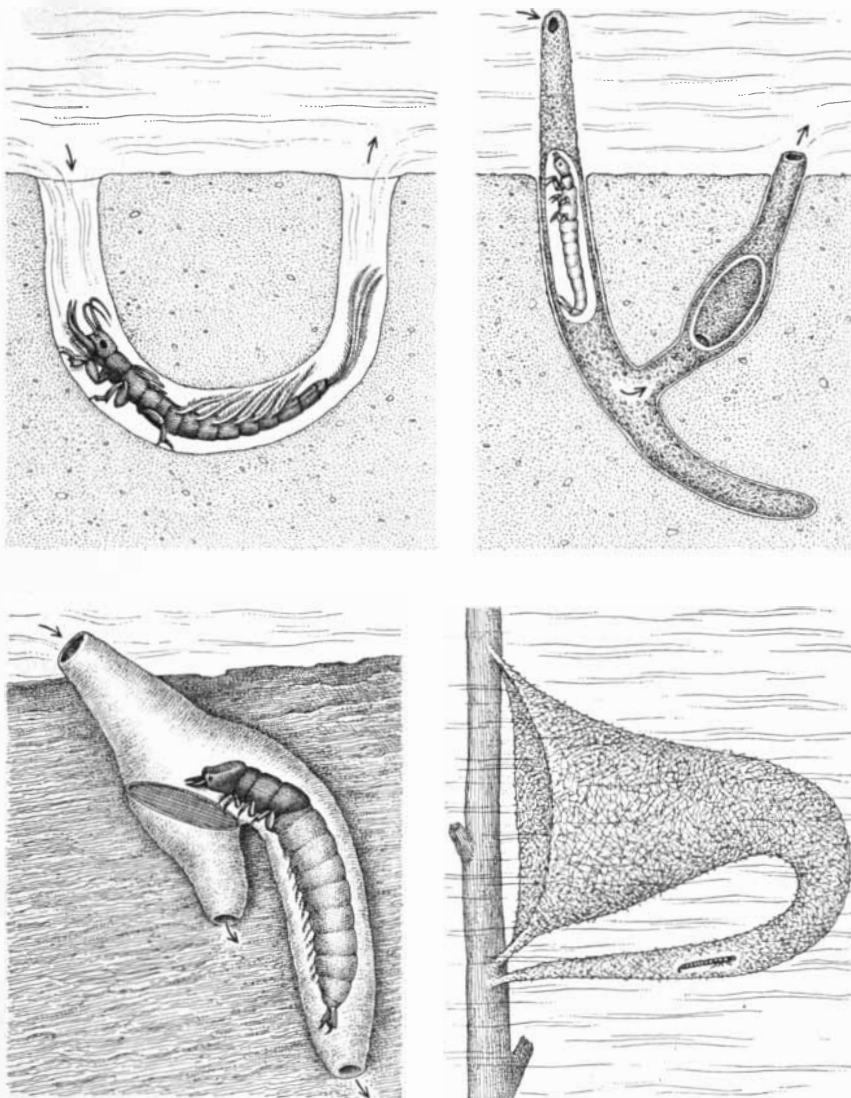
tigations have shown that the early larvae of some hydropsychid caddisflies feed mainly on fine detritus and on diatoms and other algae. As the larvae go through successive molts, however, they begin to consume increasing amounts of animal tissue. The food the filter feeder preferentially ingests during its period of maximum growth is the one it can assimilate most efficiently.

Aldo S. Leopold, a pioneer in wildlife studies, once pointed out that the processes of nature cause all materials, including organic ones, to move predominantly in a downhill direction. A corollary, Leopold suggested, was that the continuity and stability of upland communities depends on life forms storing nutrients and organic matter and participating in other processes that retard the downhill trend. It is evident that the filter-feeding aquatic insects are among these life forms, but it is difficult to measure their contribution with much precision.

The difficulty arises from the unidirectional flow of water, which is foremost among the downhill carriers of material Leopold had in mind. With an open-ended system such as a flowing stream the relative value of inputs and outputs is not easily calculated. Some commonsense evaluations are nonetheless possible. For example, in comparison with the total downstream transport there is little upstream movement of materials. One can cite instances of net upstream travel by fishes and by bottom-dwelling animals and even by aquatic insects after they have reached maturity and taken wing, but all these movements are trivial compared with the downstream one.

Exactly what do the filter feeders accomplish by way of retarding the downhill process? One of their major contributions may be to retain part of the food they ingest and to alter the rest and pass it along. For example, studies of six species of net-spinning caddisfly larvae in a stream in southern Appalachia indicate that these filter feeders actually add more detritus to the stream than they withdraw from it. Only between 2 and 20 percent of the filter feeder's intake of detritus is actually assimilated. At the same time the larva is also ingesting, but far from completely assimilating, high-quality animal tissue and somewhat lower-quality plant matter. The filter feeder's feces, although they may contain between 80 and 98 percent of the larva's own intake of low-grade detritus, will also contain some unassimilated animal and plant matter. By assimilation the larvae lower the net food value of what they ingest, but their feces, together with whatever colonizing microorganisms the feces may acquire in their travels, are available for reingestion by other filter feeders farther downstream.

This process can be considered the



SLOW-WATER FEEDERS include two, a nymph of the mayfly genus *Hexagenia* at the top left and a larva of the caddisfly genus *Phyllocentropus* at the top right, that occupy tubes or burrows and pump water currents through their retreats by undulating their bodies. The mayfly nymph sets up the flow of water by waving its dorsal abdominal gills and collects food particles with the setal brushes on its forelegs and mouth parts. The caddisfly larva builds its branched tube with silk and sand grains and weaves an irregular net in the shorter branch; body movements pull water into the long branch and push it out of the short one, and the larva periodically sweeps the net and the inner wall of the tube clean with the brushes of its forelegs and mouth parts. Two other caddisflies depend on weak currents to bring food particles to their nets. The larva of the genus *Macronema*, at the bottom left, builds a shelter with its entrance directed upstream and cleans its net with the brushes of its forelegs and mouth parts. The larva of the genus *Neureclipsis*, at the bottom right, constructs a large cornucopia-shaped net that can be 20 centimeters long. The larva usually feeds on live prey that accumulate at the narrow end of the net.

Bridging The Gap!



R&D MEXICO

P.O. Box 992, Farmingdale, NY 11737

- YES, let's bridge the gap. Send me the latest on science, industry, business and more with R&D MEXICO.
 - Enter my 1 year subscription, 11 issues, just \$12* (33% Charter Discount)
 - Enter my 2 year subscription, 22 issues, just \$21* (41% Charter Discount)
- Bill me later Payment enclosed

Name _____ Title _____

Company _____

Address _____

Home Office Signature _____

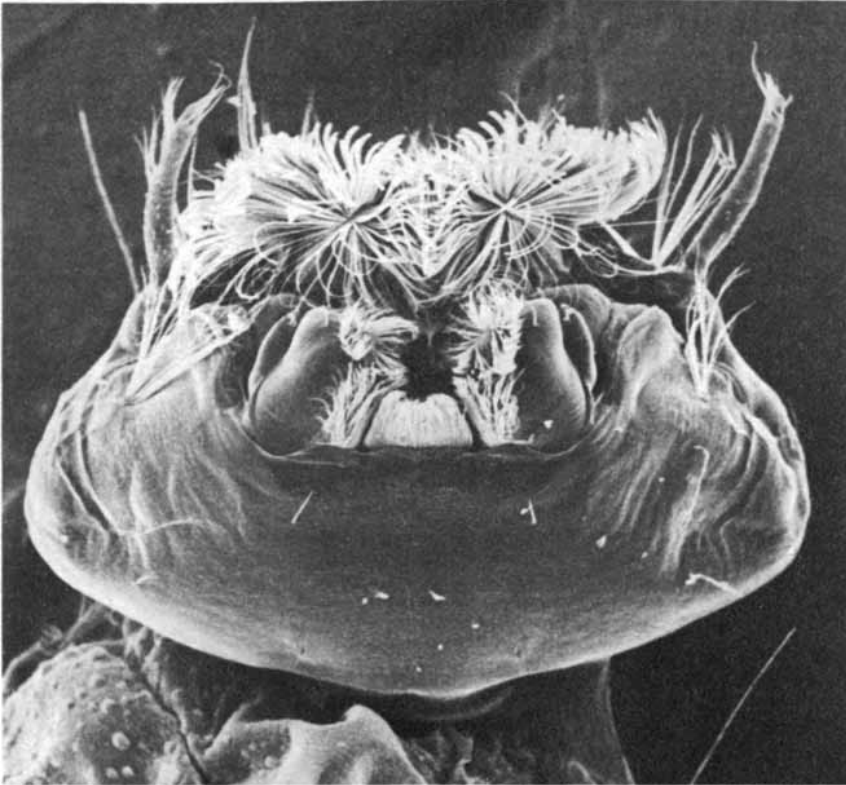
CHARTER GUARANTEE: 100% satisfaction or full refund on unmailed copies.

010401

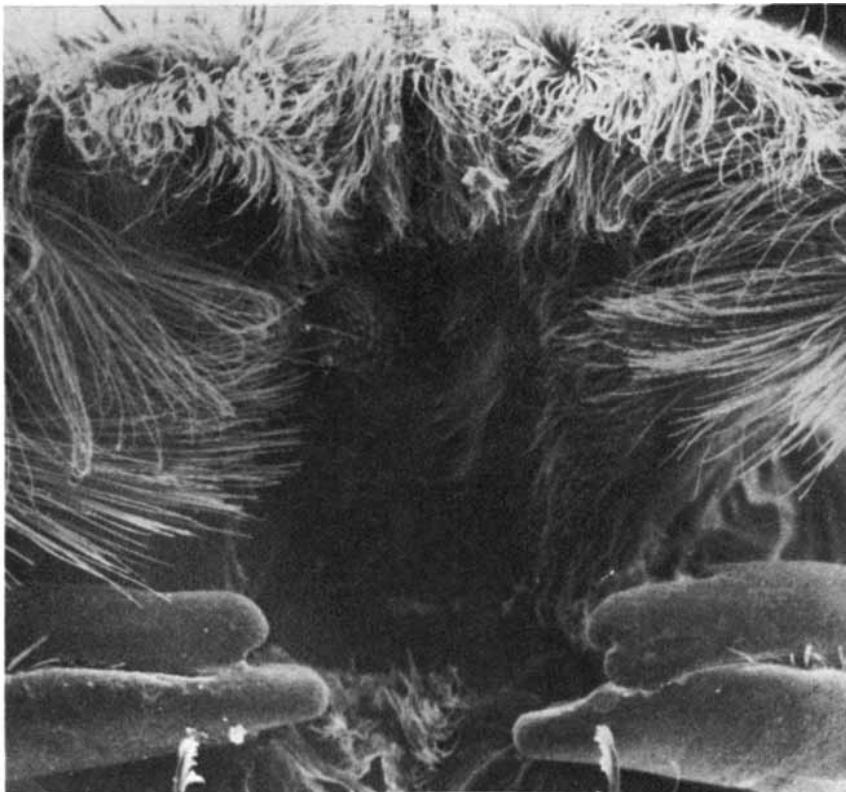
** All rates U.S. funds. Add \$5 (U.S.) per year for delivery outside U.S., possessions and Mexico.*

R&D MEXICO bridges the information gap between scientists, specialists and business decision-makers. Inside every issue you'll find valuable information on developments from Mexico affecting industry, technology, and science — domestic and foreign. Now's the time to get up-to-date with R&D MEXICO at these SPECIAL CHARTER DISCOUNTS.

A publication of the National Council of Science and Technology ● CONACYT



MOUTH BRUSHES of a mosquito larva are seen from below in this scanning electron micrograph made by Craig. A filter feeder of still-water habitats, the mosquito (*Culiseta inornata*) moves the brushes rhythmically to draw a current of nutrient-bearing water toward its mouth.



CADDISFLY MOUTH BRUSHES are used to sweep fine particles from its catch net into its mouth. Seen in this scanning electron micrograph is the larva of a hydropsychid caddisfly of the genus *Macronema*. Objects at the bottom of the micrograph are the larva's mandibles.

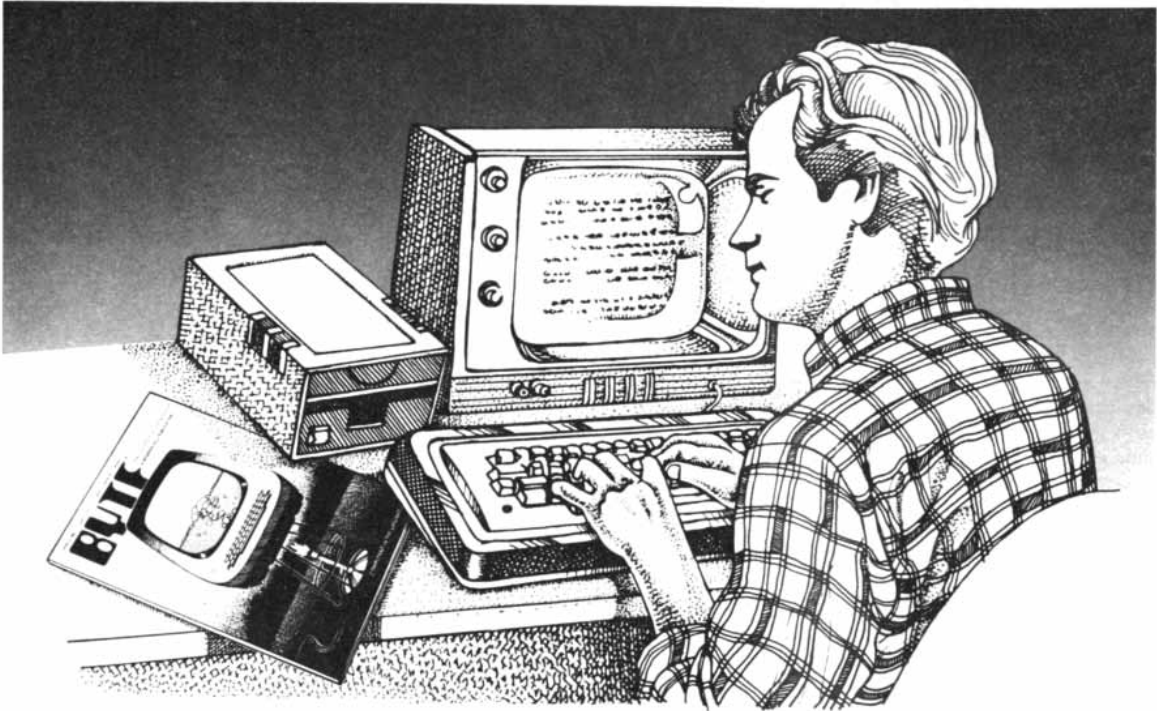
starting point of a recycling process that increases the efficiency of a stream ecosystem in terms of the utilization of its organic inputs. If it were not for the filter feeders, much of the organic matter being transported downstream would be utilized only by the microbial components of the ecosystem. In a stream where successive populations of filter feeders are active, however, the organic matter may yield fractions of its stored energy repeatedly on its long journey to the sea.

Jackson R. Webster of the Virginia Polytechnic Institute and State University has applied to this recycling process the term "spiraling." He has done so to emphasize the longitudinal and unidirectional aspects of cycling within a stream environment. For example, particulate matter of high quality such as animal tissue and algae may be utilized quickly and thus can be characterized as having a short spiraling distance. Low-quality detritus can be characterized as having a longer spiraling distance. The shorter the spiraling distance is, the greater will be the proportion of organic matter that is converted by the metabolism of filter feeders into carbon dioxide and is thereby removed from the total amount of organic matter downstream. And the more diverse the trapping mechanisms of the various filter feeders are, the greater will be the efficiency with which organic matter is removed.

This, however, is only one of the factors relating to the efficiency of food utilization. For example, the particulate matter that is most abundant in streams has a particle diameter of 25 micrometers or less. The filter-feeding insects that capture particles in this size range are black flies, midges and certain caddisflies. These are the same filter feeders that select food almost entirely by particle size. Therefore in the overall efficiency of the ecosystem these insects may accomplish more in retaining organic matter than the filter feeders that selectively trap larger particles.

The few studies made so far indicate that over short downstream distances the filter-feeding insects utilize only a small proportion of the passing organic matter. They convert that proportion, however, into organic matter of a more complex form and with a higher food value. That organic matter consists of their own bodies, which are potential food for predators, such as carnivorous insects and fishes higher up the food chain. Hence it is clear that the filter feeders, evolving to occupy a variety of habitats and to employ many kinds of capture mechanisms, act to retard the dominant downhill movement of organic matter by both retaining it and altering it.

BYTE...for the Serious Microcomputer User.



Do you use a personal computer for business or pleasure? If so, join the over one-quarter million professionals, beginners, engineers, and business people who get the latest microcomputer information in **BYTE**.

With a circulation of 170,000 and a total readership of 320,000, **BYTE** is the world's largest computer magazine. Each month's issue brings you the latest about graphics, mass storage, personal data

bases, word processing, speech synthesis, new integrated circuits, new software for engineering, business and fun, plus do-it-yourself projects of all kinds.

And if you are in the market for hardware or software, our new expanded product review section gives you the unbiased facts.

So, if you are serious about micro-computing, **get serious about BYTE.**

Mail coupon or
call us toll-free

800-258-5485

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.

BYTE Subscription Dept. P.O. Box 590 Martinsville, NJ 08836

Please enter my subscription for:

One year \$19 (12 issues) Two years \$34 Three years \$49

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 _____

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 \$21 (12 issues) Two years \$38 Three years \$55

Europe, one year (air delivered) \$43

All other countries, one year (surface delivered) \$35. Air delivery available on request.

BYTE[®]
the small systems journal

Please allow 6-8 weeks for processing.

7S 41



© BYTE Publications, Inc. 1981

JOIN THE 1900 STAFF MEMBERS AT SOUTHWEST RESEARCH INSTITUTE IN SAN ANTONIO

the freedom to think, the freedom to question,
the freedom to answer.

Our staff of internationally recognized research engineers and scientists and supporting technical and administrative personnel enjoy a professional working environment that enhances creative research and encourages excellence. Southwest Research Institute's 471 acre campus houses 11 research divisions. Each division is involved in a variety of engineering and applied research projects for individuals, private industry and government. Some of the areas of research we are active in, and seeking researchers for at the present time, include alternate energy and synthetic fuels, rotor dynamics and vibrations, explosive hazards and ballistics, electronic and electrical engineering, engine research and development,

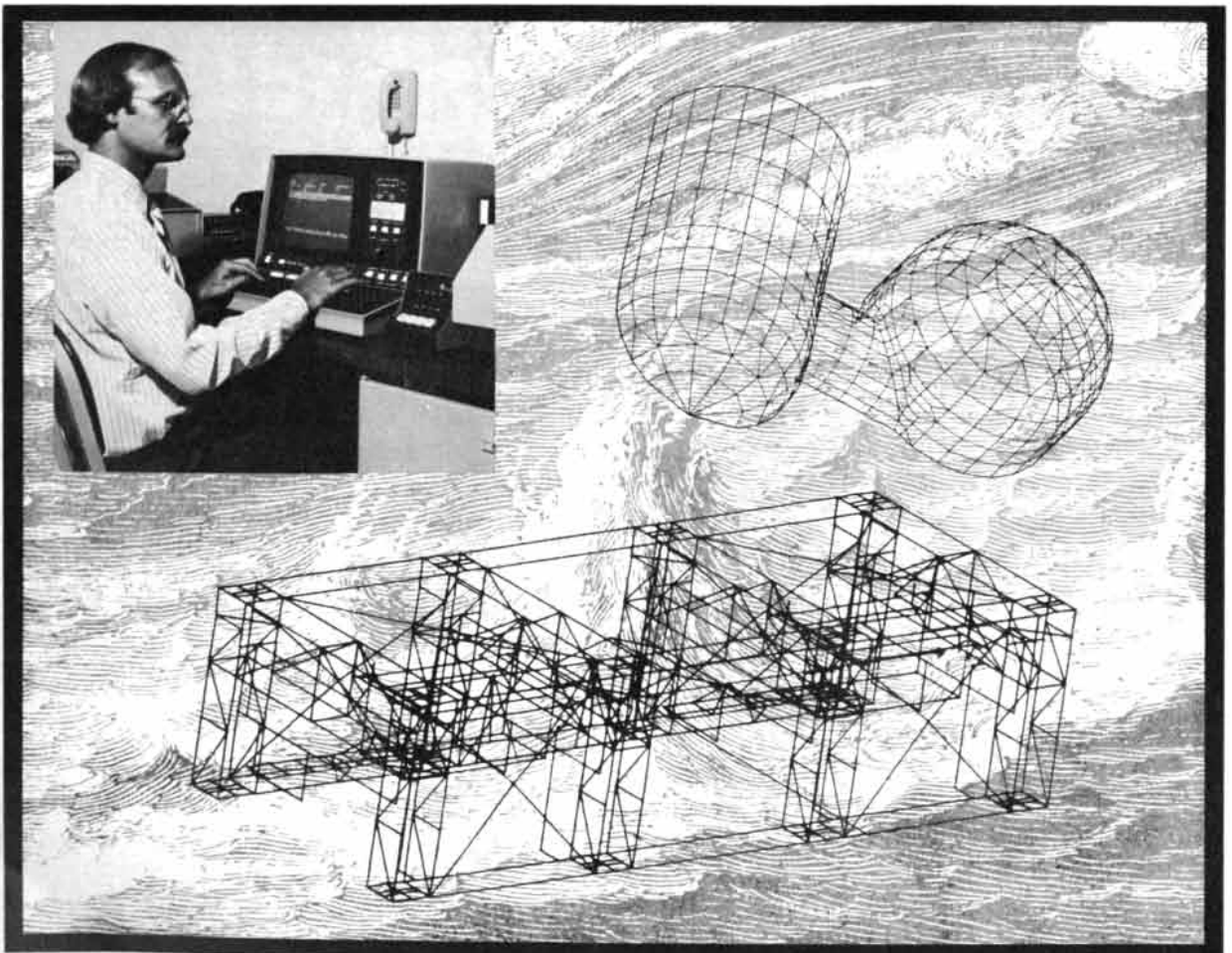
mechanical design, material science, fracture mechanics, fluid dynamics, mining technology, fuels and lubricants, fire technology, non-destructive evaluation, nuclear safety, structural analysis and design, ocean engineering, and highway safety and transportation systems.

Historic San Antonio, the nation's 10th largest city, is an ideal setting for our constant search for new ideas and their development. Our modern complex is surrounded by the beautiful southwestern ranchlands that make San Antonio a unique place to live and work.

We have current opportunities for qualified individuals who want to contribute to the technological excellence and innovation to which

Southwest Research Institute is dedicated. For consideration, please forward your resume to:

**Personnel Department No. 326
Southwest Research Institute
P.O. Drawer 28510
San Antonio, Texas 78284**



We Are An Equal Opportunity Employer M/F

MUSIC FROM MARLBORO

Anyone who has heard Marlboro music-making

- at Rudolf Serkin's unique musical community in Vermont
- in one of 300 cities that have presented Music from Marlboro performances
- on one of the 39 recordings released by CBS Records and the Marlboro Recording Society
- on radio broadcast series in over 60 cities
- will welcome this invitation to join in celebration of the 30th Anniversary of Marlboro Music.

Here is how you celebrate:

- You enjoy the matchless performances to be heard on one or more of the Marlboro Recording Society records of lesser-known masterworks listed below.
- You make as large a contribution as you can to the Marlboro 30th Anniversary Endowment Fund, helping to match a National Endowment for the Arts Challenge Grant.

MRS-1 BRAHMS: Serenade in A Major, Op. 16—Pablo Casals, Conducting The Marlboro Festival Orchestra

MOZART: Sonata in A Major, K. 12; Sonata in F Major, K. 13—Pina Carmirelli, violin; Rudolf Serkin, piano; David Cole, cello

MRS-2 SCHOENBERG: Suite, Op. 29—Marlboro Instrumentalists; conducted by Leon Kirchner

BOCCHERINI: Quintet in F Minor, Op. 42, No. 1—Pina Carmirelli, Jon Toth, violins; Philipp Naegele, viola; Fortunato Arico, Dorothy Reichenberger, cellos

MRS-3 SCHOENBERG: Serenade, Op. 24—Marlboro Instrumentalists, Conducted by Leon Kirchner

SCHUBERT: Introduction & Variations, Op. 160—Paula Robison, flute; Rudolf Serkin, piano

MRS-4 BEETHOVEN: Trio in C Major, Op. 87—John Mack, Joseph Turner, oboes; Patricia Grignet, English horn

BEETHOVEN: Variations in Eb Major, Op. 44—Hidetaro Suzuki, violin; Ronald Leonard, cello; Rudolf Serkin, piano

MRS-5 NIELSEN: Woodwind Quintet, Op. 43—Paula Robison, flute; Joseph Turner, oboe; Larry Combs, clarinet; William Winstead, bassoon; Richard Solis, horn

WEBER: Clarinet Quintet in Bb Major, Op. 34—Harold Wright, clarinet; Vera Beths, Mari Tsumura, violins; Philipp Naegele, viola; Jeffrey Solow, cello

MRS-6 BEETHOVEN: Trio in G Major, Op. 9, No. 1—Isidore Cohen, violin; Nobuko Imai, viola; Nathaniel Rosen, cello

BEETHOVEN: Quartet in C Major, No. 3 (WoO 36)—Rudolf Serkin, piano; Felix Galimir, violin; Nobuko Imai, viola; Nathaniel Rosen, cello

MRS-7 HAYDN: Trio in G Major, Hob. XV:15—Rudolf Serkin, piano; Michel Debost, flute; Peter Wiley, cello

BEETHOVEN: Trio in Bb Major, Op. 11—Rudolf Serkin, piano; Richard Stolzman, clarinet; Alain Meunier, cello

MRS-8 MOZART: Serenade No. 11 in Eb Major, K. 375 Marlboro Instrumentalists

GOUNOD: Petite Symphonie—Marlboro Instrumentalists

MRS-9 ADOLF BUSCH: Sonata No. 2 in A Minor, Op. 56—Pina Carmirelli, violin; Rudolf Serkin, piano

MRS-10 BOCCHERINI: Quintet in E Minor, for Guitar and Strings, G. 451—David Starobin, guitar; Pina Carmirelli, Philip Setzer, violins; Philipp Naegele, viola; Peter Wiley, cello

RAVEL: Chansons Madécasses—Katherine Ciesinski, mezzo-soprano; Judith Mendenhall, flute; Paul Tobias, cello; Luis Batlle, piano

MOZART: Sonata in Bb Major, K. 292—Alexander Heller, bassoon; Yo-Yo Ma, cello

MRS-11 MOZART: Serenade No. 10 in Bb Major, K. 361—Marlboro Instrumentalists; conducted by Marcel Moyse

MRS-12 BACH: Fourteen Canons on the First Eight Notes of the Aria Ground from the "Goldberg Variations," BWV 1087—Marlboro Instrumentalists; Rudolf Serkin, piano

REGER: Sonata in Bb Major, Op. 107—David Singer, clarinet; Rudolf Serkin, piano

MRS-13 DVORAK: TRIO IN E MINOR, OP. 90. "DUMKY" Bruno Canino, piano; Daniel Phillips, violin; Jerry Grossman, cello

BOCCHERINI: QUINTET IN D MAJOR FOR GUITAR AND STRINGS, G. 449. David Starobin, guitar; Pina Carmirelli, Joseph Genualdi, violins; Philipp Naegele, viola; Marcy Rosen, cello

Mischa Schneider, Artistic Director

Paola Saffiotti, Administrative Coordinator

The Recording Society is made possible through the generosity of Dr. and Mrs. Andre A. Aisenstadt.

If you would also like to help Marlboro match a National Endowment for the Arts Challenge Grant, please enclose your tax deductible contribution to Marlboro's 30th Anniversary Endowment Fund Drive.

Marlboro Music		
201 West 54 St., N. Y. C. 10019.		
Telephone: (212) 581-5197		
Please send me the records I have checked below at \$7.50 each plus \$1.50 each for postage and handling (\$9.00) total and accept my contribution (tax deductible) to the Marlboro 30th Anniversary Endowment Fund to help match the National Endowment for the Arts Challenge Grant.		
<input type="checkbox"/> MRS-1	<input type="checkbox"/> MRS-5	<input type="checkbox"/> MRS- 9
<input type="checkbox"/> MRS-2	<input type="checkbox"/> MRS-6	<input type="checkbox"/> MRS-10
<input type="checkbox"/> MRS-3	<input type="checkbox"/> MRS-7	<input type="checkbox"/> MRS-11
<input type="checkbox"/> MRS-4	<input type="checkbox"/> MRS-8	<input type="checkbox"/> MRS-12
		<input type="checkbox"/> MRS-13
		I enclose my check in payment for _____ \$ _____
		records at \$9.00 each \$ _____
		and for my contribution to the 30th Anniversary Endowment Fund \$ _____
		Total \$ _____

		Name _____
		Street _____
		City, State, Zip Code _____

Ancient Oared Warships

The city-states of classical Greece deployed fleets of swift ramming galleys equipped with as many as three banks of oars. These vessels later evolved into huge weapons platforms with thousands of rowers

by Vernard Foley and Werner Soedel

During the golden ages of Greece and Rome warships designed chiefly to be rowed played crucial roles in supporting long-distance trade and sustaining the bonds of empire. Sails, which had been introduced in about 3500 B.C., were adapted primarily for merchant ships, since labor costs dictated that their crews be kept to a minimum. Merchantmen resorted to oars only for clearing and entering harbor or for making some small progress during calms. Warships also relied on sails as much as possible, but when battle was imminent, their main rigging was struck and left on some convenient beach. Oars delivered less energy than a good wind, but for brief periods they gave warships adequate speed and superior maneuverability. The metabolism of the rower provided a way of storing energy in significant amounts and then releasing it quickly, an advantage that was unattainable with sails.

In spite of the comparative inefficiency of rowing for long-distance travel the ancient oared warship achieved an impressive level of engineering sophistication. The standard fighting ship of the era of Greek ascendancy was the trireme, as it is usually called in English. The word is derived from the Latin *triremis*, which in turn comes from the Greek *trieres*, meaning roughly "equipped by threes." The physical meaning of the name was once a matter of controversy among classicists. Today the issue is fairly well resolved, thanks largely to the work of the British scholar J. S. Morrison, who devised a plausible picture of these famous boats from such data as literary references, dockyard inscriptions, pictorial evidence and measurements of the slips where they were stored.

The trireme had evolved from simpler designs, chiefly Greek and Phoenician. These early craft were open boats rowed by a line of oarsmen along each side of the hull. From vase paintings and passages in Homer it appears that they were built for speed and the efficient utilization of the crew's energy, but a further element was necessary to bring them to

full effectiveness. This innovation came when the ram appeared in about 800 B.C., touching off a revolution in ancient marine architecture. Earlier naval combat was man-to-man, largely through boarding. Now one could aim instead at the destruction of the opposing ship's hull; inflicting casualties on the crew became a secondary consideration.

The advent of the ram greatly increased the need for speed and maneuverability. Accordingly the simple galleys of pre-Homeric times soon evolved into ships with long, low, slender hulls, which had room for as many as 25 oarsmen on each side. These craft were called penteconters, or 50-oared ships. The dimensions of their hulls, which have been determined with some certainty, reveal several design principles retained in the later development of more elaborate galleys.

The resistance of a ship to motion through the water is governed by four major factors. One of them is frictional resistance, which arises from the fact that water, like all other fluids, is viscous. Another is form resistance, or the degree of streamlining. When water molecules cannot flow close past the hull in smooth, unbroken lines, they separate from it and increase the volume of the water that must be displaced. If they separate sufficiently, they form eddies. The energy needed to create the eddies is lost from the amount available to move the ship, so that eddy resistance is a third form of hindrance.

All three of these factors are closely interrelated. A fourth kind of resistance, that of wave making, can be treated more independently. Like the other factors it increases with the ship's speed, but the rate at which it builds up is so much greater that ultimately it becomes the main impeding factor. It is chiefly determined by the ratio between the ship's length and the length of the waves its motion creates. To see why this is so one must visualize the way the waves made by a ship interact with its hull as the speed increases.

As the bow meets new water it gives

some of it an upward acceleration that varies with the speed. Gravity counteracts this motion and eventually the water falls to the surface level and below it. Thus a wave is formed. Meanwhile the ship has moved forward. The combined effect of the two motions is to set up one or more waves that are stationary with respect to the hull when the ship's speed is constant. At low speeds there will be a number of wave crests, as the individual water molecules will have time to rise and fall several times before the ship has passed.

The pressure increase at the bow that sets up these waves is matched by a pressure decrease at the stern that tends initially to lower the surface there. Both sets of disturbances transfer energy from the ship to the water and further increase the drag on the hull by enlarging its wetted area. If the hull speed is such that the standing wave created by the bow is in phase with the trough at the stern, these pressure-created differences in water level are reinforced. If the bow wave and the stern trough are completely out of phase, they substantially cancel. Higher hull speeds increase the wavelength of the disturbance system, so that a boat accelerating from zero passes through successive speed regions where the resistance grows first at a greater rate and then at a lesser one. In other words, the resistance curve climbs in a zigzag fashion.

Ultimately the boat's trim is affected. Its stern is lowered and its bow raised so that it must try to climb a hill of water of its own creation. The effect becomes critical when the wavelength of the standing bow wave reaches the length of the hull. When that happens, vastly more power than has been provided up to this point is needed to increase the ship's speed further. It follows that the greater the length of the hull is, the longer this speed crisis can be deferred. A long, slender hull also minimizes bow and stern disturbances by taking a larger amount of the total pressure necessary to support the boat away from these two points and distributing it instead along the sides, where it contributes through

skin friction at a lower rate to the total increase in resistance.

The crucial limit on the top speed of a boat that has a fixed power supply is given by the Froude relation: the ratio of the hull length to the square of the speed (since the resistance tends to rise with velocity squared). The relation is named after William Froude, the British marine engineer who first clarified these matters in the 1880's.

Naturally one cannot expect the ancients to have performed the foregoing analysis. Nevertheless, through trial and error they arrived at an effective understanding of the major impediments to high ship speed. The hull of a penteconter was as long as 38 meters, with a beam probably no greater than four meters. The approximate 10:1 ratio of length to beam typical of galleys designed for maximum speed persisted until the end of the ancient era of oared warships. It is generally agreed that for such ships the lengths were very close to the maximum feasible for construction in wood. In fact, in the case of the even slenderer trireme the limits of wood construction appear to have been exceeded. Even though an intricate system of mortises, tenons and pegs joined the

planking of the ship in such a way that strains were distributed to a considerable degree throughout the ship's skin, it was not safe to put a trireme into the water unless it was fitted with large cables run about the ship from stem to stern and then put under heavy tension by a windlass. Exactly where the cables were placed is not known, but their compressive effect on the ship seems to have been necessary to keep it from weakening dangerously when it was stressed. Wood by itself is difficult to adapt to joints subject to tension.

The width of these hulls was nearly the minimum needed to make room for two rowers sitting abreast and having their oar pivots on the gunwale (the top plank of the hull). The fulcrum of an oar cannot be too close to the handle or the rowing effort becomes excessive. Modern practice is to extend a third of the oar (or a little less) inboard of the fulcrum. Since the oars of triremes were of a length still common in oars today, the same allowance would seem appropriate for antiquity. Considering finally that some clearance is necessary between the oar ends for the shoulders of the adjacent rowers, one can see that the penteconter hull enclosed its propulsive units with scant side room to spare. The

draft of the hull was minimal as well, perhaps half a meter. Hence there was little fluid displacement and as a result minimal frictional resistance.

The hulls of these galleys could be of shallow draft because they were made of soft, light woods sparingly employed. The hull was only about 3.5 centimeters thick, and in some places it was even thinner. In writing of the hulls of merchant ships, which were usually more heavily built, ancient poets noted that only three fingers of wood separated the sailor from his doom. The minimization of the hull seems to have been carried to the point where the crew constituted about a third of the total mass of the system. We estimate that the displacement of the trireme was less than 40 metric tons, including the rowers.

The detailed configuration of the hull made further contributions to its speed. Today the lines of supertankers are a familiar sight, with their bulbous bow projecting forward underwater to minimize the formation of a bow wave by avoiding abrupt momentum changes in the water. Ancient rams varied in their configuration, but at least some of them would have had the same effect. At the stern galleys rose as gradually and cleanly from the water as any modern



BEST PICTORIAL EVIDENCE of the appearance of a Greek trireme at its prime is this fragment of a stone relief carved in the late fifth century B.C. The carving, known as the Lenormant relief after the French archaeologist Charles Lenormant, who discovered it in

1852, is now in the Acropolis Museum in Athens. It was presumably done to scale. As in the painting on the cover, only the top bank of rowers are visible; the other two banks are either down inside the hull or sitting inboard of the top rowers and thus obscured by them.

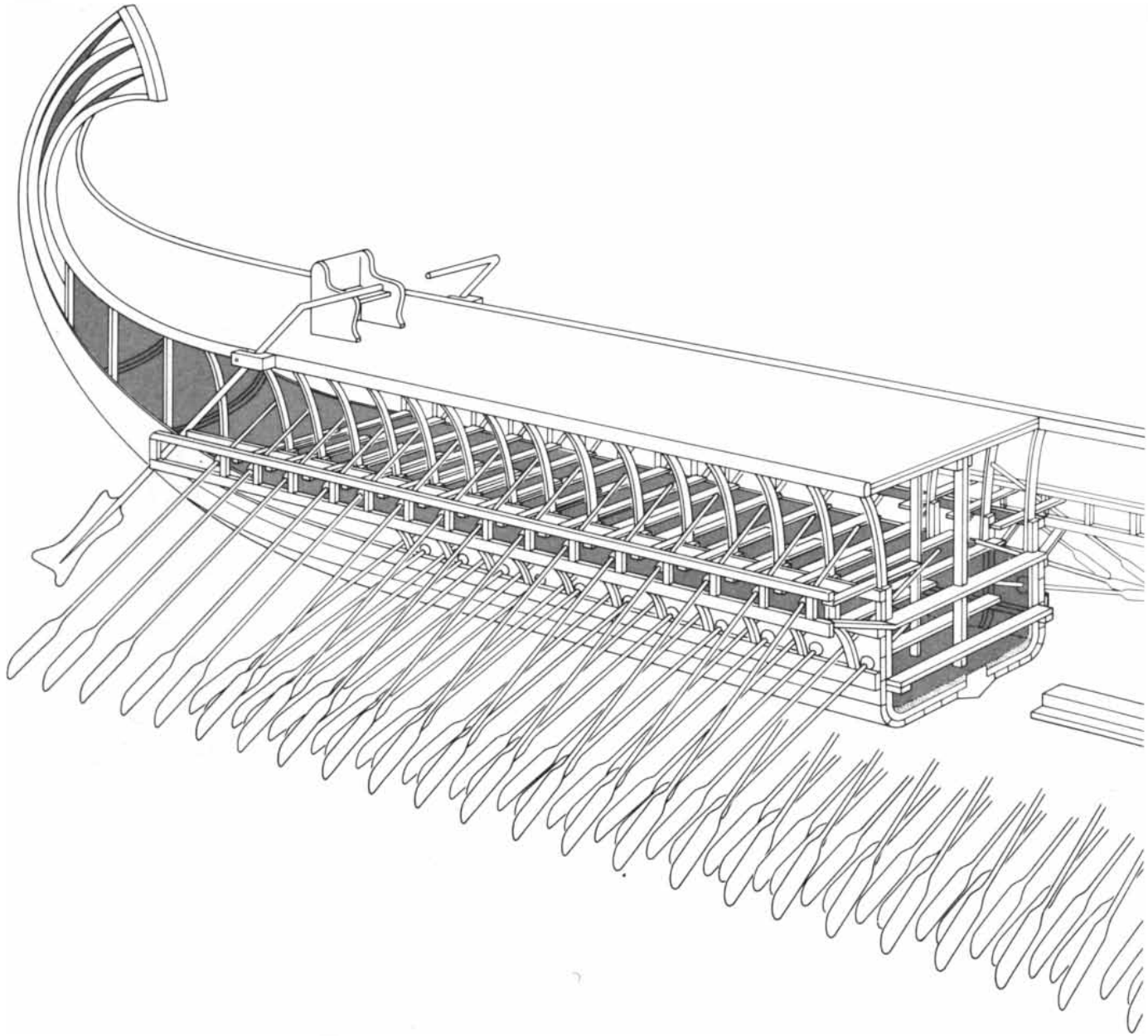
racing yacht, thereby minimizing the formation of eddies and a wake. Indeed, the galley may have had a cleaner departure than modern craft, which must often incorporate a fin to support the rudder. Ancient ships were steered with oars slung on each side of the stern. These probably blocked eddy formation and at the same time kept the wetted area to a minimum.

Above the water these craft showed equal thoughtfulness of design: their

low hulls minimized wind resistance, except at the stern, where the hull's rising curve was prolonged into a crest or a fan. This feature not only helped to protect the low hull from being swamped by a following sea but also, by increasing the hull's moment of rotational inertia, reduced the rolling that the typical lightly ballasted conditions of such ships tended to promote. The stern crest also helped to turn the boat into the wind if it was hit by a gust, reducing the boat's

chances of being swamped by a wave striking broadside.

Hence even before the age of the trireme the galley had come to embody considerable marine expertise. This paid off in speed: the fastest penteconters had a maximum speed estimated at 9.5 knots (17.6 kilometers per hour), which is only about a knot slower than the best performance of a modern racing shell.



CUTAWAY VIEW of a Greek trireme of the fifth century B.C. shows how the three banks of rowers were arranged in echelon to achieve maximum close packing. The 170 oarsmen were distributed

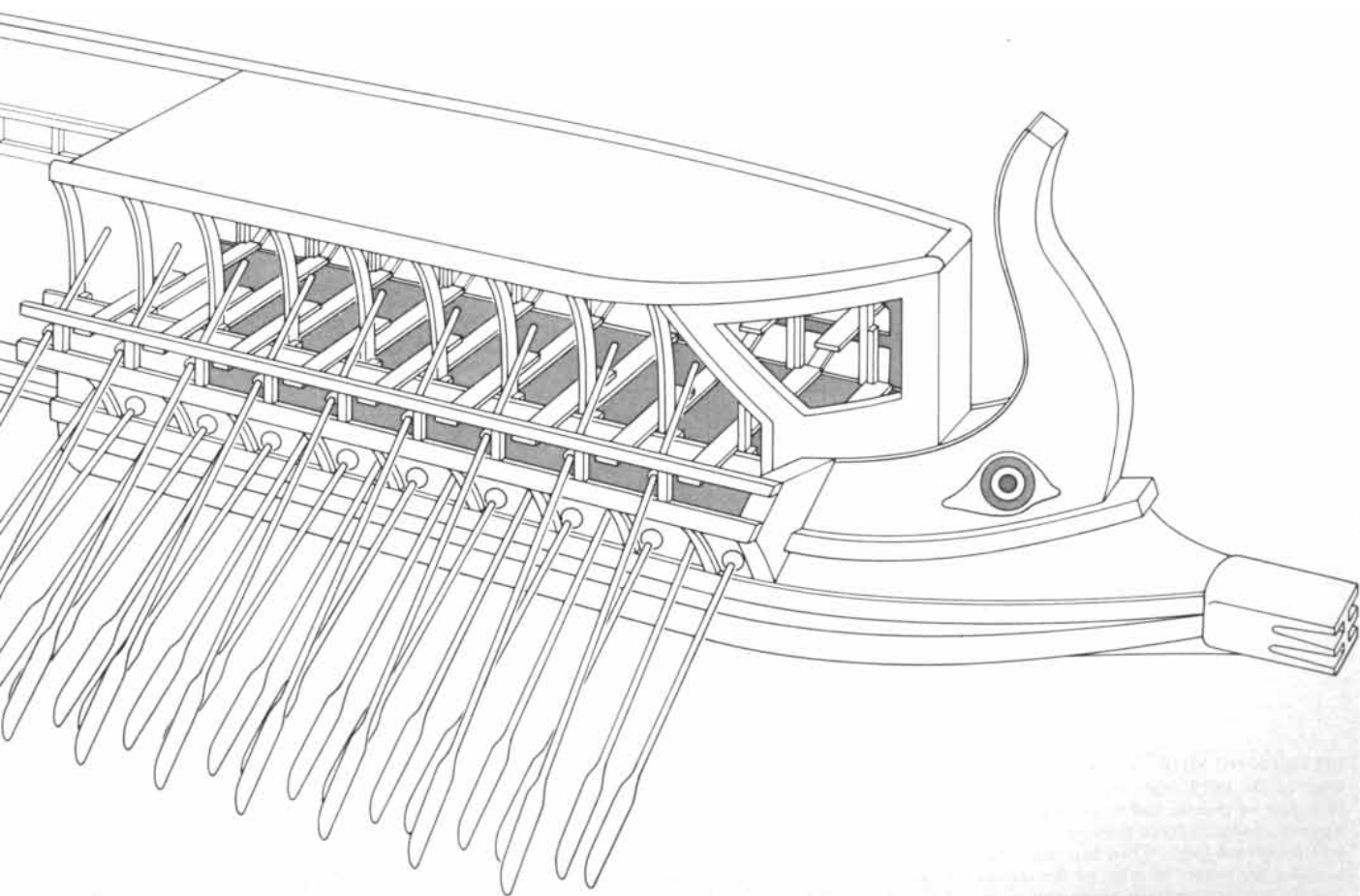
as follows: 31 in each top bank and 27 in each of the lower banks. Because of the narrowing of the hull at the stern there were no men in the lower banks there. Overall the vessel was approximately 35

In addition to their functional aspect galleys had an emotive aspect that seems to attach itself to most of the great ship designs of history. Their hulls were waterproofed by painting them with pitch, and ventilating holes or patches of contrasting color painted on their bows just above and behind the ram evolved into the eyes that some ships still sport today. The ram was often sheathed in bronze and decorated with multiple dagger designs or shaped

like the snout of a gigantic boar. Thus a galley closing to ram was black and staring, its crest upraised like the tail of a wild beast. Sometimes it carried wave screens along its upper sides made of hides tanned with the hair left on. The oars extended the zoomorphic effect; the synchronized flashing up and down of their salt-whitened shafts was often compared by ancient poets to the motion of a bird's wings. Even animal iridescence was imitated; Aristotle noted

that if the light was right as the oars broke water, their spray made rainbows.

It was chiefly the functional aspect of galleys that motivated their designers, however, and between Homer's time and about 500 B.C. several fundamental steps were taken in an effort to pack more motive power into the basic pent-conter hull. Not all the details are clear, and the full story is too complex for a brief presentation, but in essence the hulls were extended upward by the addi-



meters long by 3.5 meters wide. The front of the hull was extended into a ram, which was the ship's chief offensive weapon. Steering was by means of a pair of special oars slung at the stern. This recon-

structed view is based in large part on the work of the British classical scholar J. S. Morrison, who in turn relied heavily on the proportions of the Lenormant relief, shown in the photograph on page 149.

tion of superposed decks or seats to hold additional rowers. A second tier created a bireme, or "double-fitted" ship, and adding the third set of rowers resulted in the trireme.

To take these steps without endangering the stability of a hull that was al-

ready minimal required proceeding with care. Even in a bireme if the second set of rowers had simply been positioned above the heads of the first, the ship might have become dangerously top-heavy. The final design involved only a slight upward extension of the ship's

sides, probably less than half a meter. At the gunwale of this hull rowed a file of men, as before, except that now they were the middle layer in a group of three. Below them in the hold of the ship rowed men who were too close to the surface—about half a meter—to risk lowering the lip of the hull to that level. Hence their oars protruded from portholes in the sides, and leather gaskets secured to the porthole rim and the oar shaft kept the sea out.

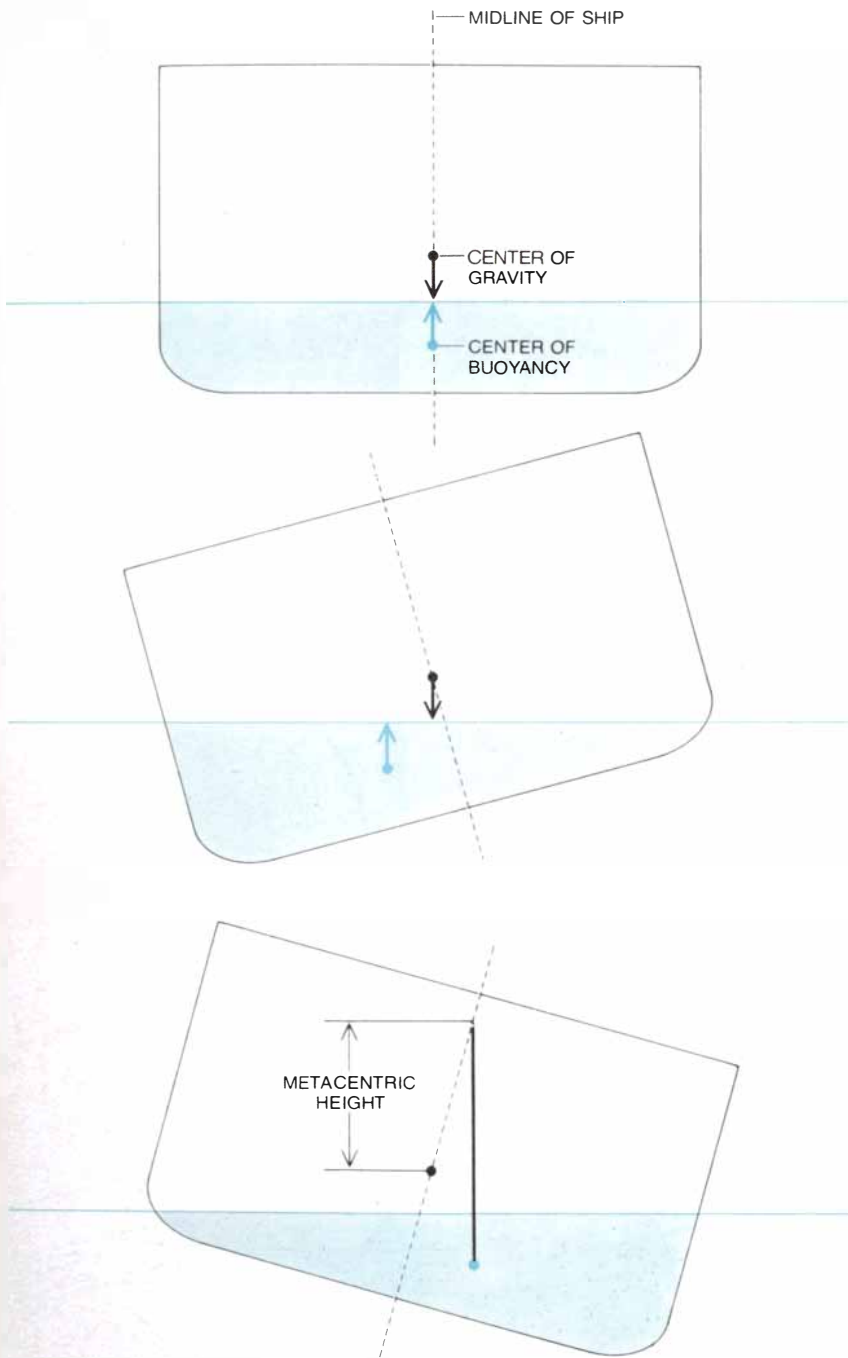
At the top rowing level of the ship was an outrigger, the innovation that had made possible the change from bireme to trireme. The outrigger enabled shipwrights to move the fulcrum of the top oars out about 60 centimeters from the gunwale. Now the top bank of rowers could sit alongside the shoulders of the middle rowers rather than over their heads, trimming the height of the superstructure by perhaps 50 centimeters.

The close packing of the rowers in the hull was further effected by arranging them in echelon, with the top men sitting about half a meter ahead of those in the middle bank and the middle men in turn about the same distance above and ahead of the men in the hold. The men were closely packed in each layer too, with only about a meter's space between them. This meant that if someone was rowing even slightly out of time, he would drive his back into the knuckles of the man at his rear or his knuckles into the back of the man in front.

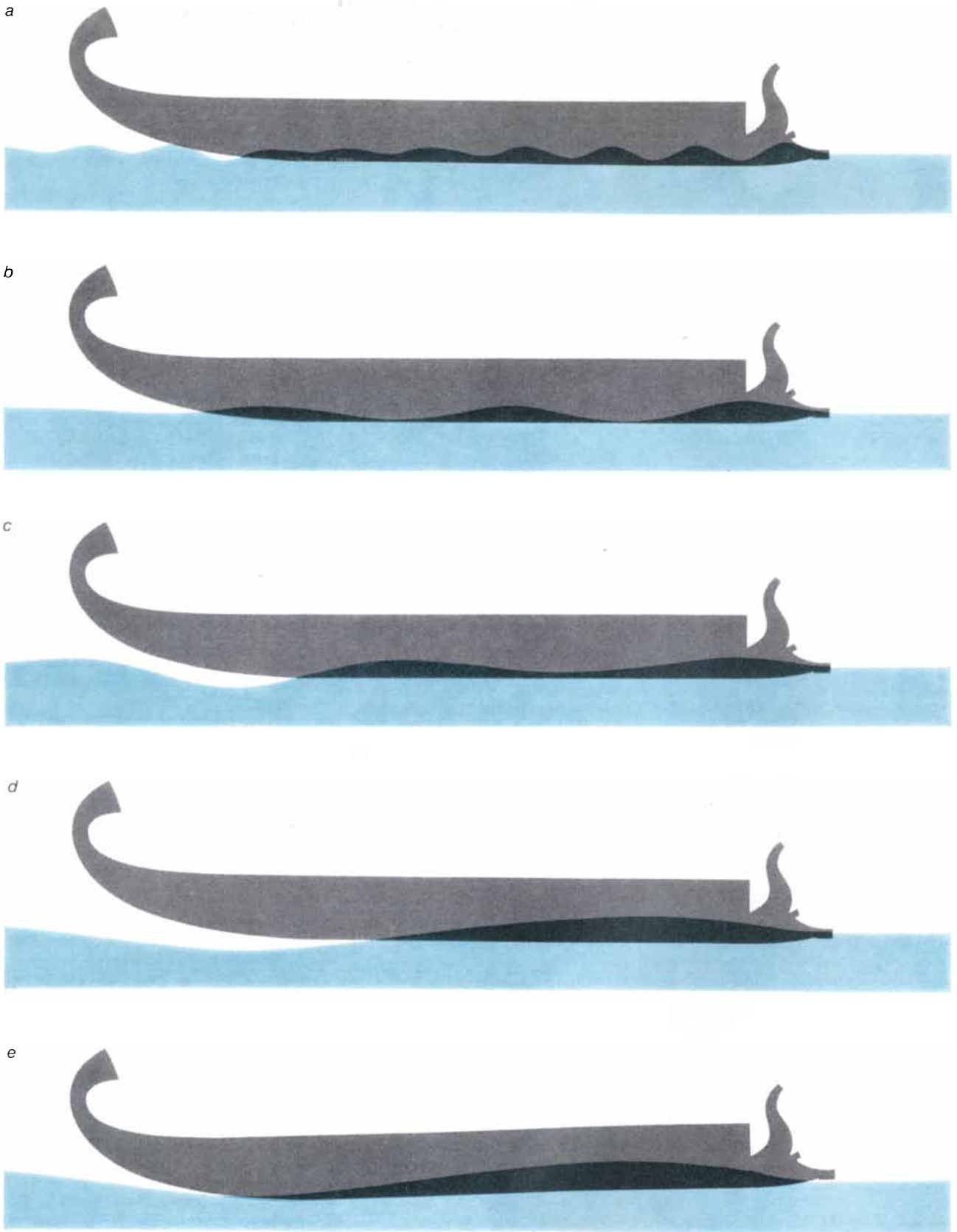
The men rowing in the hold had a further clearance problem. As the comic playwright Aristophanes remarked in his play *The Frogs*, their noses came close to the bottoms of the men seated just above and ahead of them as they reached forward to begin their stroke. The effort of the pull sometimes made the rowers break wind at just this moment.

The most severe coordination requirements involved the three men sitting in echelon. From the trireme shown in side view it can be seen that the oars in each trio must maintain the parallelism of their shafts very closely during the stroke, to well within 30 centimeters, if the blades of the oars are not to become entangled. If such entangling were to occur, it might by a domino effect spread to most of the remaining oars on that side, and thereby critically affect not only the speed of the boat but also its direction.

To control the system properly continuous training and drill were necessary, and due attention was paid to the incentive factor. Rowers in antiquity were recruited mainly from among the ranks of the free citizens, who therefore had a stake in their city-state's survival. Generally the use of slaves was limited to emergencies and then usually only after they had been freed. It can easily be seen that only one malcontent on a ship

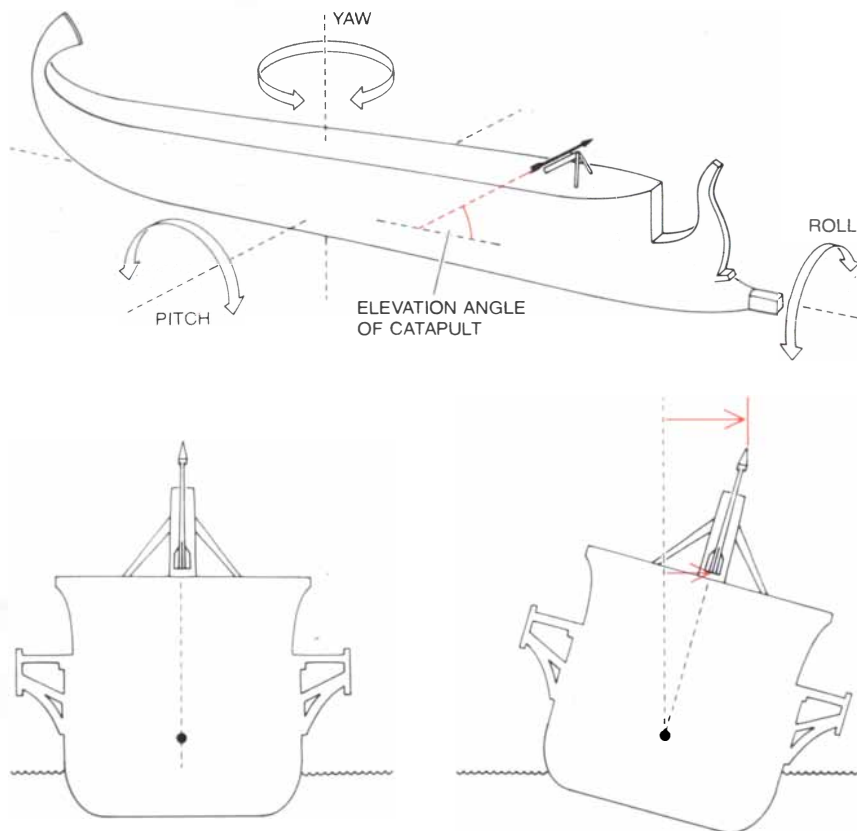


PHYSICS OF SHIP STABILITY was first investigated systematically by Archimedes, who lived in the third century B.C. He thought of the weight of a ship as being concentrated at its center of gravity and directed downward toward the center of the earth (black arrow). The upward, buoyant force (colored arrow) was at the center of mass of the part of the hull that was immersed (top). When the ship rolled, these two forces acted to return it to the vertical because the center of mass of the immersed part moved in the direction of the rolling (middle). In modern nautical terms the center of buoyancy of the ship is said to be at a certain vertical distance from the midline of the ship; the vertical distance between the ship's center of mass and the intersection of the ship's midline with a line projected upward from the buoyancy center is the metacentric height (bottom). The greater this height, the stabler the ship.



SYSTEM OF WAVES created by the passage of a ship through the water is the most important form of fluid resistance to the ship's progress. At low speed (*a*) wave formation affects the vessel only slightly. As the speed rises both the wavelength and the amplitude of the wave system are increased. The bow wave can either diminish or augment the stern wave, depending on whether the two systems are out of

phase (*b*) or in phase (*c*). Ultimately the crest of the bow wave will tend to raise the bow of the ship at the same time that the trough of the stern wave lowers the stern (*d*). When this happens, the ship will lose trim and a growing fraction of its power will be needed to push it uphill (*e*). To defer this condition the boat should be made as long as possible. All the waves in the diagram are exaggerated for clarity.



SHIP MOTION, which can be categorized into yawing, pitching and rolling (top), has a strong effect on the optimum direction of firing of a weapon such as the catapult. In the case of the ancient galleys, for example, the high length-to-beam ratio made them as much as 50 times more sensitive to rolling than to pitching. Since catapult firing at any appreciable range requires elevating the weapon, the head of a catapult bolt aimed straight ahead would be farther from the ship's center of rotation than the butt (bottom left). As the ship rolled, the projectile head would therefore be swept farther to the side than the butt (bottom right). Initial aiming could be done easily only when the arrow and the target were in a common vertical plane. A projectile fired after the ship had rolled only slightly off the vertical would fly wide of its mark. Hence the effectiveness of catapult fire must have been significantly reduced when the crew aimed more or less straight ahead, for example at a target they were also attempting to ram.

could ruin its performance. The lash was never used; a chanter called out the timing of the stroke. As a further incentive rowers received a comparatively high wage. It acted the more strongly on their performance because they were typically recruited from the poorer classes of the city, for whom the cost of the elaborate armor worn by the heavy infantry was beyond reach.

One can imagine that a certain social tension must have existed between the sweating rowers on a trireme and the heavily armed marines carried grandly on the deck for boarding actions. It was only in the more democratic states, such as Athens, that reliance on the rower-powered ram was carried to its extreme and the marine contingent trimmed to a minimum. In the triremes used in the Peloponnesian War the Athenians carried only 14 marines. The close packing of the crew, in contrast, made it possible to squeeze 170 rowers into a hull not much altered in length and breadth from that of the penteconter.

The payoff of this elaborate effort lay

in a surprising level of speed and maneuverability. Reasonable estimates of the top speed of a trireme range as high as 11.5 knots (21.3 kilometers per hour). Indeed, these estimates may be overly conservative, because they assume that the hull displaced water as it advanced, and there are marine engineers who argue that triremes were light enough and fast enough to achieve a semiplaning condition, rather like a modern speedboat. Under such conditions the top speed might be increased by as much as 50 percent. Such rates could be maintained for only five or 10 minutes before the crew tired, but at their peak they compare well with the speeds reached by the heavy horses of a medieval cavalry charge.

It probably took triremes about 30 seconds to reach top speed from a standing start, but since half speed could be reached in perhaps eight seconds and quarter speed in as little as two, it did not take long to give the boat enough energy to strike a decisive blow. Triremes were clearly very good for darting. More-

over, they were highly maneuverable. By rowing backward on one side and forward on the other, for example, their crews could spin them in little more than their own length. To see a fleet of some dozens of these ships cutting through the water on maneuvers or sprinting home after drill to determine which one could get into the harbor first must have been impressive.

One of the most famous long-distance trireme dashes of the period took place in 427 B.C. The city of Mitylene on the island of Lesbos had revolted against its Athenian masters and had then been retaken. The Athenian demagogue Cleon proposed that the entire population of the subject city be put to death, and his rhetoric carried the day in the public assembly. Accordingly a trireme was sent out to carry the order to the Athenian garrison. Given the high state of political passions in Athens, it probably set out early in the afternoon, soon after the assembly vote. As Thucydides writes, however, because of the horrid nature of its errand, it made no haste. Cruising on one or two banks and setting a slow stroke, it probably made no more than four or five knots. The next morning the assembly met again, cooler heads prevailed and the massacre order was repealed. Hoping for this change, the Mitylenian ambassadors in Athens had arranged for a fast ship and a crack crew, providing them with high-energy foods for the trip and promising large sums of money if they could catch the first galley before its dispatch could be carried out.

The second ship apparently left for Lesbos, a trip of some 345 kilometers, about 24 hours after the first. It made the open sea before nightfall, and the crew rowed continuously through the night. They were even fed as they rowed, with wine-moistened barley cakes. The night was clear enough for navigation and there was no head wind. In order to maximize their speed the commanders either took on enough rowers for an extra bank to rotate with the original three or kept two full banks out of three going continuously in watches during the night, with the third sleeping. Whichever technique was used, they reached Mitylene at midday, just after the arrival of the first galley. It appears, therefore, that they spent less than 24 hours in transit, cruising at close to nine knots (16.6 kilometers per hour). The dispatch had been delivered but the garrison had not had time to begin putting it into effect. A modern ferry makes the trip in 14 hours.

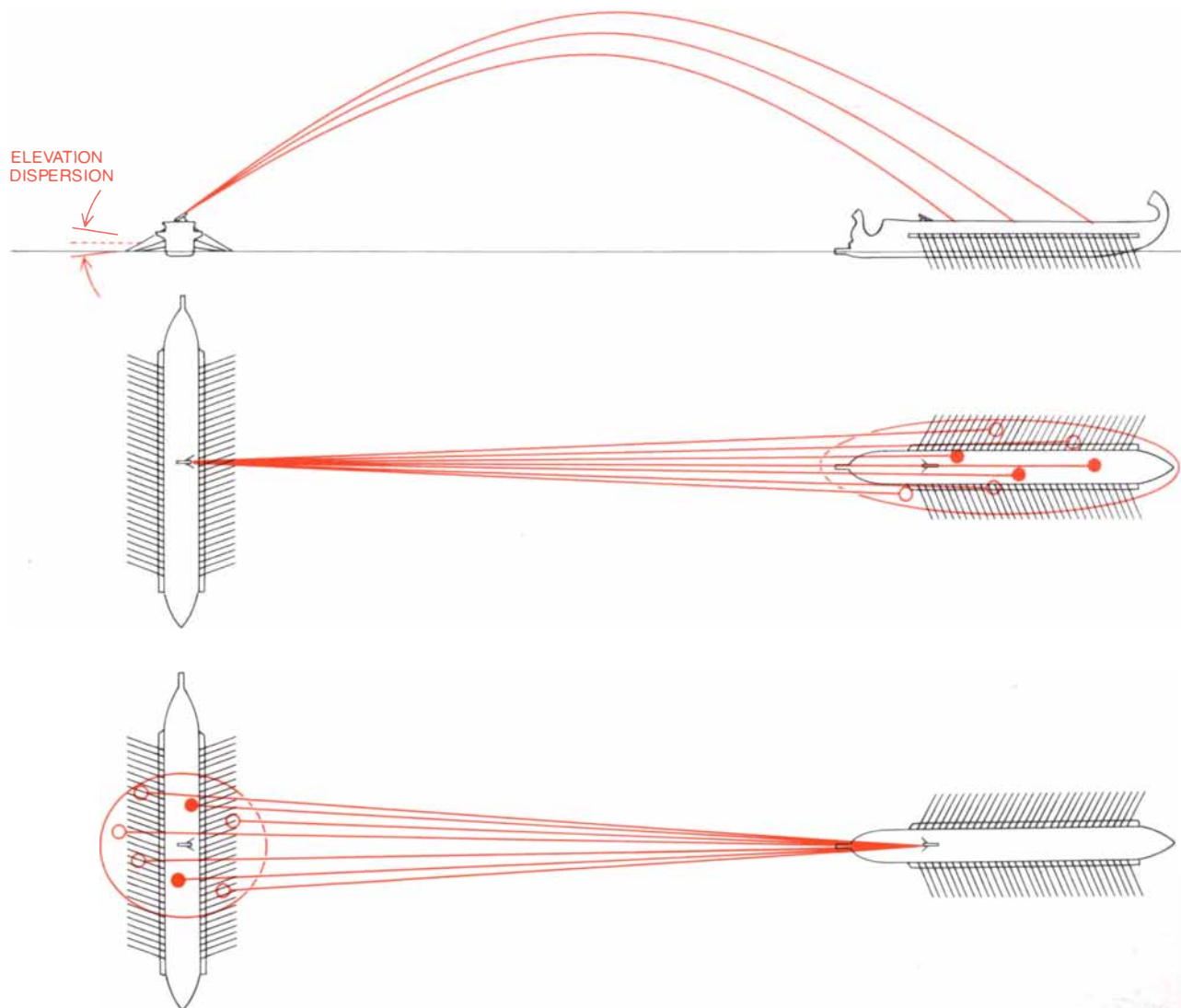
The next major step forward in galley design was taken in about 400 B.C. By that time ships of four banks, called quadriremes, were being developed, and in 399 an engineering team assembled by Dionysius of Syracuse built the first quinquereme, or five-banked vessel.

These bank numbers initially appear to be straightforward continuations of the trend previously established by single-banked galleys, biremes and triremes, but such was not the case. A number of considerations indicate that no ancient galleys were ever built with more than three superposed tiers of oars. Physically this conclusion can be justified because the oar angle necessitated by a fourth tier would have made the effort there severe. Even on a trireme the top bank's job was appreciably harder than the others, and those rowers sometimes received extra pay.

What seems to have happened instead is that one or more of the banks was

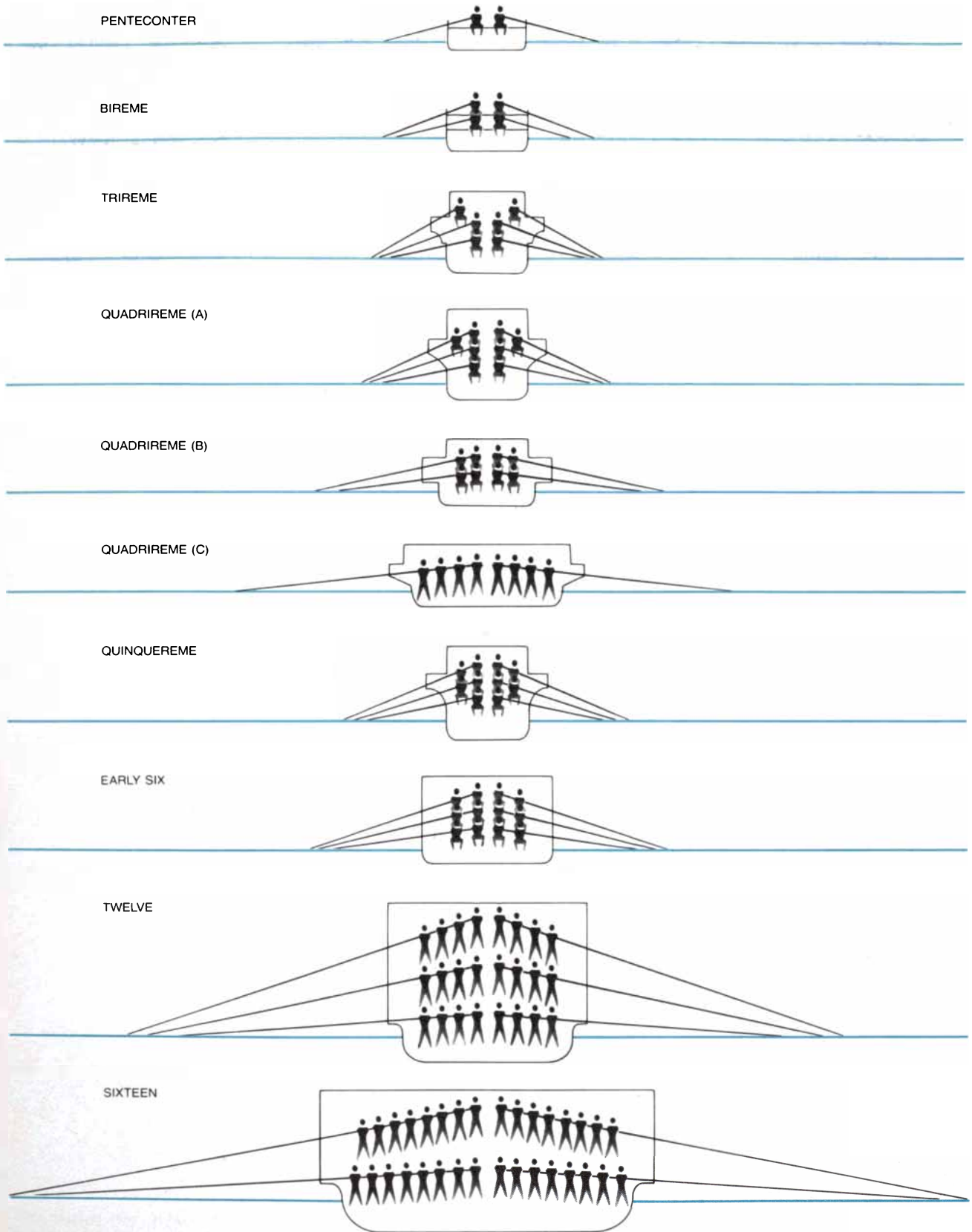
now double-manned. Given the minimum configuration of the trireme hull, double-manning was probably adopted first for the top bank, making the ship into a four. Only there could enough extra room for two more files of rowers be found without altering the design greatly. An alternative approach seems to have been taken by the Carthaginians, whose galleys probably followed the Phoenician preference for beamy hulls; they apparently began with a galley wide enough to accommodate four rowers abreast in any of its levels. The main line of galley development, however, seems to have stayed at first with the narrower hull.

To pass from the quadrireme to the quinquereme called for redesigning the hull, unless one assumes that shipbuilders were content to put the extra rowers close to the oar fulcrums, where their efforts would have added little. The most plausible design change to assume is that the builders deepened the outrigger, making it a two-story affair and enabling both the top and the middle banks to be double-manned. The bottom bank probably remained single-manned to preserve the speed advantages of a narrow immersed section as much as possible, although of course the draft of quadriremes and quinqueremes would have been increased by the weight of the



EFFECTIVENESS OF CATAPULT FIRE from a galley that is trying to avoid ramming would be enhanced by its own stability characteristics and by the shape of its target. Since such ships were particularly sensitive to rolling, the projectiles of the defending ship, fired more or less broadside, would undergo considerable dispersion in elevation, which would translate into a "footprint," or pattern of impacts, in the shape of an elongated ellipse (*top and middle*). Since the attacking galley in this case would itself be very long and narrow, however, there would be a good match between the impact footprint and the shape of the target and hence a greater probability of making

a hit. (The echelon arrangement of the rowers inside the hull of the attacking ship would in turn increase the chances that more than one might be injured by a single hit.) In contrast, the rolling of the attacking ship would cause considerable lateral dispersion of its projectiles, since it would be firing more or less straight ahead (*bottom*). If the target ship was broadside, this dispersion would not be very serious, but the small depth of the target would mean that even minor changes in elevation, such as those caused by rowing, could make the projectiles fall long or short. Catapult fire between ships, it seems, was more effective in a defensive mode than it was in an offensive one.



DEVELOPMENT OF ANCIENT GALLEYS went through three main stages. The invention of the ram in about 800 B.C. led from simple one-level ships such as the 50-oared penteconter (top) to the bireme and then, with the addition of outriggers, to the trireme. The advent of the trireme in effect ended the first stage of development, since there was no practical way of going higher than three levels of oarsmen. The next stage focused on increasing the number of men

per oar. Greater design flexibility now set in, since ship designers could combine numbers of rowers and banks in various ways. This approach was probably feasible up to a maximum of about eight rowers per oar (bottom), although the rowers in such cases must have had to stand and walk back and forth during each stroke. The final stage in this evolutionary process involved the construction of giant catamarans, or twin-hulled galleys (see illustration on pages 160 and 161).

additional rowers. Since the top bank on triremes held 31 rowers per side originally, and the other two banks held 27 rowers each, the new quadriremes probably had a total of 232 rowers and the quinqueremes 286.

The earliest pictorial evidence for this series of changes does not begin until a century after the introduction of quadriremes and quinqueremes. In conformity with the physical requirements of rowing, however, these illustrations show that the larger Greek galleys of the third and second centuries B.C. had most or all of their oars pivoted on a large outrigger, commonly referred to as an oar box or a rowing frame, rather than on the hull. The Romans tended to follow the Carthaginian preference for a wider hull with simple oar ports in it. They were not particularly good seamen and preferred whenever possible to grapple with enemy ships, pour troops on board and so convert sea fights into struggles where infantry tactics, at which they excelled, could be employed. Hence speed was not as important to them as troop-carrying capacity.

During the first decades of the existence of quadriremes and quinqueremes these designs spread slowly through the Mediterranean navies. In 330 B.C. Athens had only 18 quadriremes in her fleet, compared with 492 triremes. In the next six years this low number was more than doubled, however, and the construction of quinqueremes was begun. In assessing the reasons for the slow initial adoption of the heavier galleys it is helpful to compare some of the physical characteristics of triremes, quadriremes and quinqueremes in closer detail.

It is obvious from the design of the trireme that one of the chief problems faced by its designers was the avoidance of topheaviness. The conditions that govern whether a ship will be stable and capable of righting itself when it is tipped are usually considered today in terms of the vessel's metacentric height. The concept as such goes back no further than the 18th century, although most of the underlying physics and mathematics was known to Archimedes. Indeed, there are indications that Archimedes' investigation of floating bodies was motivated by his work on problems of ship stability. Consider the case of a typical vessel seen in cross section [see illustration on page 152]. Its center of gravity will be at some point within the hull and because of symmetry will lie somewhere on the midline.

One can imagine that the ship would behave as if its mass were concentrated at this point and were pulled downward toward the center of the earth by gravity. If the ship is tipped to one side, the center of gravity of the smaller section of the hull that is immersed will be moved in the same direction. The buoyant force of the water will act through the center of gravity of the immersed



If you'd like to know more about this unusual old store, drop us a line.

AT THE LYNCHBURG HARDWARE & GENERAL STORE, you'll find everything from darning thread to duck decoys.

And, just a short walk away, you'll find Jack Daniel's Distillery, where we still make whiskey the same way we did 114 years ago: gentling every drop with a process called charcoal mellowing. If you live in a big city, you won't find a store much like Lynchburg's. No matter where you live, you won't find a distillery much like Jack Daniel's.



CHARCOAL
MELLOWED



DROP



BY DROP

Tennessee Whiskey • 90 Proof • Distilled and Bottled by Jack Daniel Distillery
Lem Motlow, Prop., Inc., Route 1, Lynchburg (Pop. 361), Tennessee 37352

Placed in the National Register of Historic Places by the United States Government.

Garden Camera CALCULATORS

Hewlett Packard

HP 67	\$288.50
HP 97	568.50
HP 31E	34.50
HP 32E	42.50
HP 33E	69.95
HP 37E	57.95
HP 38E	89.95
HP 33C	68.50
HP 34C	114.50
HP 38C	124.50
HP 41C	188.50
Accessories	Call

Texas Instruments

TI 57	27.95
TI 58C	87.50
TI 59	198.50
PC 100C	155.00
TI MBA	52.50

Cameras

Canon A-1 w/50mm f1.8	379.50
Canon AE-1 w/50mm f1.8	257.50
Nikon FE w/50mm f1.8	299.50
Minolta XD-11 w/50mm f1.7	358.50
Olympus XA Kit w/Flash	149.50
Minolta 20 1 w/45mm f2.0	159.50



Call for low prices on Nikon, Minolta, Olympus and all Major Brand Cameras

Prices subject to change without notice
Speed your order
TOLL FREE!

Call 1 (800) 223-0595

Or Send postage and handling to

GARDEN CAMERA

345 Seventh Avenue, N.Y., N.Y. 10001

New York, Alaska & Hawaii Call:

Tel: (212) 868-1420 Open Weekdays 8:30-6:00

OPEN SUNDAYS 10-4 p.m. Closed Saturdays

WILD FINNISH ADVENTURES FOR HIKERS, BIKERS, CANOERS & SAILORS.

To: Finnair, P.O. Box 32,
Irvington, N.Y. 10533

Please rush me your free color brochure on Finland Adventure Tours for hikers, bikers, canoers and sailors before I go wild.

Name

Street

City

State Zip



SCA4

Great vacations start with the Flying Finnish

section, pushing vertically. As long as the center of mass of the immersed section moves sufficiently in the direction of the tipping so that its upward-acting buoyant force clears the downward-acting force of gravity by passing it on the depressed side of the hull, the net effect of the two forces acting together will be to cause the boat to right itself. The metacentric height of the boat is the distance between its center of mass and the place where the upward-directed force of buoyancy cuts the ship's midline. As can be seen from the illustration, the greater this height, the stabler the design of the ship.

Enough is known about the dimensions of triremes and about the placement of their crews and other stores inside the hull to allow estimates of their metacentric height to be made. The numbers one obtains are sensitive to changes in assumptions regarding such factors as the draft of the boat, the amount of wood used in the hull and the placement of rowers inside the hull. Even slight shifts in one or more of these factors can give rise to changes of five or six kilometers per hour in the wind force capable of capsizing the boat. As long as one keeps the parameters constant from one type of ship to another, however, the ratio between the metacentric values will be correct even though neither number has an absolutely accurate magnitude. For this approach to be valid, then, one must assume that in passing from the trireme to the quadrireme or the quinquereme the hull and its contents were changed as little as possible. Since marine engineering has historically been extremely conservative in making such changes, the assumption seems quite plausible.

The chief imponderable in this analysis is the amount of ballast carried by the ship, because there was enough hull space in ancient galleys to accommodate sufficient ballast to considerably alter the stability characteristics of the hull. In estimating the ballast we have chosen to keep the hull as light as possible, in keeping with the other obvious attempts by the ancient shipwrights to produce a hull that was maximally responsive to the rowers' efforts. Our estimate is 13,000 kilograms, a mass sufficient to prevent the pressure of winds of less than 60 kilometers per hour from blowing the hull over on its beam ends.

The outcome of this process of estimation is that triremes had a metacentric height of about .4 meter, which is marginally adequate but by no means a generous allowance. Hence the caution displayed by the ancients in avoiding battle unless the water was quite smooth is amply justified. In the case of the quinquereme the addition of four extra files of rowers to the top two tiers of the ship creates a dilemma that pits the vessel's stability against its speed. With a hull

of the same submerged dimensions as those of the trireme and with the same remaining assumptions as those given above, the metacentric height of the quinquereme goes to only .1 meter. Such a vessel could stand little in the way of wave pressure, and it would overturn in winds of more than 30 kilometers per hour.

This obviously unsatisfactory condition meant that shipbuilders were under heavy pressure to modify their hull designs. For reasons we have already noted the only good option open to them was to increase the beam of their ships. This approach, however, would sacrifice some of their speed, which was a key reason for designing ships of higher bank number in the first place. In the case of a quinquereme, if its hull were widened to achieve a stability equal to that of a trireme, its speed gain over the trireme would be only 14 percent and its acceleration would be poorer. The increase in the size of the crew, assuming the addition of 62 rowers on the top bank and 54 on the middle one, would be nearly 70 percent. If the stability of the quinquereme were reduced as postulated above, the gain in speed would be as great as 29 percent.

Given these problems, one can appreciate more clearly why two-thirds of a century passed before quadriremes and quinqueremes began to be built in appreciable numbers. A tendency by the Athenians to cling to their historic trireme tradition may also explain part of the delay. The ships that had won at Salamis would not lightly be discarded. This analysis, however, leaves unexplained the apparent rise in the popularity of quadriremes and quinqueremes beginning in about 330 B.C. and the extraordinary ancient naval arms race that ensued.

The development of ships rated as sixes had been undertaken in Macedon and Syracuse as early as 340 or 350 B.C., but little is heard of them until after the death of Alexander the Great in 323 B.C. and the breakup of his empire into rival factions. By about 315 one of Alexander's successors, Antigonos, was building sevens. By 301 his son, Demetrius the Besieger, had in his fleet ships with ratings as high as 13. Before Demetrius died he had built a 16. His rivals built ships that were meant to equal or better these, but strangely some of them had lower bank numbers. Lysimachus, for instance, built an eight that was supposed to withstand Demetrius' 16. This match clearly implies that something further was happening to the meaning of the bank numbers. The movement peaked with a 20 and two 30's built at mid-century, and a gigantic 40 launched perhaps four decades later.

Even less is known of these ships than of triremes, and they have been equally controversial. Recently, however, satis-

DESCRIPTION	JAN 1980	FEB 1980	MAR 1980	APR 1980	MAY 1980	JUNE 1980	JULY 1980	AUG 1980	SEPT 1980	OCT 1980	NOV 1980	DEC 1980
ASSETS												
CASH	5118.68	5221.85	5325.48	5431.99	5540.62	5651.44	5764.47	5879.76	5997.31	6117.14	6239.17	6363.38
RECEIVABLES	689.62	703.41	717.48	731.81	746.46	761.37	776.54	791.97	807.68	823.66	839.91	856.33
INVENTORY	2874.69	2727.97	2782.53	2838.18	2894.94	2952.84	3011.90	3071.14	3131.56	3193.15	3255.91	3319.84
FIXED ASSETS	8289.48	8248.51	8208.54	8168.57	8128.60	8088.63	8048.66	8008.69	7968.72	7928.75	7888.78	7848.81
LIABILITIES												
ACCOUNTS PAYABLE	111.78	111.78	111.78	111.78	111.78	111.78	111.78	111.78	111.78	111.78	111.78	111.78
LONG TERM DEBT	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52	2794.52
EQUITY	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94	2499.94
NET PROFIT	2147.17	2298.11	2435.92	2584.63	2736.33	2891.85	3049.18	3208.35	3369.38	3532.28	3697.05	3863.68

Radio Shack's New VisiCalc™ Program is The Planning Tool for Scientists, Engineers and Managers!

Use It with TRS-80™ Computer Models I, II, and III

Designed for the Decision Maker, VisiCalc is so effective that if no other computer program existed, VisiCalc alone would justify your owning a TRS-80. It's an all-purpose planning and modeling program that dramatically changes the way you make calculations. Ideal for scientists, chemists, engineers, even businessmen — anyone wanting to solve, examine and revise complex, formula-oriented computations with just a few simple keystrokes!

The "Electronic Worksheet." Think of VisiCalc as a giant columnar pad divided into as many as 63 columns and 254 rows (there are some restrictions depending on amount of available memory). This huge worksheet with its built-in calculator easily adapts to your special numerical problems — big or small. Simply insert the numbers and the formulas needed to perform the desired calculations. Then VisiCalc automatically performs all computations and instantly displays the results!

Here's Where the Magic Begins. Suppose you have a set of complex design equations. VisiCalc lets you define and store them, then see the effect of chang-



ing any variable. With every change, VisiCalc instantly recalculates all affected parameters updating the entire worksheet. This makes "What If..." forecasting a snap. You can predict and analyze the result of changing circumstances more quickly and accurately than ever before possible! Built-in transcendental, statistical and special functions, plus 11-digit accuracy and scientific notation make VisiCalc perfect for a wide range of engineering and scientific applications. Special screen formats allow you to specify your requirements for accuracy, and its graphing capabilities are ideal for easy analysis of variables.

Amazing Versatility. VisiCalc is a powerful and versatile tool for everything from engineering analysis to interpretation of statistical results to cost projections — and hundreds of other time-consuming chores. The possible applications are almost limitless. In fact, "special" studies can be developed since

the computations can be saved to disk for access by your computer programs. What's more, you can easily print specified portions of your worksheet with an optional printer. You get the speed, storage, and printout capabilities of the computer — without the programming work!

And It's So Affordable. VisiCalc programs are priced from just \$99.95 to \$299, and are available for all TRS-80 computer systems (starting as low as \$2495). Available at Radio Shack stores, dealers, and Computer Centers nationwide. Leasing and credit plans available. Come see what VisiCalc can do for you today!

Radio Shack® The biggest name in little computers™

Retail prices may vary at individual stores and dealers. Special order may be required initially.

™VisiCalc is a Trademark of Personal Software, Inc.

**Send For More Information on
VisiCalc or TRS-80 Computers**

**Radio Shack Dept. 81-A-288
1300 One Tandy Center
Fort Worth, Texas 76102**

NAME _____

STREET _____

CITY _____ STATE _____

ZIP _____ PHONE _____

RADIO SHACK DIVISION OF TANDY CORPORATION • 6100 STORES AND DEALERS, 123 COMPUTER CENTERS AND 135 SERVICE CENTERS NATIONWIDE

fying explanations of their design have been advanced, notably by Lionel Casson of New York University. In his view double-manning the oars was carried to its completion in the design of the six, which would logically have had its lowest bank laid out in this way. At that point the wider hull became inescapable, whether or not individual ship-builders or city traditions had adjusted their designs in such a direction earlier. From the six onward some new oaring arrangement had to be devised. By putting even more men on each oar, up to a maximum of about eight, larger equivalent bank numbers could be obtained. (Eight was found to be the largest useful number when such rowing arrangements came into fashion again in the Renaissance.)

With more than two men per oar the rowers could no longer operate while seated. Their boat was provided with benches, but to begin a stroke the rowers had to rise, step forward and (if the oar was very long) step up on a kind of stool to dip the outboard end of the oar into the water. To complete the stroke they stepped backward and finished by throwing themselves down on the bench. To judge from later European experience such arrangements made it possible for a crew to consist of fewer skilled rowers, the inboard men being the most critical. The problems of rower coordination rose regardless, and the greater physical exertion, partly made necessary by oars now as long as 17.5 meters, doubtless meant that ship performance dropped even more than the widened hull alone would suggest.

Once a multiple-rower arrangement has been assumed, bank numbers take on a greater degree of uncertainty than they had before. A four could now be a trireme with its top bank double-manned, a bireme with both banks doubled or even a single-banked ship with four men per oar. Demetrius' 16 similar-

ly could have had two banks of eight men each, or three banks adding to 16 in some combination, such as six men each on the top two tiers and four on the bottom. As it happened, there were reasons other than stability for preferring the wider, shallower options.

Beyond Demetrius' 16 there are indications that bank numbers changed their meaning for yet another reason. Witness the fact that this craft was challenged by an eight built by Lysimachus. Casson has adduced a number of reasons for believing this eight and all other galleys rated at more than 16 were built as catamarans, with twin hulls connected by struts. In Lysimachus' ship perhaps two decks of rowers labored in each hull, with four men at each oar. The arrangement suffices to account for all the larger classes of oared warships, including the 40, which had three banks in each hull, arranged perhaps with eight, seven and five rowers respectively in descending order, giving for each side of a hull a total of 20 rowers. The more generous width and depth of the hulls of such ships made it possible to build them longer than triremes. The 40 had hulls 128 meters long and could hold 4,000 rowers.

Most of the reports of the performance of such giant galleys suggest that they were stable but slow. It seems reasonable, from the results obtained above for the quinqueremes, to assume that their performance levels declined steadily as the bank numbers rose above five. A rough estimate indicates that the ratio of their power to their displacement was as low as a sixth of the corresponding value for triremes. Why, then, if the earlier period had been dominated by the trireme and its ram, did the age of Alexander and his successors turn to ships with lower speed and inferior maneuverability?

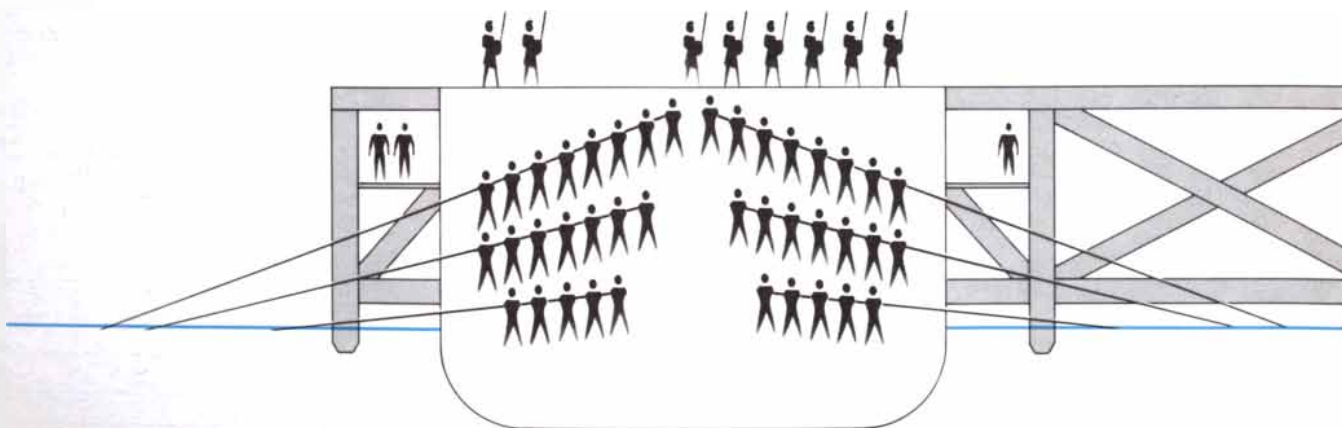
A partial answer is that ramming declined in popularity and was replaced by grappling and boarding. The Romans in

particular put as many as 120 marines on the decks of their quinqueremes. This number implies, and surviving illustrations confirm, that their ships were broad-beamed and only double- or single-banked. The giant catamaran galleys carried considerably more troops. Lysimachus' eight held 1,200 marines on its deck, and the 40 was once loaded with 2,850 marines and 400 deckhands. Clearly in a ship-to-ship encounter decided by the marines one of these supergalleys could overwhelm any number of triremes, each carrying anywhere from 15 to 30 soldiers. The spatial facts indicate that it would have been impossible to surround such a ship with enough triremes to carry a force equal to the catamaran's.

On the other hand, this argument pays insufficient attention to the ram. Unless such a supergalley could be surrounded by a screen of smaller ships, the blows of a single trireme might be enough to sink it. And if its offensive capabilities were limited by low speed, why were such ships built and in increasingly large sizes? Clearly the shift to slower ships and boarding actions implies that some way of neutralizing the ram had been found by about 330 B.C.

The ancient catapult appears to supply much of the answer to the puzzle. It has long been noted that the heavier galleys carried catapults, and it is known that the most powerful and developed form of the weapon, relying on the elastic force of sinew ropes rather than conventional flexible bows, first made its appearance in about 332 B.C. At the siege of Tyre in that year Alexander mounted some of his heavy catapults on ships to pound the city walls. Later Demetrius, who increased the rank of galleys from seven to 16, also had ship-board catapults.

Freed of the constraints of overland travel, such catapults could reach large dimensions. Archimedes built a



4,000 ROWERS manned the oars of this monstrous twin-hulled warship built under the direction of Ptolemy IV in Alexandria near the end of the third century B.C. According to the description attributed to the Greek Callixenus, preserved in the writings of Athenaeus

and Plutarch, the *tessarakonteres*, or 40, had two hulls and was 280 cubits (128 meters) long; it was once loaded with 2,850 troops and 400 deckhands. In this reconstruction by the authors, following a scheme worked out by Lionel Casson of New York University, eight

shipboard catapult that fired stones weighing 78.5 kilograms or darts 5.5 meters long. The latter were doubtless made merely by lopping the limbs from medium-size trees and tipping the trunks with iron points. The machine's range with either projectile was just under 200 meters, but smaller devices could reach more than twice as far. Most of the attention directed to the effect of such projectiles on ships has centered on the stone ball. It has been doubted, with good reason, that these could hull a galley and sink it. Even if balls could penetrate the upper works, the layer of sand or stones used as ballast would halt the missile effectively.

This view, however, ignores the vulnerability of the marines and the rowers to catapult fire. For such an attack arrows are as good as stones, particularly since a new form of arrowhead had become established in Greece in about 500 B.C. Pyramidal in form, it had great penetrating power. Julius Caesar wrote of catapult arrows able to penetrate nearly a foot of solid oak. Ancient galleys certainly offered much less resistance. Theophrastus reports that the top decks of triremes were made of limewood, which has modest mechanical strength but is one of the lightest suitable woods available in the Mediterranean basin.

Our colleague James F. Doyle of the School of Aeronautical and Astronautical Engineering at Purdue University has tested a modern replica of an ancient pyramidal arrowhead of a size suitable for firing from the smallest of the sinew-powered catapults whose existence can be certified from ancient ammunition dumps. He finds that the arrow could penetrate five centimeters of basswood, the American variety of limewood, with an energy equivalent to less than half of the velocity it could be expected to have when it was launched. This thickness, if anything, is too conservative, and the machine in question developed much less than Archi-

medes' shipboard catapult. Even if galley decks had been made of oak and had been as thick as the hull planks, larger ships could mount multiple catapults, firing bolts of ample penetrating power, and still carry at least half of their normal contingent of marines.

The accuracy requirements of such shooting do not appear to have been excessive. In the Middle Ages, English handbow archers practiced shooting at the "clout," a piece of cloth pegged to the ground, at ranges of more than 200 meters. The precise size of the clout is not known, but its overall diameter was certainly less than the roughly five-meter width of a trireme deck. Rolfe Smith, president of the National Crossbowmen of the U.S.A., reports that he is able with modern equipment, but firing offhand, to group his bolts within a radius of three meters at a range of 365 meters. Since ancient catapults of the larger sizes were mounted on adjustable pedestals, they were probably able to do almost as well. Aiming would be facilitated by the visibility of the larger arrows in flight.

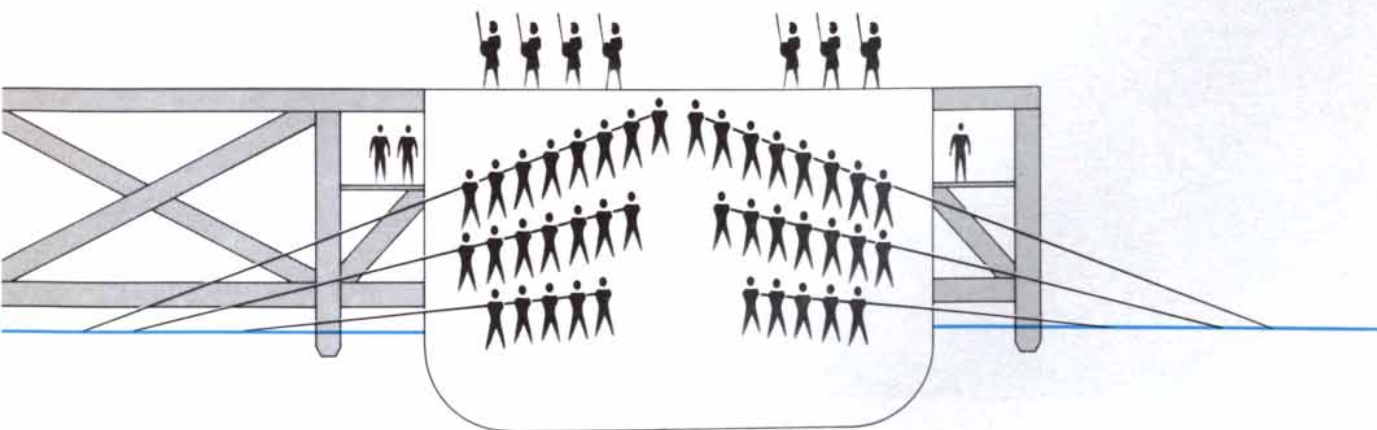
Once through the top deck of a trireme, a catapult bolt stood an excellent chance of striking a rower in the close-packed array that characterized the design. In fact, the echelon arrangement of rowers meant that bolts sufficiently large and long might strike more than one rower at a time, the incoming angle when the bolts were fired at long range being about right for this. With only one rower per oar, even a single casualty could throw an entire side out of beat for a few seconds, particularly if the stricken man was on the most vulnerable top bank and dropped his oar down among the others. Even if no one was injured, the sudden presence of a large arrow shaft in the crowded rowing space would very likely interrupt someone's stroke until it could be chopped away, with much the same result. A de-

lay of a few seconds would be enough, since a trireme could travel its own length in less than six seconds.

Accordingly there seems to have been considerable motivation to alter ship design in order to minimize the danger from catapult fire. Replacing single-man oars with multiply rowed sweeps would reduce the likelihood of the rowers' getting out of beat if a single man was wounded. The combined strength of six or eight men might well suffice to snap the shaft of a javelin-size arrow fired into their midst. Fewer oar banks would present a much less dense target to the projectiles. Hence from this vantage the main trends in galley design after the time of Alexander are justified.

Considerations of stability also help to explain the shift away from ramming. For a number of reasons ramming was best done by striking the enemy on his broadside or quarter. Therefore a ship caught standing in the water broadside to a galley attacking with a ram would be at its most vulnerable. When the ships were employing catapults, however, the tactical situation might have been nearly reversed. The stationary ship would be launching broadsides fired at a right angle to its axis. Since the ship had marginal stability, its aim would be subject to disturbance from rolling, which would cause variations in the elevation of the launched projectiles. Elevation scattering in turn would lead to variations in distance to impact. The great length of the galley would minimize its pitching, which would translate into windage errors in its shooting. Because the oncoming ship was approximately 10 times as long as it was wide, however, it would have presented a target almost perfectly adapted to the firing characteristics of the defending ship.

The attacking galley, on the other hand, would have to fire its catapults along a line virtually dead ahead. It would not pitch much either, but it



men are assigned to each oar in the four top banks, seven in the middle banks and five in the bottom banks. Given this cross-sectional arrangement, each longitudinal row of oarsmen would have had to be 100 men deep to add up to the total of 4,000. The length of the long-

est oars is given in the ancient texts as 38 cubits (17.5 meters). The great stability of the catamaran made its deck ideal for large catapults. There is no evidence, however, that this particular ship ever saw combat. Indeed, it was apparently intended only for display.

could not afford to because its target would be only about three meters wide, excluding the oars. For the attacker rolling would constitute a severe problem, because in elevating its catapults to give them sufficient range it would raise the heads of their projectiles farther from the center of rotation of the ship than their feathered ends.

It follows that as the ship rolled, the projectile heads would be carried farther to the side than the ends. If aiming was done when the ship was vertical, which would be the only time it could be done easily, and firing was done when the ship had passed the vertical, the bolts would fly wide of the target. The ship's motion would be fastest in the middle of its roll, so that determining the precise vertical would be difficult. For a catapult of the size assumed above, an error of only 1.5 degrees from the vertical would be sufficient to result in a complete miss when a bolt was fired at an oncoming trireme from a distance of 200 meters.

Since we assumed a target ship standing broadside to an attacker, this windage problem might not be thought to be severe. The ship's roll would also affect the elevation of the catapult to a small degree, however, as would the slight lurching of the ship as it was

rowed. In spite of the great length-to-width ratio of the galley the bow would still be brought up slightly during the stroke, and it would subside afterward. Even on a smooth sea these irregularities could be crucial in firing at a shallow target at long range. Furthermore, if the target ship were dead in the water, injuries to its rowers would be less crucial than they would be for a ship that was under way.

Thus on balance it seems that catapults were able to neutralize at least part of the threat presented by rams and bring boarding tactics back into favor. Catapults appeared at the right time to explain the observed changes in rowing organization; their performance levels were sufficient to injure rowers if not to sink ships, and their firing characteristics from shipboard platforms were of a kind that would discourage ramming. The enhanced stability of catamarans made them particularly adaptable as catapult platforms. It remains to be explained why after about 250 B.C. such large warships faded from the scene again, to be replaced by triremes or even smaller ships.

Part of the reason appears to be tactical. The sluggishness of the bigger galleys seems to have encouraged the evolution of interfering tactics by smaller and faster ships. At the battle of Chios in

201 B.C., for example, small open vessels called *lemboi* were effective in harassing the heavier ships of Rhodes. After the initial charge, when the bigger ships lost formation, the *lemboi* crept in among them, damaged or interfered with their oars and disrupted their steering. It was a design weakness of catapults that they could not ordinarily be fired at angles below the horizontal, which meant that a zone of partial safety surrounded a large galley for those ships small enough to creep in close. Unfortunately there appears to be no record of small craft darting in between the twin hulls of catamarans, partly protected by the overhanging outriggers, and snapping off oars on both sides as they slid through. It is an exciting idea, but at least some of the catamarans had ram assemblies that stretched between the hulls, obstructing entry there.

Part of the reason for the decline of the heavy galley can be traced in the larger political events of the day. As Rome expanded and shut down the rivalries of the successors of Alexander, its navy was increasingly limited to jobs that required only smaller craft, such as the suppression of piracy. In the final great naval battle of antiquity, at Actium in 31 B.C., the heavy ships of Antony were defeated by the lighter and speedier ones of Octavian's admiral Agrippa.



TYPICAL ROMAN GALLEY, possibly a quinquereme, is seen in this detail of a mosaic dating from the early first century B.C. The mosaic was photographed by Casson in the Palazzo Barberini at Palustrina near Rome. The Romans broke with the Greek preference for speedy warships. Instead, following the Carthaginians, they de-

veloped broad-beamed galleys that could carry larger numbers of troops. Compared with the classical Athenian trireme, which had carried only 14 marines, Roman quinqueremes contained as many as 120. In the latter ships the outrigger seems to have been enlarged and strengthened in order to support all the oars, not just the top bank.

The latter had added several tactical touches that helped his craft, such as grappling irons fired from catapults and pots of fire, some of which appear to have been delivered in the same way. The size and complexity of the new empires meant also that their navies had to be more versatile, in order to carry out such duties as blockading and patrolling. The increased cost of the heavy galleys in both men and materials also worked against them in the imperial age. Even Rome had to minimize its military expenditures.

After Rome's decline began shipbuilding lost even more of its emphasis on a sophisticated reliance on human muscle power. After about A.D. 325 the trireme is heard of no more. The standard oared warship of the late Roman Empire and Byzantine periods, the dromon, relied heavily on Greek fire for offense. This early incendiary agent could be pumped through a kind of flamethrower or contained in catapult-launched pots. Dromons usually had two banks, each oar being manned by one, two or three oarsmen.

The next major galley innovation came early in the 14th century, when the Italians introduced the *a zenzile* rowing system, which featured oars in groups of three each worked by one man, as in the trireme. The novelty lay in the seating of all three men side by side on a common bench, placed slantwise in the boat to enable them to work without interfering with one another. Probably the chief reason for the new system was the lower center of gravity and the greater stability it gave compared with two-banked ships. The compass had recently made its way into Europe, and therefore sailing in rougher weather could be done with more confidence; galley escorts had to be able to venture out in conditions that would have been impossible for their Greek forebears.

The final change in galley design came in about 1550, when the need to carry heavy cannon forced a return to the wide-hulled galley with multiple rowers on each oar. Although the guns on such vessels were mounted to fire only straight ahead, their weight entailed many of the same changes associated with catapults almost 2,000 years earlier. The future, however, belonged to warships that could fire heavy broadsides without the interfering presence of oars and their rowers along the sides of the hull. After the battle of Lepanto in 1571 the role of the galley declined, and after a last-ditch stand in shallow seas such as the Baltic they disappeared altogether. By the time of their eclipse galleys had endured on the seas of the civilized Western world for almost two and a half millennia, a remarkable survival of the direct use of human muscle power in an increasingly machine-dominated culture.

"Science writing at its best!"
 "Astrophysicist Gribbin dynamically interweaves the latest findings on the origins of the universe, our solar system, and our species. He writes as one deeply aware that the universe is alive with wonder, in a style at once precise and lyrical."
 —*Publishers Weekly*
 With 50 photos; a Book-of-the-Month Club Alternate; \$13.95 at bookstores.

GENESIS
 THE ORIGINS OF
 MAN AND THE UNIVERSE
 JOHN GRIBBIN

Delacorte Press/Eleanor Friede
 DELL PUBLISHING CO., INC.

Saturn!



POSTERS • SLIDES

Voyager has done it again! We now have two beautiful full-color posters of everyone's favorite planet, Saturn! One shows only Saturn; the other is a montage showing the ringed planet and six of its moons.

We are also offering a set of ten slides from the Voyager-Saturn encounter.

Enclosed is \$4.50 prepaid for both posters.
 Enclosed is \$5.00 prepaid for set of 10 slides.

Add \$1.00 for our complete catalogs
 Foreign orders add \$1.00 postage

DEPT. SA
Hansen Planetarium
 15 South State Street
 Salt Lake City, Utah 84111

To preserve your copies of **SCIENTIFIC AMERICAN**

A choice of handsome and durable library files—or binders—for your copies of **SCIENTIFIC AMERICAN**. Both styles bound in dark green library fabric stamped in gold leaf.

Files Each file holds 12 issues. Price per file \$5.95; three for \$17.00; six for \$30.00, postpaid. (Add \$2.50 each outside U.S.A.)

Binders Each binder holds 12 issues. Issues open flat. Price per binder \$7.50; three for \$21.75; six for \$42.00, postpaid.

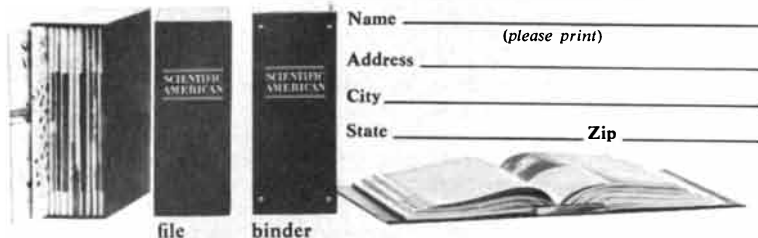
(Add \$2.50 each outside U.S.A.)

To: Jesse Jones Box Corp., P.O. Box 5120, Philadelphia, Pa. 19141

I enclose my check or money order for \$_____ (U. S. funds only).

Send me _____ **SCIENTIFIC AMERICAN** Files. Binders.

Name _____ (please print)
 Address _____
 City _____
 State _____ Zip _____



file binder

NOTE: Satisfaction guaranteed or money refunded. Allow four to six weeks for delivery.

Lithium and Mania

How is it that salts of lithium have a beneficial effect on people in a pathologically manic state? Clues to the answer may be found in the ways the lithium ion moves through the membranes of cells

by Daniel C. Tosteson

The word mania applies loosely to many kinds of excitement, but its clinical meaning is more specific. There the term describes an illness in which the patient is mentally and physically hyperactive and has an elevated mood and disorganized behavior. Often the illness has an obverse side, depression, and the patient's mood swings periodically from one extreme to the other, a syndrome usually described as a bipolar disorder. People suffering from mania or bipolar disorders often improve impressively when they are treated with salts of lithium.

These salts, like most others, dissociate into positively and negatively charged ions when they are dissolved in water. The distribution and many of the actions of lithium ions in the human body involve interactions with the membranes that envelop cells and intracellular organelles. Indeed, the transmission of impulses between and along nerve cells in the brain depends on the regulated movement of ions across the membranes of those cells. It has recently become clear that some patients with mania have a defect in a pathway for the transport of lithium across the membranes of the red cells of the blood. A plausible conjecture is that the condition also affects brain cells. I shall set forth in this article what is now known about the relations among lithium, membranes and mania. I shall also venture some speculations about the significance of this knowledge for deepening human understanding of the physicochemical basis of the human mind.

The story begins in 1949, when a brief report by J. F. J. Cade appeared in *The Medical Journal of Australia* telling of a marked improvement of patients with acute and chronic mania who had been treated with lithium salts. At that time Cade, a physician and psychiatrist, was working alone at a small hospital for chronic mental patients and had not had any previous experience in research. His interest was in the conditions that give rise to such affective disorders as mania and depression. He reasoned that the manic state might be initiated and

sustained by a substance that could be found in the body fluids of people who had the disorder.

Cade began a search for such a hypothetical compound by injecting urine from patients with mania and depression and from normal people into the peritoneal cavity of guinea pigs. Because uric acid and its salts can be present in high concentration in urine he chose to investigate the effect of injecting uric acid into the animals. And because uric acid and many of its salts are not readily soluble in water he worked with lithium urate, which is more soluble. In order to control for any effects of lithium as opposed to urate he injected a solution of lithium carbonate into other guinea pigs. To his surprise he found that after a latent period of about two hours the animals became extremely lethargic and unresponsive, remaining in that condition for an hour or two before again becoming timid and active as they are normally. "Those who have experimented with guinea pigs," Cade later wrote, "know to what extent a startle reaction is part of their makeup. It was thus even more startling to the experimenter to find that after an injection of a solution of lithium carbonate they could be turned on their backs and that, instead of their usual frantic righting behavior, they merely lay there and gazed placidly back at him."

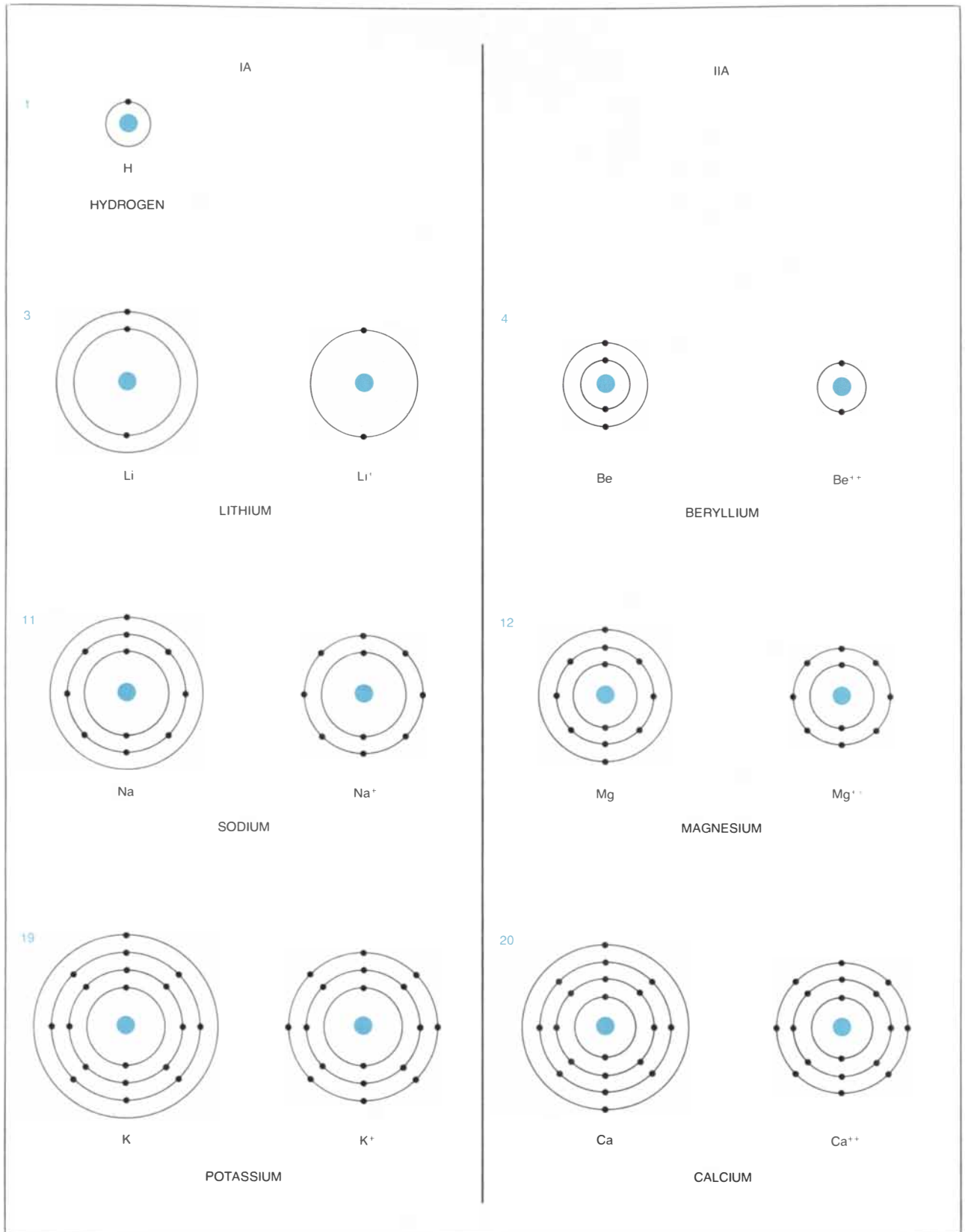
Cade speculated that the calming effect of lithium on guinea pigs might be repeated in human beings. He therefore began a study of the effect of administering lithium salts orally to a 51-year-old man who had been hospitalized with chronic mania for five years. After five days of treatment the patient showed a definite improvement; within two months he was able to leave the hospital and resume his job while remaining on lithium therapy.

After Cade's initial report the role of lithium therapy in mental disorders was most vigorously explored by Mogens Schou and his associates at the University of Aarhus in Denmark. By the early 1960's lithium was accepted as the

treatment of choice for mania in Australia and most of Europe. Another 10 years passed before lithium was accepted in the U.S. One reason for the long delay was the great caution that many American physicians had about lithium because of disastrous experience with it as a substitute for sodium in the treatment of congestive heart failure in the 1940's. Several patients had died before the toxicity of lithium, which is particularly pronounced in a person deprived of sodium, had become evident. Also contributing to the delay were economic factors. Because the chemical simplicity and ready availability of lithium made it impossible for pharmaceutical companies to develop an exclusive position for the marketing of the drug, they had little incentive to make lithium salts and to carry out all the testing required by the Food and Drug Administration before lithium could be introduced into clinical practice in the U.S.

Lithium is the third-lightest element, following hydrogen and helium in the periodic table of the elements. The atom has three protons (positive charges) in its nucleus and three electrons (negative charges) in orbit around the nucleus. One of the electrons is alone in the outer orbital and so is readily lost to interactions with other atoms. When it is lost, the lithium atom has three positive charges in its nucleus and two negative charges in its electron shell and is hence a cation: a positively charged ion that interacts with anions, or negatively charged ions. In the synthesis of heavier chemical elements out of lighter ones in the universe lithium is not on the main pathway; therefore on the earth it is present in small quantities compared with its chemical relatives sodium, potassium, magnesium and calcium. Nevertheless, it is distributed widely in minerals, largely as silicates.

From the point of view of the physical chemist the striking thing about lithium, sodium and potassium is their similarity. Indeed, it was this similarity that led to their classification in Group IA of the periodic table. From the viewpoint of the physiologist, however, the signifi-



LITHIUM ION is portrayed with other atoms and ions that move through cell membranes. They are presented according to their positions in groups IA and IIA of the periodic table of the elements and also according to their relative sizes. The colored numerals show the number of protons, or positive charges, in the nucleus of each atom; the black dots represent electrons in their orbits around the nucleus.

A lithium atom has two electrons in its first orbital and one in its second: a total of three negative charges balancing the three positive charges of the nucleus. The single electron is readily lost, however, and the atom becomes a cation: a positively charged ion (Li⁺). A number of patients with clinical mania can be restored to normality by treatment with lithium salts, which in body fluids dissociate into ions.

cance of these elements lies in their differences.

The ability of the membranes of the living cell to generate and propagate electrical signals depends on the ability of these membranes to distinguish between sodium and potassium. The two elements must also be recognized and separated by the molecular machine that pumps sodium out of and potassium into nerve and muscle cells and thereby charges the "battery" that drives the electrical signals. The effects of lithium in living organisms, including human beings with affective disorders such as mania, turn on the differences between ions of the element and ions of the other members of groups IA and IIA that are present in cells, particularly sodium, potassium, calcium and magnesium. Biological evolution on the earth apparently has taken advantage of subtle differences among some of the elements that make up the crust of the earth.

The distribution and rates of movement of ions in living organisms are largely regulated by the membranes that separate the inside and the outside of cells and of intracellular organelles such as the mitochondrion. Such a membrane consists of a bilayer of phospholipid and cholesterol molecules in which specialized protein molecules are embedded. The thickness of the membrane, including the proteins, is about 10 nanometers (10 millionths of a millimeter).

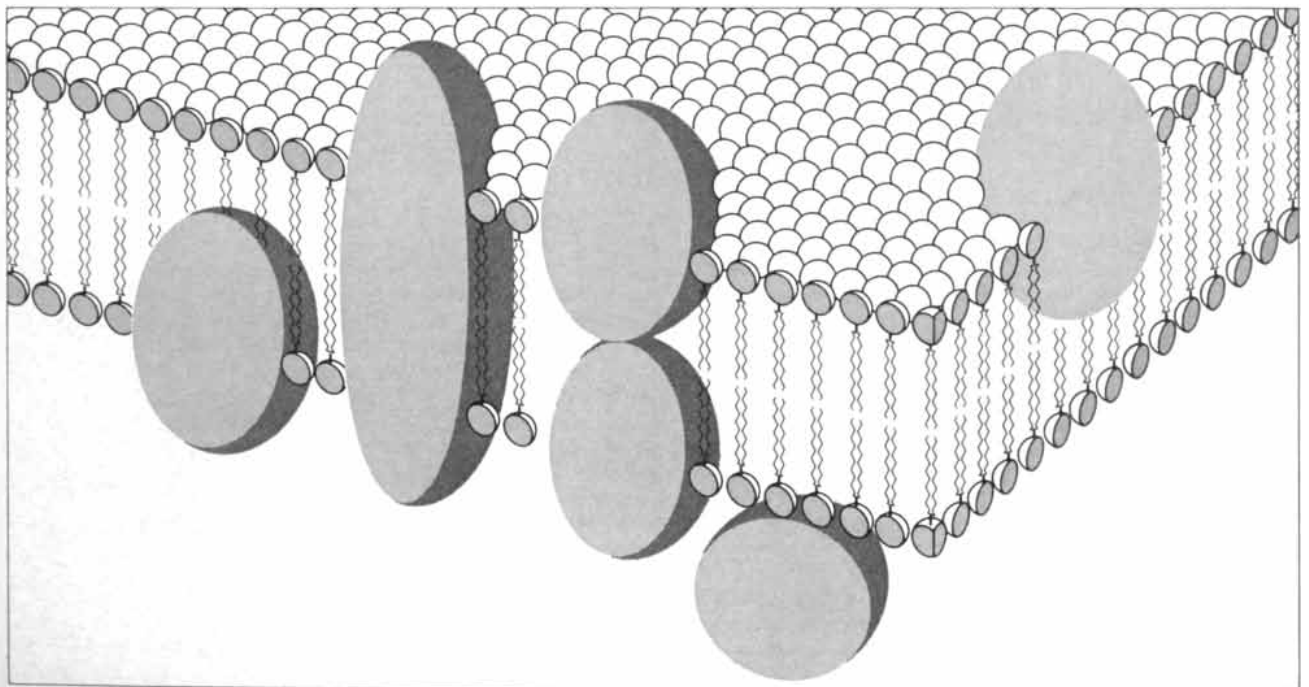
The lipid-bilayer part of these biological membranes readily allows the passage of small, uncharged, relatively lipid-soluble molecules such as those of carbon dioxide and molecular oxygen but not the passage of charged, water-soluble ions such as those of sodium, potassium and lithium. Most if not all of the transport of ions across a biological membrane is mediated by specific proteins embedded in the lipid bilayer. Four types of transport deserve mention. First, protein channels may allow the selective movement of various ions down their respective concentration and electric-potential gradients. One example is provided by the time- and voltage-dependent sodium and potassium channels that are responsible for the propagation of the action potential (the nerve impulse) in nerve and muscle fibers. Another is seen in the channels, activated by acetylcholine, that initiate the action potential in muscle fibers.

Second, there are pumps that convert the energy of chemical bonds, such as the energy stored in the pyrophosphate bond of adenosine triphosphate (ATP), into the work necessary to move ions against concentration and electric-potential gradients. Among the examples are the sodium-potassium pump present in the outer membrane of most cells and the calcium pumps present in the membranes of mitochondria and the sarcoplasmic reticulum.

A third kind of ion transport was discovered by Hans H. Ussing of the Uni-

versity of Copenhagen, who called it exchange diffusion. The process allows an exchange but no net movement of ions through the membrane. Such transport was first observed by Ussing and his colleagues as sodium-sodium exchange in frog muscle. To detect an exchange involving a single species of ion (sodium in Ussing's observation) requires a tracer such as a radioactive isotope of the transported ion. If two different kinds of ion (such as sodium and lithium) can participate in a system of this kind, the exchange proceeds until the ratio of concentrations of the two ionic species is identical in both solutions (inside and outside) bathing the membrane. Chloride ions and bicarbonate ions are exchanged across the membranes of human red blood cells by this kind of exchange mechanism.

A fourth type of molecular movement across biological membranes is cotransport, in which two substances move together. Examples are the simultaneous movement of sodium and sugars or sodium and amino acids into the epithelial cells of the intestine as part of the process of absorption and the coupled cotransport of sodium and potassium in the red cells of birds and in certain other cells. Lithium can substitute for sodium in all these pathways. Under the conditions that prevail in the blood of patients being treated with lithium, however, only one pathway—sodium-lithium exchange or countertransport—moves lithium outward across the red-



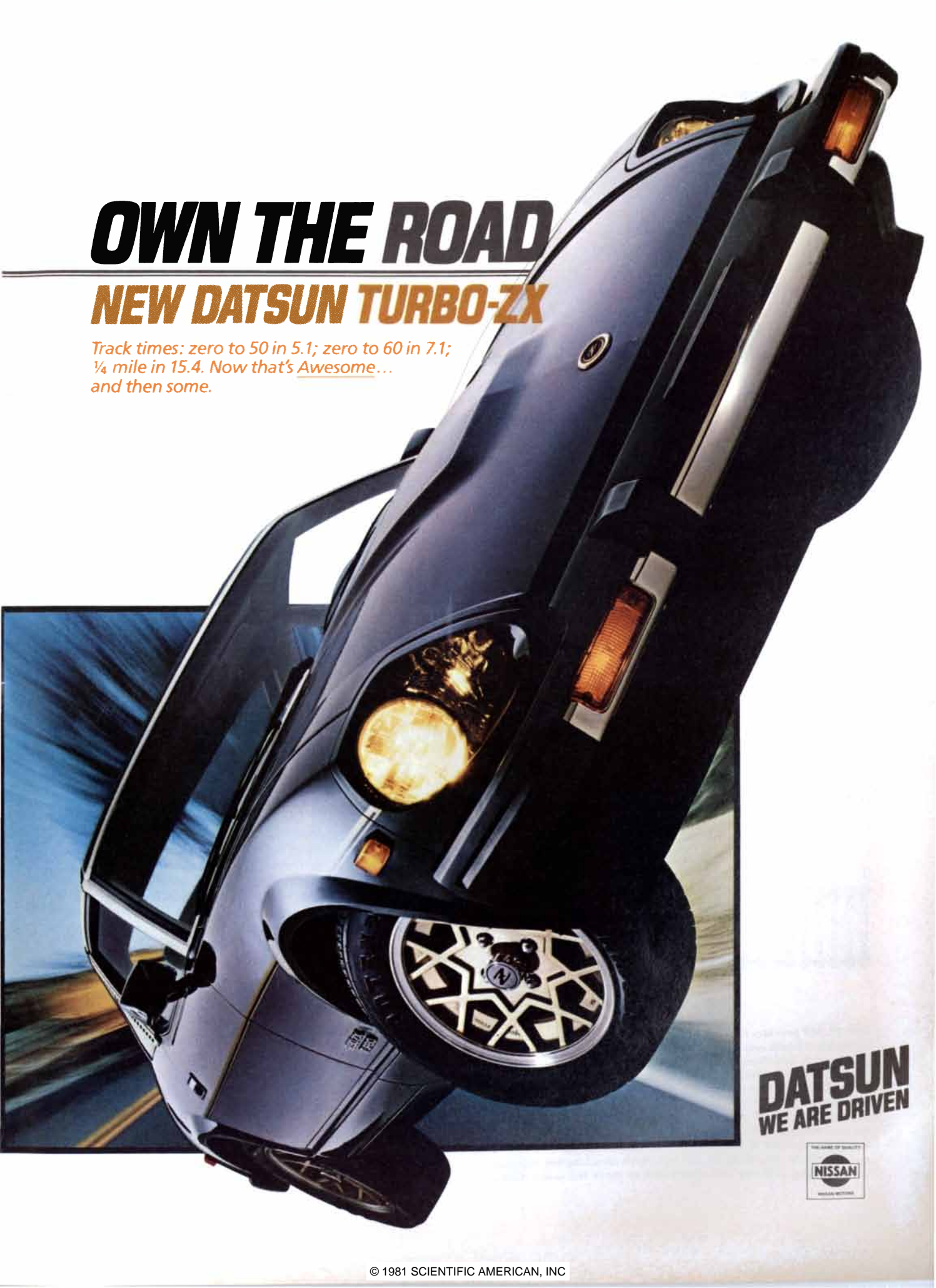
STRUCTURE OF THE CELL MEMBRANE is shown schematically according to current conceptions. The basic structure is a double layer of lipid (fat) molecules with their hydrophilic heads pointing outward and their hydrophobic tails pointing inward. Also associated

with the membrane are protein molecules, seen here as larger bodies some of which penetrate the membrane and others of which are embedded in one side or the other. Most of the movement of ions into and out of the membrane is by way of channels in the various proteins.

OWN THE ROAD

NEW DATSUN TURBO-ZX

Track times: zero to 50 in 5.1; zero to 60 in 7.1;
¼ mile in 15.4. Now that's Awesome...
and then some.



DATSUN
WE ARE DRIVEN



cell membrane against its concentration gradient.

In addition to regulating the distribution of ions and small molecules, membranes carry out many other important cellular functions. Among them are the uptake and release of large molecules and small volumes of fluid; the recognition of extracellular molecules such as

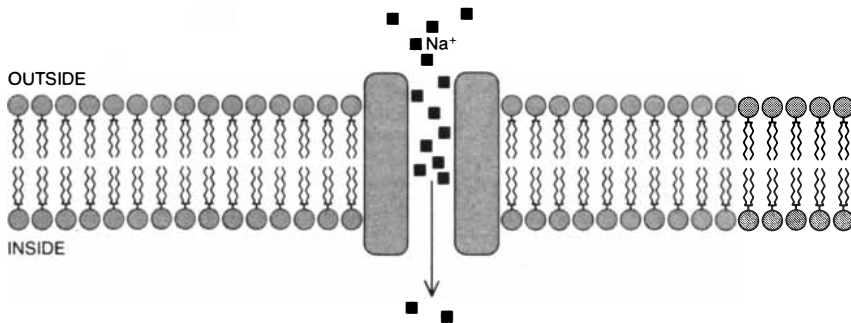
hormones, antigens and antibodies and of other cell surfaces in embryonic development and the repair of organs, and the ordering of the sequence of various biochemical reactions.

Some of the connections among lithium, membranes and mania began to become clearer when Joe Mendels and

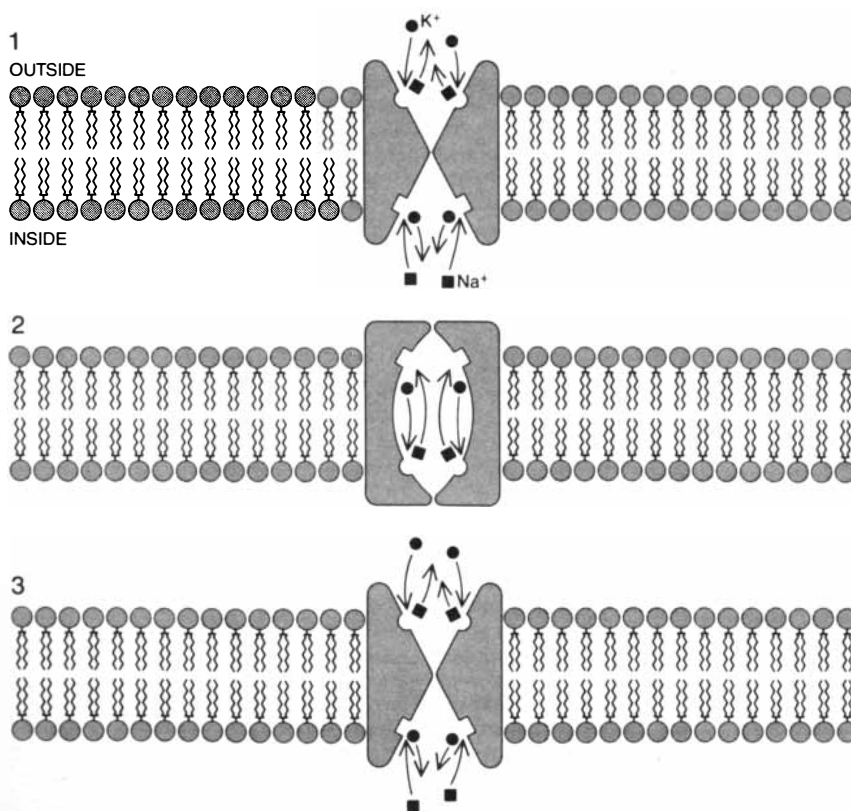
Alan Frazer of the University of Pennsylvania School of Medicine observed that the lithium ratio (the concentration of lithium inside red cells divided by the concentration in the blood plasma outside the cells) showed a considerable variation from person to person among patients who were being given lithium therapeutically. Mendels and Frazer also found that the mean value of the ratio in a group of patients with affective disorders responsive to lithium was significantly higher than it was in other groups of patients and in normal people who served as controls.

Mendels and Frazer also reported that the lithium ratio is higher in sheep red cells with a high concentration of sodium than it is in sheep red cells with a low concentration of sodium. This finding led me to suspect that a countertransport system involving sodium and lithium could be found in the membranes of human red cells. Experiments that Mark Haas, James M. Schooler and I did at the Duke University School of Medicine in 1975 supported the hypothesis. Somewhat later, in collaboration there with G. N. Pandey and John M. Davis and their colleagues, with Balasz Sakardi and R. B. Gunn at the University of Chicago and with J. Funder and J. O. Wieth of the University of Copenhagen, we investigated the transport processes that determine the distribution of lithium between human red cells and blood plasma. Simultaneously and independently Jochen Duhm and his colleagues on the medical faculty of the University of Munich made similar observations that were later confirmed and extended by Barbara E. Ehrlich and Jared M. Diamond of the University of California at Los Angeles School of Medicine. On the basis of this now quite substantial body of evidence the following picture of lithium metabolism in human red cells emerges.

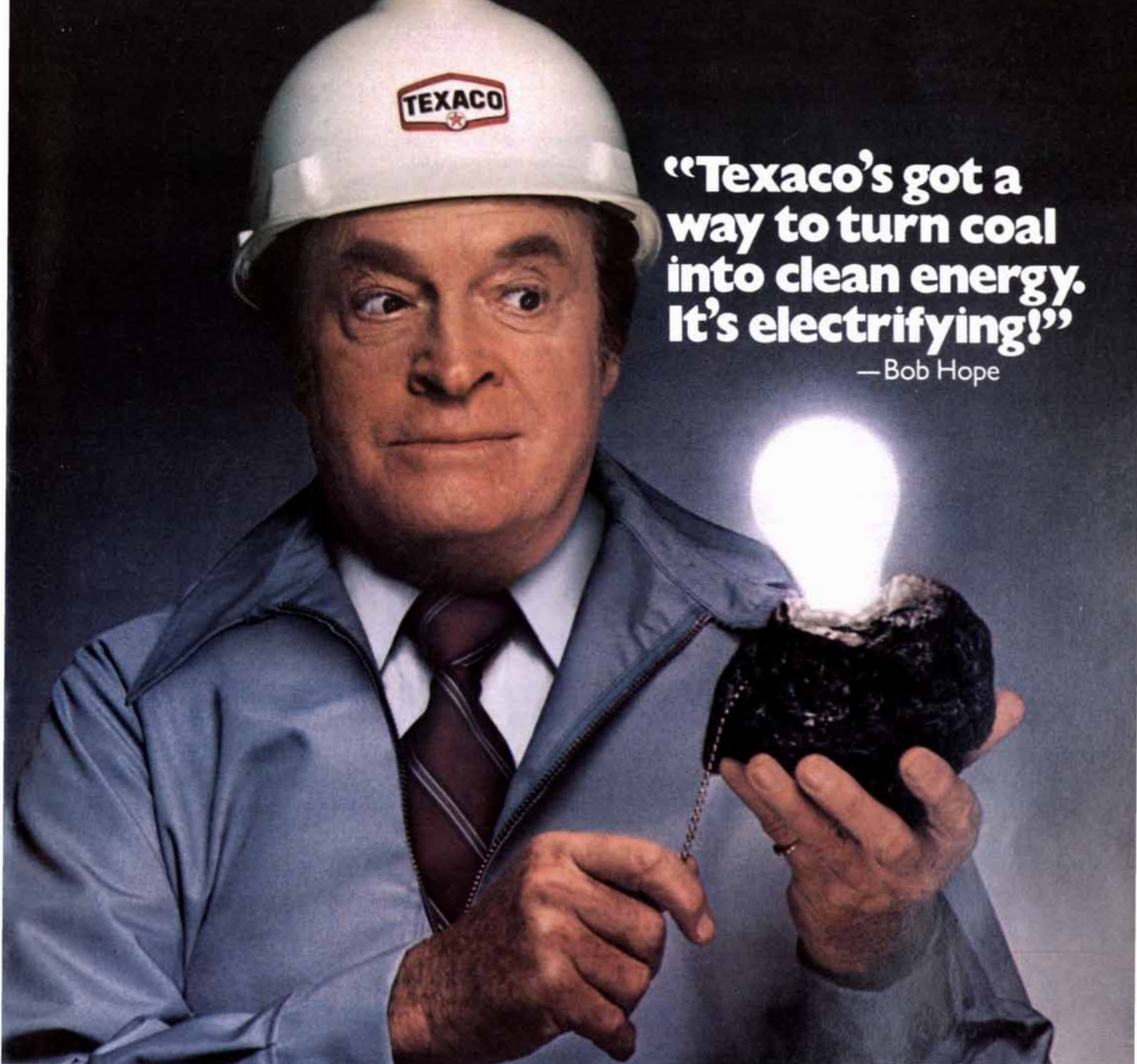
The concentration of lithium in the blood plasma of patients receiving treatment for affective disorders is about one millimole per liter (one millimolar). The steady-state concentration of lithium in the red cells of such patients is the result of a dynamic balance between inward leakage down the lithium concentration gradient and outward countertransport up the gradient. The uphill movement is driven not by the release of energy stored in chemical bonds (as in ATP) but by a dissipation of the internally directed concentration gradient of sodium. The sodium that enters the cell through the lithium-sodium exchange system is extruded by the sodium-potassium pump, which is driven by ATP. The countertransport system therefore has the effect of coupling lithium to the sodium-potassium pump. The steady-state lithium ratio approaches the sodium ratio of .08 when the lithium-sodium countertransport system is very active



PROTEIN CHANNEL through which ions such as sodium, potassium and lithium can move "downhill" involves a penetrating protein. Downhill movement, which can be either into or out of the cell depending on differences in the ion concentration and the electric potential on the two sides of the membrane, does not require metabolic energy. Here sodium ions move inward.



ION PUMP provides another means of moving ions into and out of a cell. It converts the energy of chemical bonds into the work necessary to move ions against their electrochemical-potential gradients. A two-state, four-site speculative mechanism for the sodium-potassium pump begins (1) with a membrane-penetrating protein that is open to the outside and inside of the membrane. On the outside a potassium ion from the blood plasma moves to a binding site vacated by a sodium ion; a reverse exchange takes place on the inside. The configuration of the protein channel changes to a closed state (2), the selectivity of the two sites changes and potassium on the outward-facing site exchanges with sodium on the inward-facing site. The first configuration and selectivity of sites are restored (3), and the cycle is completed. The changes of shape and selectivity are driven by the hydrolysis of adenosine triphosphate (ATP). At no time is there an open pathway that allows ions to move the entire distance through the membrane.



“Texaco’s got a way to turn coal into clean energy. It’s electrifying!”

—Bob Hope

America’s got enough coal to keep your lights on for hundreds of years. And Texaco’s coal gasification process could mean you won’t have to worry about how it affects the environment.

One of the main problems with burning coal to generate electricity has been, of course, to burn it in an environmentally acceptable manner.

But Texaco’s developed a process to turn our most plentiful energy resource into clean-burning fuel gas which can be used to produce power for generating electricity.

We’ve already successfully proven the gasification process in small-scale plants. Now Texaco is working with other companies and organizations to build a large coal gasification/electrical generating plant in the Mojave Desert. It will turn a thousand tons of coal a day into electricity.

Some years from now, Texaco’s investment in coal gasifi-

cation technology can mean you’ll have the electricity you’ll need—and without worrying about environmental effects.



You can trust the Star at home, at work, or in your car.

THE NEW NIKON F3

A masterpiece of pace-setting technology and surpassing precision, the new Nikon F3 has no equal among 35mm slrs. Strikingly handsome, it more than hints at the incredible automatic efficiency beneath its flawless finish. And its capacity for superlative photography is both awesome and inspiring.

Imagine exposures controlled to millisecond accuracy by electronic microprocessors and quartz crystal technology!

Imagine enjoying this kind of accuracy with any of the nearly 70 lenses, 4 interchangeable viewfinders and hundreds of other accessories in the Nikon F3 system!

Imagine all this and much more in a camera built to Nikon professional standards. So rugged and reliable that it has been selected by NASA for use aboard the Space Shuttle. Yet light, compact, and ingeniously designed to give you magnificent pictures – automatically.

But, why only imagine! Let your Nikon dealer put the Nikon F3 in your hands. It's like no other camera on earth!

©1980 Nikon Inc., Garden City, N.Y. 11530.

LIKE NO OTHER CAMERA ON EARTH



and there is little or no leakage. When the leakage pathway is large and the countertransport pathway is small, the lithium ratio approaches 1.2—the value for passive singly charged cations when the electric potential inside the cells is 10 millivolts negative with respect to the outside plasma, as is the case for human red cells. In most people the lithium ratio is between .2 and .6.

The countertransport system carries out a 1:1 exchange of sodium for sodium or of lithium for sodium. Depending on the sodium gradient, the system can move lithium uphill either into or out of the red cells. Since in the body the concentration of sodium inside cells is almost always much lower than the concentration outside them, the countertransport system gives rise to a net outward movement of lithium whenever the lithium ratio is higher than the sodium ratio. The affinity of the system for lithium is about 20 times greater than it is for sodium, as is shown by the fact that the intracellular concentration needed for a half-maximal stimulation of countertransport is .5 millimolar for lithium and 10 millimolar for sodium.

The countertransport system will not accept potassium, ammonium, choline, calcium, magnesium or any other cations so far tested. It is not inhibited by ouabain, which specifically blocks the ATP-driven sodium-potassium pump. It is inhibited by phloretin, a well-known inhibitor of transport systems for sugar and anions. The system does not require ATP or glycolysis (the glucose metabolism that supplies red blood cells with energy).

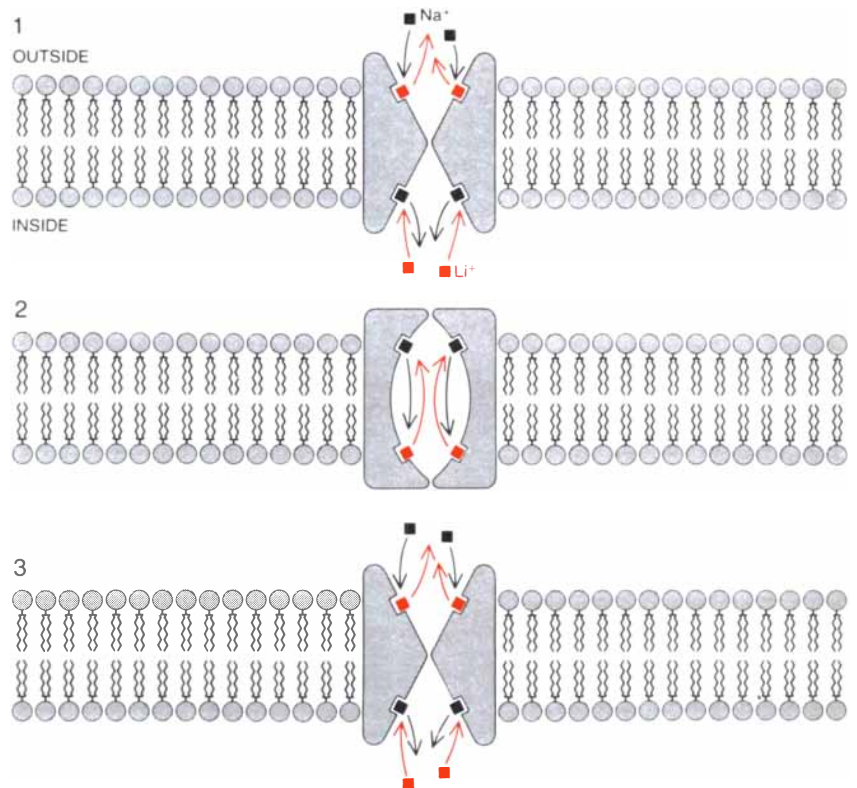
The maximum transport rate of the lithium-sodium countertransport system varies considerably from individual to individual. The range is from zero to .7 millimole per liter of red cells per hour. Pandey and his colleagues and Duhm and his have shown that these variations correlate with and account for the variation between individuals in the red-cell lithium ratio, measured both in patients receiving lithium and in vitro (with red cells incubated for 24 hours in a plasmalike medium containing 1.5 millimolar of lithium). Considerable evidence from studies of twins and other family members, involving both patients with affective disorders and normal control subjects, suggests that the variations among individuals in the maximum rate of lithium-sodium countertransport are (at least in part) genetically determined.

The downhill leakage of lithium in human red cells can take either of two pathways. One pathway is not inhibited by any compounds yet tested and does not appear to involve other ions or uncharged molecules. The other is an interesting pathway first discovered by Funder and Wieth. Lithium ions (and to a lesser extent sodium ions) can react

with carbonate, a doubly charged anion, to form the singly charged ions lithium carbonate and sodium carbonate. These anions can then be transported on the very rapid singly-charged-anion exchange system that normally carries bicarbonate and chloride ions across the membrane of the red cell. Although the concentrations of lithium carbonate and sodium carbonate and therefore their rates of transport are always quite low compared with other anions that travel on the anion-exchange system, the rate of movement of lithium by this pathway is significant compared with its rate of movement by other pathways. In patients with normal concentrations of chloride and bicarbonate in the plasma about half of the inward leakage of lithium into red cells is through the carbonate pathway.

The movement of lithium across the membrane of the human red cell in a person who is receiving lithium therapy is primarily by way of the leakage path-

ways and the lithium-sodium countertransport system. Under certain experimental conditions, however, there are other ways for lithium ions to traverse the membrane of the red cell. For example, lithium competes for both the sodium and the potassium sites in the sodium-potassium pump. When sodium and potassium are not present outside the cell, lithium substitutes for potassium and is actively transported into human red cells; when sodium and potassium are absent inside the cell, lithium substitutes for sodium and is actively transported outward. Since the pump has more affinity for sodium and potassium than it does for lithium, not much lithium goes by this pathway in the body because sodium and potassium are always present on both sides of the membrane (and at much higher concentrations than lithium). We have recently shown that lithium can substitute for sodium in the cotransport of sodium and potassium in human red cells. Still oth-



ION-EXCHANGE OR COUNTERTRANSPORT SYSTEM that is thought to contribute to the movement of lithium ions out of red blood cells in therapy for mania is depicted. The system normally carries out a 1:1 exchange of sodium for sodium, but it can also exchange lithium for sodium. The hypothetical two-state, four-site model shown here resembles the pumping mechanism depicted in the lower illustration on page 168. In the open state (1) the two sites can exchange sodium or lithium ions present in the solutions bathing the inner and outer surfaces of the membrane. In the closed state (2) the two sites can exchange ions with each other but not with the two solutions. A restoration of the initial state completes the cycle (3). The selectivity of the sites is the same in the two states. In the body the concentration of sodium is lower inside cells than it is outside; the countertransport system produces a net outward movement of lithium when the ratio of cell-to-plasma concentration is greater for lithium than it is for sodium. The countertransport system appears to be defective in certain patients with mania, resulting in a higher cell-to-plasma ratio of lithium than is normal.

er pathways may be important for the transport of lithium in other types of cell. Halvor N. Christensen and his colleagues at the University of Michigan Medical School have recently shown that lithium moves together with certain amino acids in some tumor cells of mice.

What, then, are the connections among lithium, membranes and mania? First, the occurrence of extremely low maximum rates of lithium-sodium countertransport in red cells seems to be greater among people suffering from mania or bipolar disorders than it is among normal people or people with other types of mental disorder. Studies of the families of people with and without affective disorders by Elizabeth Dorus and her colleagues at the University of Chicago Medical School suggest that reduced lithium-sodium countertransport appears more often in the first-degree relatives (parent or child) of patients with mania or bipolar disorders than it does in other people. On the basis of this finding it is possible to entertain

the hypothesis that a reduced lithium-sodium countertransport is an inherited disorder and that the genes responsible for the expression are also involved in a predisposition to mania and bipolar disorders. Since most of the people suffering from those affective disorders do not have a reduced lithium-sodium countertransport, it is clear that this hypothesis can apply only to a subgroup of manic and bipolar patients. This is not surprising, because it is well established that an inherited predisposition to affective disorders involves many genes. More extensive studies are needed to elucidate the relation between the inheritability of the lithium-sodium countertransport system and a predisposition to mania and bipolar disease.

If the hypothesis that a reduced lithium-sodium countertransport in red cells and the incidence of mania are positively correlated is correct, what could be the connection between these apparently separate phenomena? It is unlikely that events in the membranes of red cells determine mood. One possibility is that the genes coding for the lithium-sodium

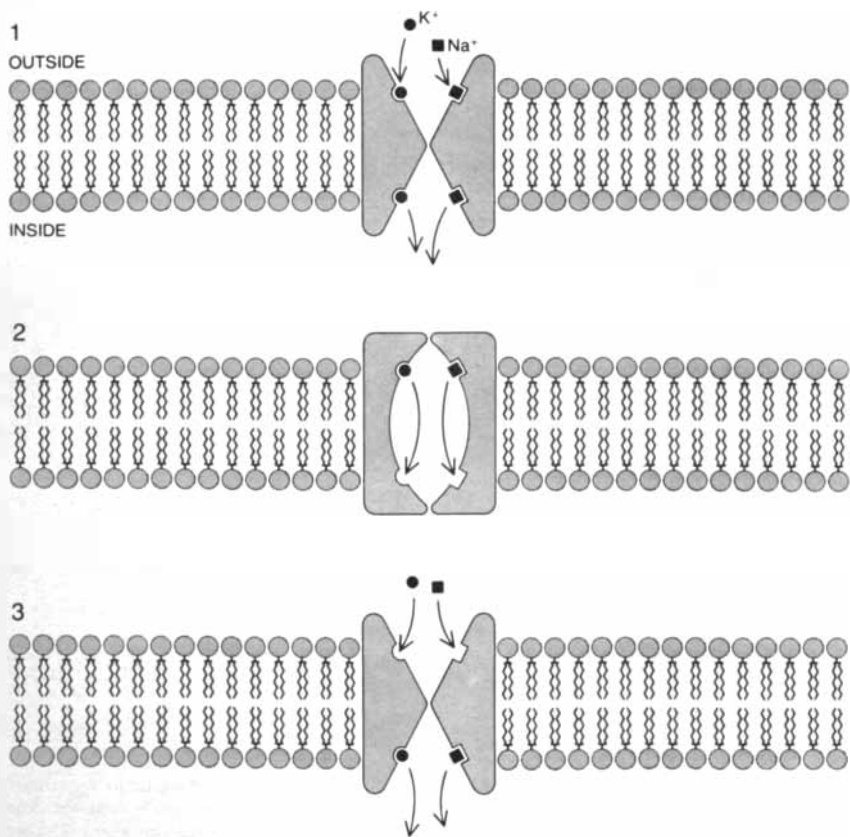
countertransport system also express a similar system in other cells, including cells in the brain.

The ionic selectivity of the sodium-exchange pathway in such brain cells might be slightly different, allowing the exchange of sodium with calcium or with various singly-charged-cation neurotransmitters: substances that transmit impulses from one nerve cell to another and from a nerve cell to a muscle cell. Lithium has been shown to interact with the sodium-calcium exchange system in the plasma membranes of nerve and muscle cells and with the systems that take up neurotransmitters in the brain. An inherited defect in such a transport system in the brain might contribute to the abnormal state of brain and mind called mania.

I should emphasize that any discussion of the relation between the genetic determination of lithium-sodium countertransport and that of mania, or of the mechanism of the therapeutic action of lithium in mania, is highly speculative given the present state of ignorance. It is not known whether mania is the manifestation of an abnormal function of a small group of cells in the brain or whether it is a generalized disorder. It is also not known whether lithium improves mania by changing the function of a small group of brain cells or of many or all of them. The therapeutic action of lithium does not seem to be related to the variations among individuals in the maximum rate of lithium-sodium countertransport in red blood cells; it is equally effective in patients with an active countertransport system and in patients lacking a countertransport system.

Given the physicochemical similarities between lithium ions and the physiologically important, naturally occurring singly charged cations sodium and potassium (and to a lesser extent the doubly charged cations magnesium and calcium), it is not surprising that lithium at sufficiently high concentrations interacts with virtually all components of a living organism: proteins (including enzymes), nucleic acids and the charged groups of atoms on lipids and carbohydrates. It is also not surprising that the list of reported biological effects of lithium is quite long.

Even if attention is limited to the effects of concentrations of lithium in the therapeutic range, there are many promising avenues for the exploration of the mechanism of the ion's action in mania. Among the more attractive ones for me are (1) its effects on the synthesis of prostaglandins, (2) its effects on the production of white blood cells, (3) its effects on the transport of neurotransmitters and the sensitivity of receptors to them, (4) its inhibition of the choline transport system during the maturation of human red cells, (5) its action of in-



COTRANSPORT SYSTEM moves two kinds of ions simultaneously into or out of a membrane. In this hypothetical two-state, four-site model sodium and potassium ions move downhill into a red blood cell. With the membrane-penetrating protein open to the outer and inner surfaces of the membrane (1) a sodium ion and a potassium ion move from the blood plasma to binding sites near the outside opening of the protein and two similar ions move from interior binding sites to the inside of the cell (3) as the cycle starts again. In contrast to the models depicted for the sodium-potassium pump and for sodium-sodium exchange, sites can be empty in the closed state of this system.

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading. It's easy to learn...lasts a lifetime...applies to everything you read...and is the only accredited course with the option of college or continuing education 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 any-

thing 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, explains 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.

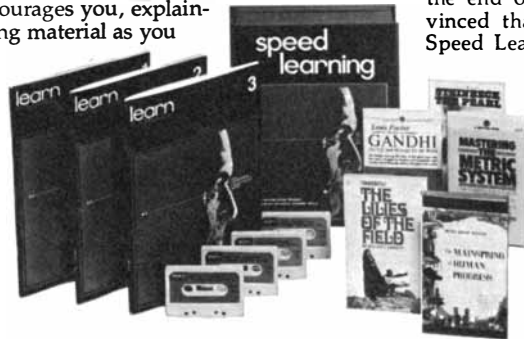
Execute Speed Learning FREE for 15 days

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 15 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.



COLLEGE CREDITS

You may obtain 2 full semester hour credits for course completion, wherever you reside. Credits offered through Whittier College (California). Details included in your program.

CONTINUING EDUCATION UNITS

National Management Association, the world's largest association of professional managers, awards 3.0 CEU's for course completion. CEU's can be applied toward the certificate in Management Studies.

PROFESSIONAL SOCIETIES

Speed Learning is offered internationally to members of professional associations such as: American Chemical Society, Foundation for Accounting Education, Institute of Electrical and Electronics Engineers and dozens more. Consult your Education Director for information.

BUSINESS, INDUSTRY, GOVERNMENT

Many companies and government agencies offer *Speed Learning* as a wholly-paid or tuition reimbursement program. Consult your Training or Personnel Director for details.

learn
INCORPORATED

113 Gaither Drive, Mt. Laurel, NJ 08054

31SAE

YES! Please rush me the materials checked below:

- Please send the *Speed Learning* program @ \$99.95 plus \$4 postage and handling.
- Please send the *Speed Learning Medical Edition* @ \$109.95 plus \$4 postage and handling.
- Please send the *Junior Speed Learning* program (ages 11 to 16) @ \$89.95 plus \$4 postage and handling.

Check method of payment below:

- Check or money order enclosed (payable to learn incorporated)
 - Charge my credit card under the regular payment terms
 - Visa Master Card Interbank No. American Express
- Card No. _____ Exp. Date _____

I understand that if after 15 days I am not delighted in every way, that I may return the materials in their original condition for a full refund. No questions asked.

Name _____
Address _____
City _____ State _____ Zip _____
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 USA add \$10 per item—Airmail extra —

creasing the concentration of choline in red cells, as recently reported by Ehrlich and Diamond, and (6) its inhibition of the activation of adenylate cyclase (the enzyme that catalyzes the action of intracellular "messenger" adenosine monophosphate, or AMP) by catecholamines, thyroid-stimulating hormone and antidiuretic hormone.

In view of the many sites of action of lithium, it is not surprising that it has many toxic effects. The safe range

of therapeutic doses is fairly narrow. Among the more serious complications of treatment with lithium is that it can cause the kidneys to fail to respond to the antidiuretic hormone, giving rise to the disorder diabetes insipidus.

What does this incomplete chapter on lithium, membranes and mania portend for the future of medicine? It is another example of the increasing precision with which the physicochemical di-

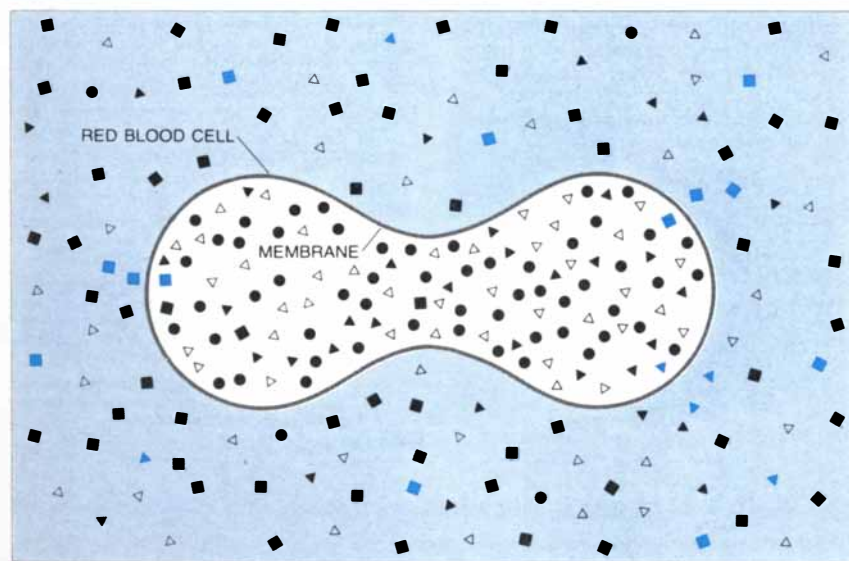
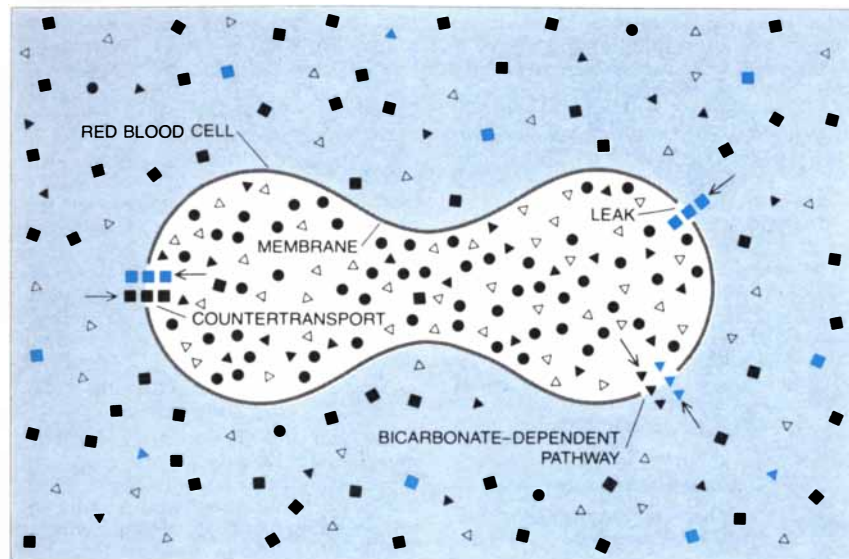
mensions of disease can be determined. Only 30 years have passed since the first example of that kind of definition, the description of sickle-cell anemia as a molecular disease by Linus Pauling and Harvey A. Itano of the California Institute of Technology. Since then workers in many laboratories have done experiments that have identified not only the precise substitutions of amino acids taking place in sickle-cell anemia and other abnormal forms of hemoglobin but also the chemical identity of the genes directing the synthesis of this essential protein.

Similar perceptions are now emerging in the understanding of the molecular events in the development of such complex disorders as atherosclerosis and diabetes. For example, Joseph L. Goldstein and Michael S. Brown of the University of Texas Health Science Center at Dallas have identified the receptors responsible for the uptake by human fibroblast cells of low-density lipoproteins incorporating cholesterol and have shown that there is an inherited defect of these receptors in certain patients with a high level of cholesterol in their blood. Such patients are predisposed to develop atherosclerosis.

Similarly, it is now realized that diabetes mellitus (the common disorder characterized by increased concentrations of sugar in the urine and the blood plasma) has not one but many causes that may involve defects in the synthesis and processing of insulin or in its binding by receptors on cells and its subsequent action on the target cells. The defects may be inherited or may be due to an attack on the system by viruses.

All these advances suggest that the present classifying of disease into categories such as hypertension, diabetes, cancer and mania is crude. Classifications of this broad kind are in fact being supplanted by much more precise identifications of the molecular defects that underlie those general manifestations of physiological disorder. It is also being recognized in increasing detail how such molecular defects result from the interaction of genes and specific conditions associated with the environment. Rational and effective strategies for the prevention and control of complex diseases will be built on the basis of this kind of understanding.

To me the special appeal of the story of lithium, membranes and mania is its stark contrasts. It is somehow surprising and fascinating that a simple salt, an ion, an extract of rock, is able to alter such an ephemeral and subtle property of mind as mood. People are more accustomed to the idea that states of feeling are affected by relatively complex organic substances such as opium, marijuana, cocaine and alcohol. The physicochemical simplicity of lithium arouses the hope that it will provide a light to clarify the neuronal basis of mood.



LITHIUM IN HUMAN BLOOD is shown schematically in a patient who is being treated with a lithium salt. The ions that figure in the process are lithium (colored rectangles), sodium (black rectangles), potassium (black circles), lithium carbonate (colored triangles), bicarbonate (black triangles) and chloride (open triangles). The chemical symbols for these cations and anions are respectively Li^+ , Na^+ and K^+ , LiCO_3^- , HCO_3^- and Cl^- . In a patient under lithium therapy the concentration of Li^+ in the blood plasma (light color) is held at about one millimolar. The concentration inside a cell is kept at a lower value (about .3 millimolar) by the operation of a lithium-sodium countertransport system. When a steady state is reached, the outward movement of lithium through the countertransport system is equaled by inward movement of the ions through two pathways: a leak and an anion-exchange pathway that exchanges LiCO_3^- for HCO_3^- or Cl^- . Several movements through the membrane are shown at the top; the steady-state situation at any given moment is shown at the bottom. In each representation the relative proportions of the various cations and anions in the cell and in the plasma are indicated.

Meet some of the Stars of our show.

Take the little charmer to your right.

She's not exactly your run-of-the-mill starlet.

But then, NOVA isn't exactly your run-of-the-mill TV show.

NOVA is the PBS TV show about science—for people who didn't know they were interested in science. The show that explores everything from the underwater social life of a coastal lagoon to why boys "play rough" and girls don't.

And we're delighted to help make it possible in part by a grant from TRW.

NOVA

Our version of a star-studded evening.

This season on NOVA, you'll spend an hour star-gazing at other galaxies of our universe. Discover some new theories about dinosaurs and



circulatory system, and see for yourself just what happens when a heart goes haywire.

This year NOVA will also take you rock-hunting on the moon.

Show you how sex got started and why some species manage quite nicely without it, thank you.

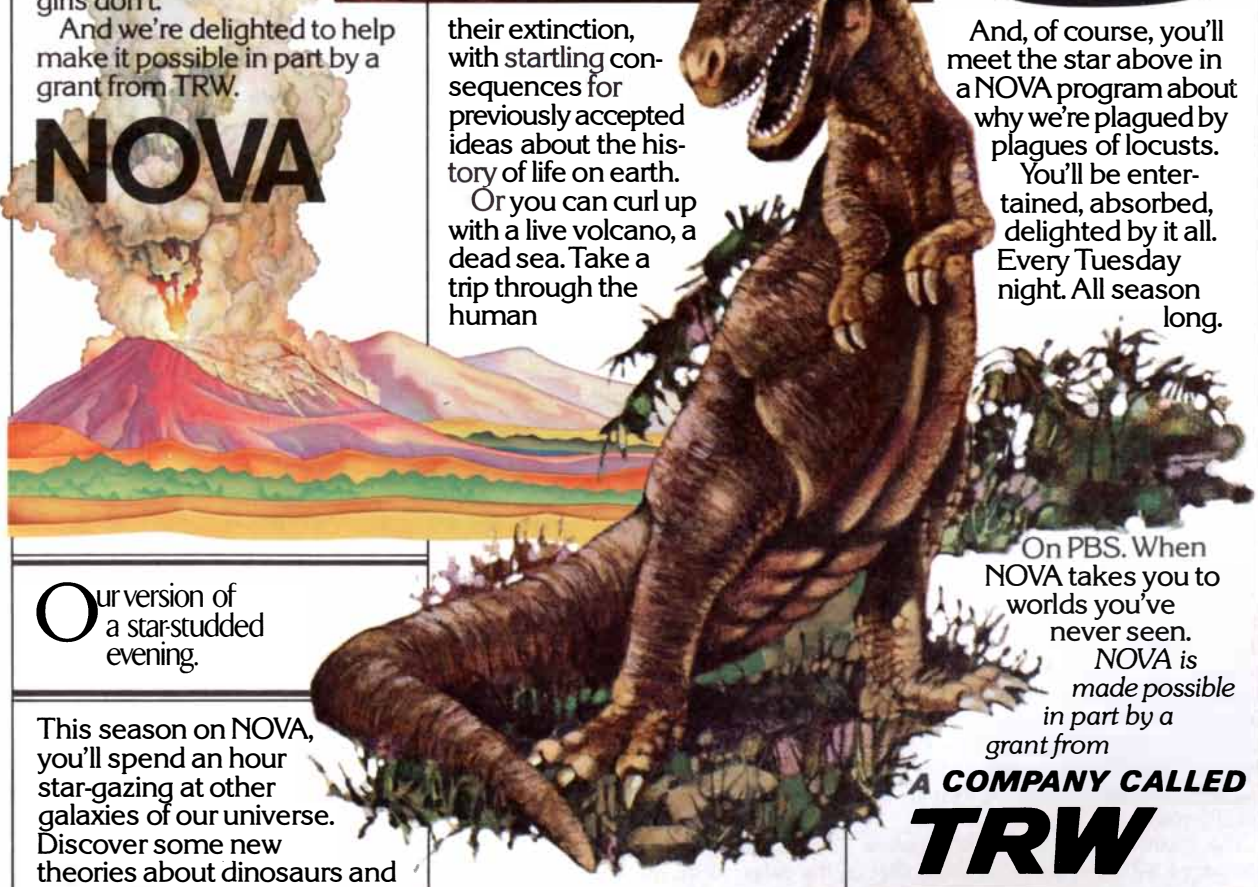


their extinction, with startling consequences for previously accepted ideas about the history of life on earth.

Or you can curl up with a live volcano, a dead sea. Take a trip through the human

And, of course, you'll meet the star above in a NOVA program about why we're plagued by plagues of locusts.

You'll be entertained, absorbed, delighted by it all. Every Tuesday night. All season long.



On PBS. When NOVA takes you to worlds you've never seen.

NOVA is made possible in part by a grant from

**A COMPANY CALLED
TRW**

THE AMATEUR SCIENTIST

The charm of hydraulic jumps, starting with those observed in the kitchen sink

by Jearl Walker

When a stream of water falls from a faucet onto a smooth sink with an open drain, a circle forms around the point of impact. The radius of the circle depends on the amount of water in the stream. The smaller the flow, the smaller the circle. Inside the circle the water is shallower than it is outside. The transition between the two heights is called a hydraulic jump. It is a stationary shock wave that is the hydraulic analogue of the atmospheric shock wave created by a supersonic airplane.

Stationary hydraulic jumps develop in many common situations but probably go unnoticed. They may arise in water flowing down a driveway or along a curb. Some can be seen in irrigation canals or small streams. The most dramatic hydraulic jumps occur in certain rivers when the tide comes in from the sea. These large-scale hydraulic jumps, called bores, move upstream at speeds of as much as 12 knots, are as much as 20 feet in height and extend across the full width of the river. The sudden and unexpected appearance of a bore can cause trouble for vessels on the river.

Of the three general types of surface waves on water, only shallow-water gravity waves matter in the explanation of hydraulic jumps. In these waves the motion is controlled by the gravitational pull on the water after the water is displaced at first. The speed of such a wave depends primarily on the depth of the water. Deep-water gravity waves, which develop on ocean surfaces, do not depend on the depth and play no role in hydraulic jumps, since the jumps occur in relatively shallow water. Ripples of the kind caused by small insects on the surface of a pond are controlled by the surface tension of the water rather than by gravity. Such ripples, which are sometimes called capillary waves, have relatively small wavelengths (a few centimeters or less) and play no role in hydraulic jumps.

The speed of flowing water is often compared with the speed at which shallow-water gravity waves move over still water of the same depth. If the water is

moving faster than the waves, the flow is said to be supercritical. It is subcritical if the flow is slower than the waves. A critical flow is seen when the speeds match. A hydraulic jump develops when a supercritical flow switches to a subcritical one. The transition is sudden and possibly chaotic because during the change the flow becomes quite unstable.

Shallow-water gravity waves are generated when a barrier interferes with the normal flow of the water. If the flow is subcritical, the waves can travel both upstream and downstream. In some situations one wave might remain stationary just downstream of the barrier. Although the speed of a wave depends primarily on the depth of the water, it also depends on the wavelength of the wave. In the range of wavelengths that can be generated by a barrier one upstream wave may have a speed matching the speed of flow of the water. That wave remains in place because it travels upstream at the same speed as the water travels downstream. The wave is constantly reinforced by the interference of the barrier with the flow of water. The other waves generated by the barrier travel away from it and die out as their energy is dissipated.

When the flow is supercritical, none of the waves move faster than the water and thus none can move upstream. They are all swept downstream by the flow. If the flow is changing from supercritical to subcritical, a stationary wave can be formed when the flow passes through the critical state. Then a match is possible between the speed of an upstream wave and the speed of the flow. From such a stationary wave a hydraulic jump can develop.

A ridge across a streambed provides an example of how a barrier can create supercritical flow, a standing wave and a hydraulic jump in an initially subcritical flow. If the ridge is small and provides only a slight resistance to the flow, the movement remains subcritical, but a small stationary wave may arise behind the ridge. With more resistance from a larger ridge the wave is larger.

A further increase in the ridge and

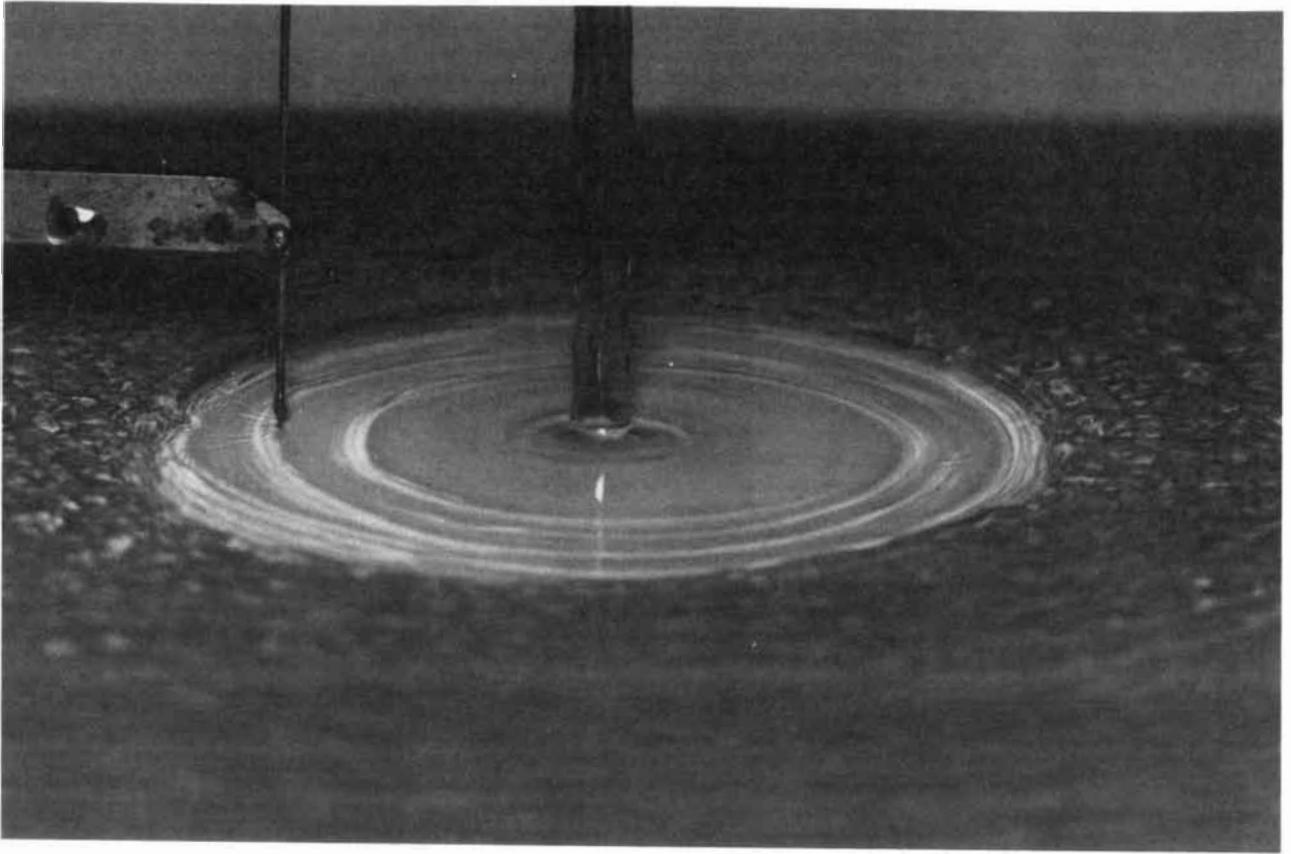
its resistance might create a wave with a relatively shallow region behind the ridge and in front of the first crest. Because the speed of shallow-water gravity waves depends on the depth of the water, the speed of a wave would be relatively small in such a shallow region, forcing the flow there to be supercritical. The first crest downstream of the shallow region would restore subcritical flow because of the increase in depth. If the shape of the water surface is maintained, the transition from the shallow region (supercritical flow) to the deeper one (subcritical flow) is a hydraulic jump. If the ridge does not itself create such a stationary wave, one might still develop in this way if a second ridge farther downstream prevents a wave from the first ridge from being carried downstream.

The shape of a hydraulic jump is normally classified according to a condition of flow called the Froude number. It is the ratio of the square of the speed of flow to the square of the speed of shallow-water gravity waves over still water of the same depth. The Froude number is greater than 1 for supercritical flow, equal to 1 for critical flow and less than 1 for subcritical flow.

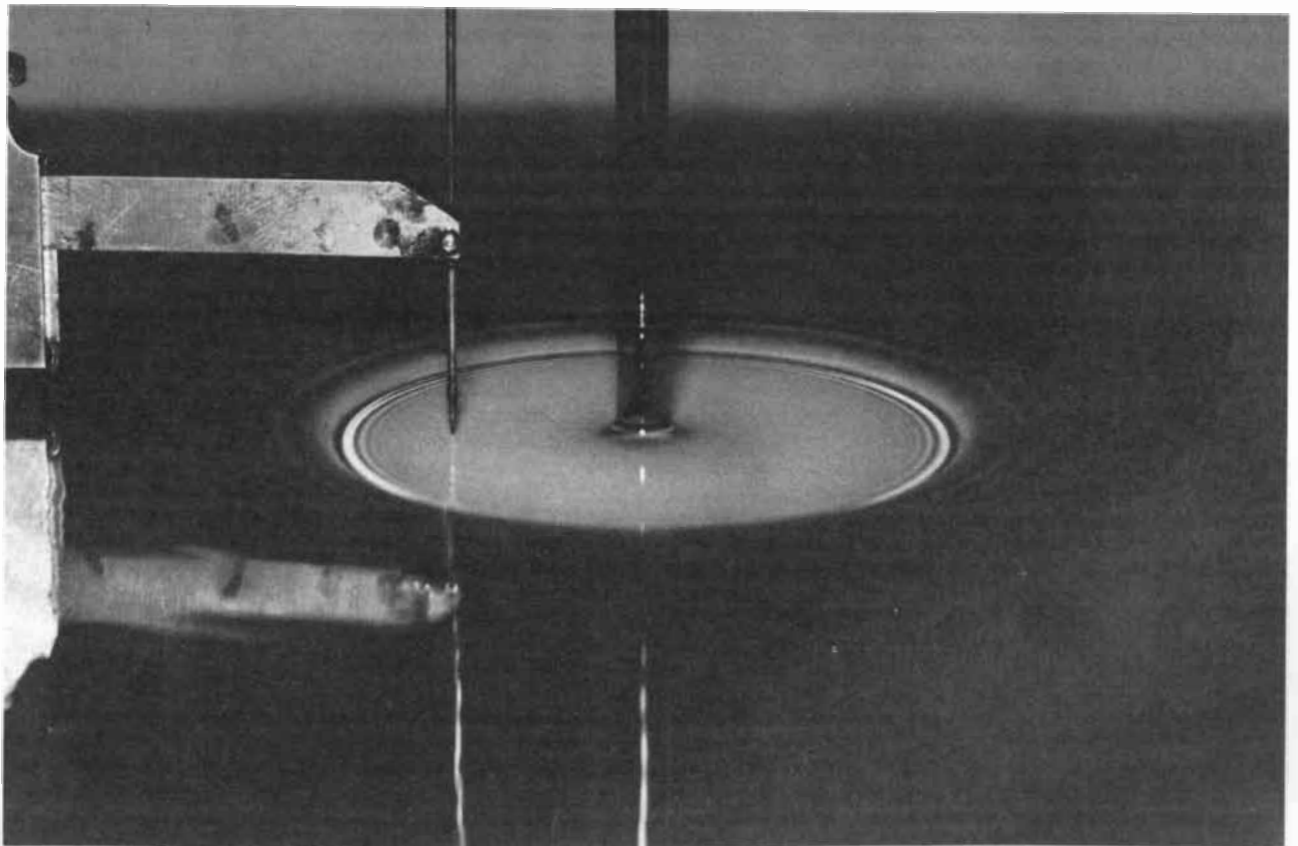
The shape of a jump depends on the Froude number of the supercritical flow ahead of the jump. (In principle the flow is unstable only at a Froude number of 1, but the instability is so chaotic that there is no precise location for this event. Hence the Froude number ahead of the jump is employed for classification.) If the number is between 1 and 3, the jump is said to be undular; it consists of a large first crest and smaller crests trailing downstream. After them the water surface is relatively smooth. At an initial Froude number between 3 and 6 the transition between the two water levels is smoother and the trailing waves are absent. The jump is said to be weak.

If the initial Froude number lies between 6 and 21, the transition gives rise to large unstable waves that may travel for considerable distances downstream from the jump. This kind of jump is called an oscillating jump because of the irregular waves. A steady jump, which lacks the destructive waves, is created when the initial Froude number is between 21 and 80. Roughly half of the kinetic energy of the inflowing water is dissipated in the turbulence of the jump. At higher initial Froude numbers the jump is again rough and irregular, dissipating as much as 85 percent of the kinetic energy of the water and sending potentially destructive waves downstream.

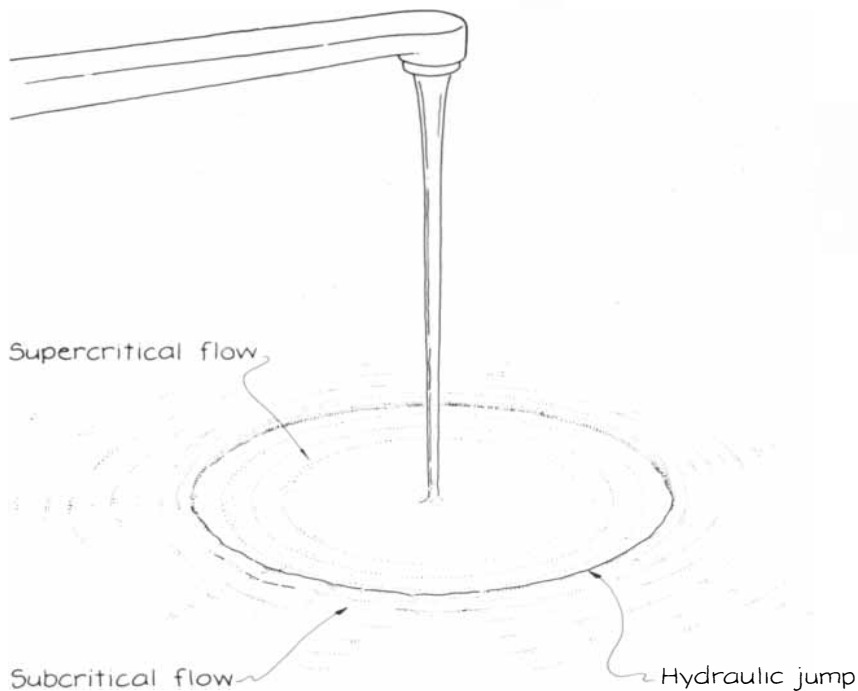
The amount of dissipation of energy can be crucial in the design of a sluice for a dam and of other systems for draining water. It is sometimes necessary to decrease the kinetic energy of a flow in order to avoid damaging the water channel. Placing a barrier across the



A hydraulic jump photographed by R. G. Olsson and E. T. Turkdogan



The effect of adding glycerine to the water



Details of a hydraulic jump

bottom or a sluice gate across the top might be beneficial. The barrier either must be designed carefully or must be adjustable so that the Froude number of the supercritical flow can be controlled and the hydraulic jump does not send potentially destructive waves downstream.

Whether or not a gate on a channel creates a hydraulic jump depends on the steepness of the channel. For a given volume of water flowing down a channel each second the slope determines the

speed of flow and the depth of the water. The two are related in a simple way. When the channel is steep, the flow is rapid and the depth is shallow. With a more gradual slope the flow is slower and the water is deeper. A steep channel is defined as one that develops a supercritical flow of water. A critical channel generates a critical flow and a gradual channel a subcritical flow.

If a sluice gate is lowered into the water of a gradual channel in such a way that the flow emerging from below the

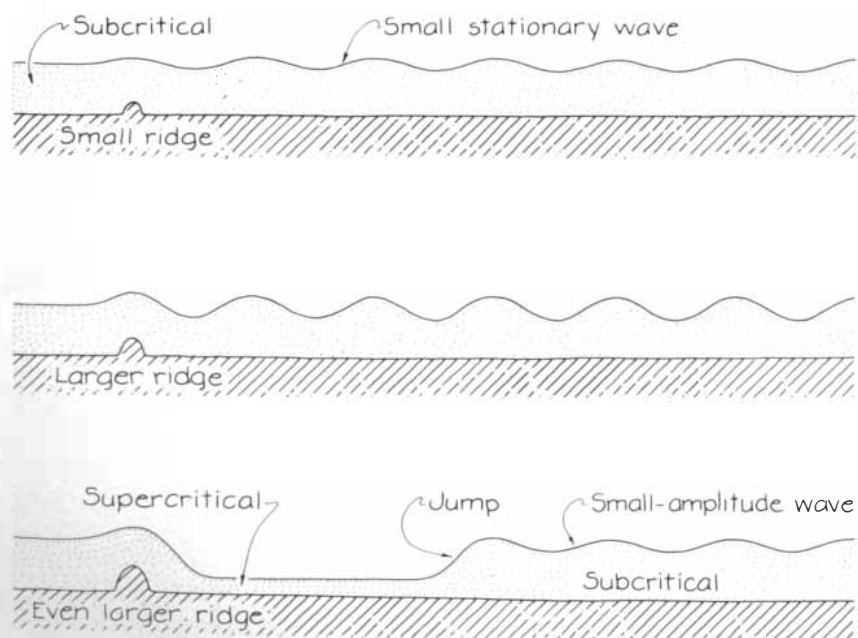
gate is supercritical, a hydraulic jump will form to restore the subcritical flow the water would normally have in the channel. If the supercritical flow from the gate builds up in a critical channel, there may not be a distinct jump. The flow changes from supercritical to critical and remains there, making the surface unstable. If the sluice gate causes a supercritical flow in a steep channel, the flow remains supercritical. It also remains relatively stable down the entire channel.

A hydraulic jump can also turn up when a steep channel joins a gradual one, forcing the water to make the transition from a supercritical to a subcritical flow. In the steep channel the water has a rather large Froude number and thus remains supercritical and relatively steady because it is fairly stable against small perturbations from any obstacles. When the water flows into the gradual channel, it must go slower and become deeper in keeping with the normal flow down such a slope. The Froude number is reduced to approximately the value where the flow becomes unstable against perturbations from obstacles. Surface waves are formed, and the hydraulic jump develops to complete the transition to subcritical flow. The result is the stationary wave that brings a sudden and dramatic change in the depth of the flowing water.

This jump may not form near the place where the steep channel joins the gradual one. Its location depends in part on the slope of the gradual channel. The greater the slope of that channel is, the more gradually the Froude number of the flow is reduced. The critical value of 1 is reached farther downstream. Only then will the jump form. If the gradual channel is only slightly angled, the Froude number is reduced sooner. Then the jump forms nearer the junction of the two channels or even up in the steep channel.

When I began investigating hydraulic jumps recently, I struggled with several fundamental questions. Why do jumps occur? As I have mentioned, a jump serves as a transition from a supercritical to a subcritical flow, but why is the transition necessary and why must it be made suddenly in a jump? If jumps suddenly increase the depth of flow, why do they not form in subcritical flows? And why are there no jumps at transitions from one supercritical flow to another with a lower Froude number?

Let me tackle the questions by considering a supercritical flow from a sluice gate or a steep channel that joins a gradual channel. The speed at which water can run uniformly down the gradual channel is governed by the slope and roughness of the channel. The problem is that the water entering the gradual channel is moving faster than the water in it. Forces from the water already in the channel begin to slow the entering



Supercritical flow from a ridge



The Ultimate Wagon

Imagine having gas mileage about the same as a full-size 2-wheel drive wagon . . . and 4-wheel drive for added security when you need it most. 21 HWY EST. (16) EPA EST. MPG.*

That's what makes Jeep Wagoneer the ultimate family wagon. No conventional 2-wheel drive wagon can perform so confidently under pressure when the rains are torrential, the streets unplowed . . . or when the beauty of the off-road beckons.

Wagoneer may take a tough-minded approach to conditions outside . . . but inside it's a soft touch, with room, luxuries and driving comfort you'd marvel at in any station wagon. And with a full range of Wagoneer models to choose from, you'll find the price of owning one surprisingly affordable.

The mileage of 2-wheel drive full-size wagons . . . the security of 4-wheel drive. Doesn't your family deserve the ultimate?



 **Jeep Wagoneer**

AT AMERICAN MOTORS

*Figures are for comparison. Your results may differ due to driving speed, weather conditions and trip length. Actual highway mileage will be less. California mileage will be different. Jeep Corporation, a subsidiary of American Motors Corporation.



What you don't know might irk you.

So we've compiled all the facts on Oregon for you, and backed them up with color photographs in our free, 64-page, master travel-guide. It's proof positive that Oregon's incredible beauty is no myth. All you have to do is call us on it, TOLL-FREE, and we'll send you the wonderful truth.

PHONE TOLL FREE for your guide. 1-800-547-7842
(outside Oregon) Or write: Oregon Travel Information, Room AT 200, Department of Transportation, Salem, OR 97310.

PROPOSALS SOLICITED

New community in isolated rural area in midst of southcentral Montana coal development region requests submittal of proposals to supply power to the community. Energy efficient, passive solar designs are being actively promoted in the town's estimated 300 residences. Electricity is needed for both the residential and accompanying commercial sectors (motel, grocery store, restaurant, service station and administration building).

Reliable renewable energy techniques are encouraged. (Photovoltaics?) A 200 KV, single-phase line services the town now.

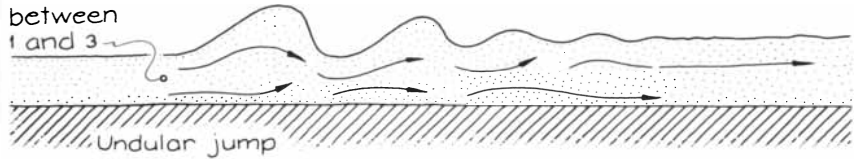
Applicants may have sole management authority or consider partnership with developers.

Submit proposal by April 30, 1981 to:

Energy Project
Box 20762A
Billings, MT 59104
tele: (406) 245-6144

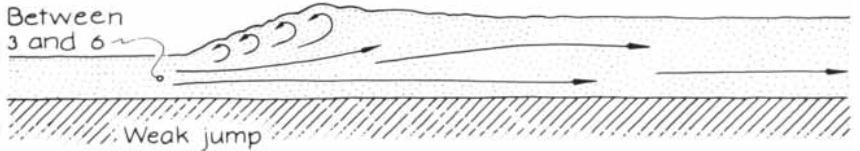
For more information, write or phone the above address.

Froude number
between
1 and 3



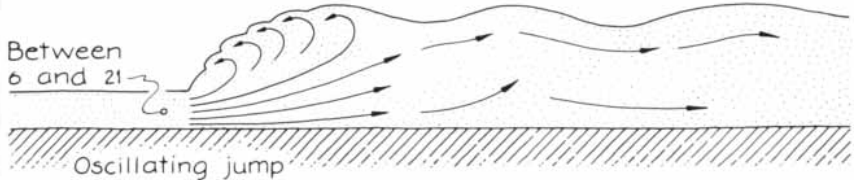
Undular jump

Between
3 and 6



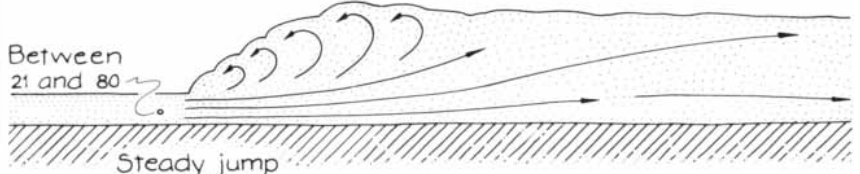
Weak jump

Between
6 and 21



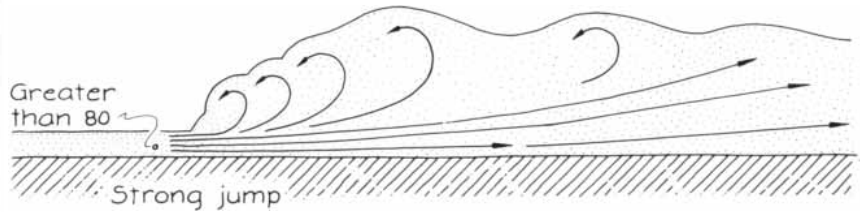
Oscillating jump

Between
21 and 80



Steady jump

Greater
than 80



Strong jump

Types of hydraulic jump

water. In order for all the entering water to be moved downstream at the slower speed the depth must increase. That is why supercritical water flowing onto a gradual slope begins to move slower and get deeper.

During this rather gradual transition the Froude number of the water decreases to nearly 1 and the flow becomes sensitive to perturbations it encounters in the channel. By sensitive I mean that if the flow meets even a modest resistance from an obstacle, the height of the surface increases considerably. Suppose the flow (at a Froude number near 1) meets a small ridge in the channel. The resistance of the ridge to the flow may be small, but it forces the surface of the water upward by a relatively large amount. If the flow had encountered the ridge when it had a different Froude number, the resulting rise would not have been as dramatic.

When the surface of water is suddenly pushed up, waves are created. (You can achieve the same result in a sink or a tub by pushing suddenly down on the sur-

face of the water or by raising your submerged hand to the surface.) The uplifting rapidly completes the transition to the greater depth required by the gradual channel, and some of the waves it creates remain as part of the jump.

Such a rapid transition is not made between two subcritical or two supercritical flows for several reasons. Suppose a subcritical flow moves from a gradual channel to a more gradual one. In the second channel the water would have to slow down and become deeper so that all of it could flow at the speed dictated by the forces on it. The transition between the two depths, however, would not form a jump. As the entering water began to move slower and get deeper its Froude number would decrease, moving farther from the critical value of 1 where the flow is sensitive. If the flow meets a small obstacle during the transition, the resistance of the obstacle to the flow would not force the surface of the water to rise much. The flow would be disturbed by the obstacle but would quickly regain its stability.

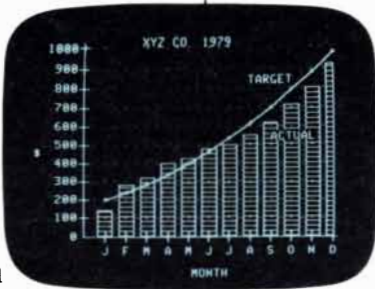
Three good reasons why professionals pick Apples.

1 In research

Apple personal computer systems help you collect, store and analyze data as fast as you can load a disk and execute a program. Because more than 100 companies offer software for Apple, you have the largest program library for manipulating your data in the personal computing world. Need special programs? Use any of Apple's development languages — BASIC, FORTRAN, Pascal.

2 In engineering

Apple personal computer systems let you define models, make trade-offs and refine prototypes. Want to study cause and effect of several variables? Apple computes new results instantly and displays them in colorful, easy-to-read graphs, charts or plots on a video monitor.



3 In production management

Apple personal computer systems make it easy to gather data, analyze productivity, measure yields and facilitate all phases of production control. Want to speed up repetitive tasks?

Rely on Apple's word processing capabilities to write, edit and print your reports.

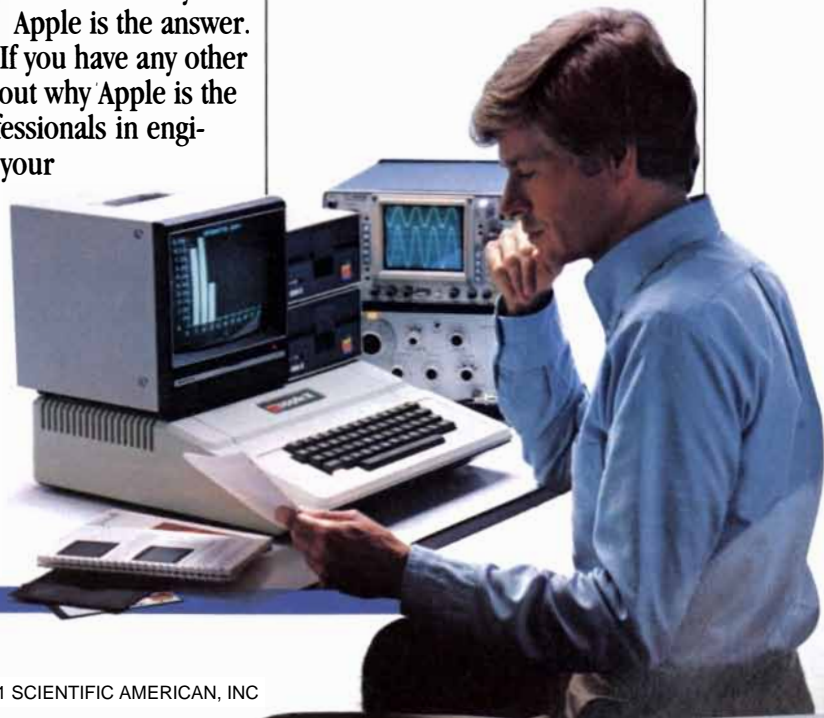
Apples grow with you.

Whichever system you pick, Apple never locks you into a single configuration. You can use up to four or eight I/O accessory expansion slots to add an IEEE bus, Apple's Silentype™ printer, a modem or a graphics tablet. Add memory up to 64K bytes or 128K bytes. Add up

to four or six 5¼" disk drives without adding any overhead.

For support, service and the best extended warranty in the industry — Apple is the answer.

If you have any other questions about why Apple is the pick for professionals in engineering, see your nearest Apple computer dealer or



	Apple II	Apple III
Maximum Memory Size	64K bytes	128K bytes
Screen Display	40 column (80 column with peripheral card) 24 Lines Upper Case	80 column 24 Lines Upper Case/Lower Case
Screen Resolution (B&W)	280 x 192	560 x 192
Screen Resolution (Color)	140 x 192 (6 colors)	280 x 192 (16 colors)
Keyboard	Fixed	Programmable
Numeric Key Pad	Accessory	Built-in
Input/Output	8 expansion slots	4 expansion slots plus built-in: disk interface RS-232 interface Silentype™ printer interface clock/calendar
Disk Drives	Add-on one to six drives	One drive built-in, plus interface to support three more drives
Languages	BASIC Fortran 77 Pascal Assembly Pilot	Enhanced BASIC Fortran 77 Pascal Assembly
Typical Configuration Pricing	CPU, 48K RAM, single disk drive, B&W Monitor (9"), Silentype™ printer, and BASIC. \$2875.00*	CPU, 96K RAM, integrated disk drive, B&W Monitor (12"), Silentype™ printer, SCS, Enhanced BASIC. \$4865.00*

* Suggested retail price.

call **800-538-9696**. In California, **800-662-9238**. Or write: Apple Computer, 10260 Bandle Drive, Cupertino, CA 95014.



apple computer inc.



PIONEER SYSCOM.

FINALLY, ALL THAT'S INVOLVED IN OWNING GREAT HI-FI IS A LOVE OF MUSIC.

Now the music-lovers of our world can have all the wonders of great music without the complexity of great hi-fi.

Pioneer, the leading maker of high fidelity components in America, has created a system for you. It's called Syscom.™

Syscom is serious, substantial, superb hi-fi. Each component was created specifically for Syscom.

And since the entire system was both built and matched by Pioneer engineers, you can't make a mistake.

Syscom is everything you need to maximize your music. Turntable, amplifier and tuner (or receiver), tape deck, and speakers. Syscom even comes with its own specially designed cabinet. So it's as easy to look at as it is to listen to.

Headphone, microphone and digital timer optional.

There are seven Syscom systems to choose from.

One is perfect for your budget, your room, your music.

For the name of the Pioneer Syscom dealer nearest you, call us at **800-621-5199***. It'll be the beginning of a great new love.

*(In Illinois 800-972-5855). © 1981 U.S. Pioneer Electronics Corp.



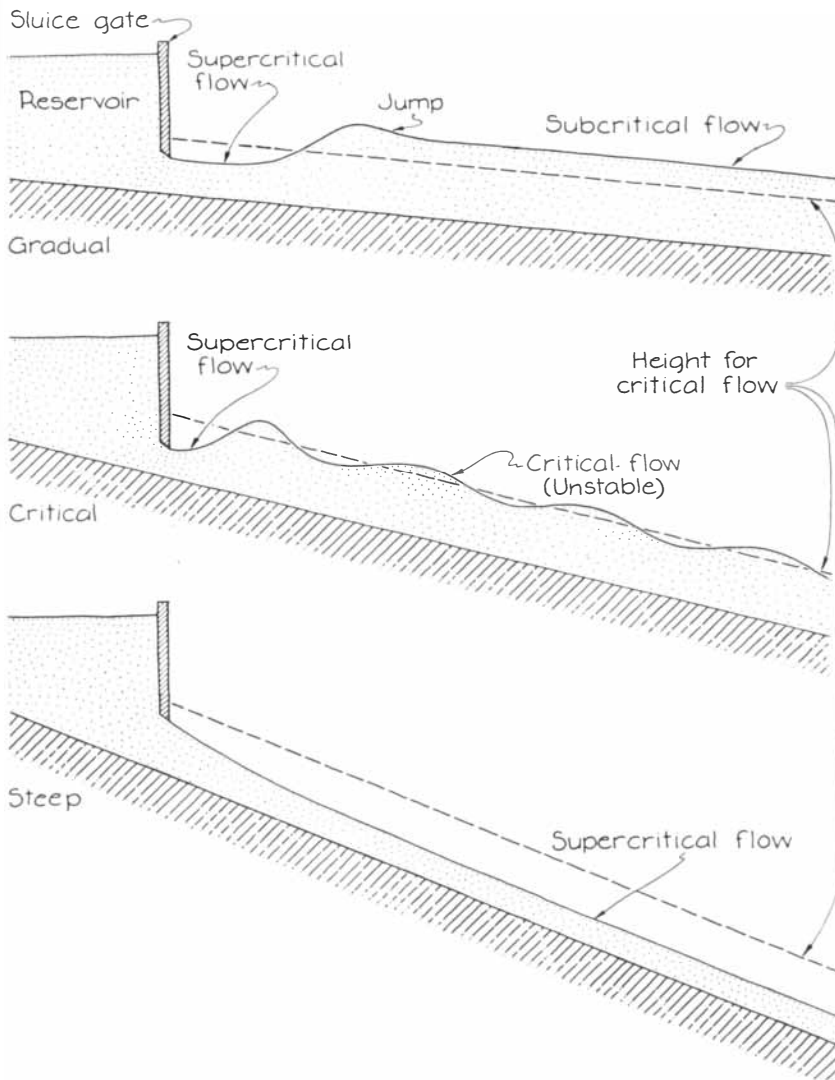
Something similar happens when water flows from one supercritical channel into another one with a smaller slope. The Froude number is reduced, but it remains above 1 and the flow is not sensitive to small obstacles. Resistance to the flow might cause a negligible increase in height, but the change would not be dramatic and any wave generated would be swept downstream by the supercritical flow.

Some of the properties of hydraulic jumps can be observed by simply positioning two flat planes in the water flowing from a faucet. I held a glass windowpane in the stream from my kitchen faucet, tilting the glass so that the water poured onto a flat piece of wood I held snugly against the lower end of the glass. The water ran down the glass onto the wood and then off the edge of the wood into the sink. This arrangement of two tilted surfaces differs from the flow of water out of a steep channel into a gradual one because the water is not confined to a channel by side walls. Nevertheless,

the dependence of the jump on the angle of tilt can be observed.

By adjusting the angle of the glass and the wood I could create a small hydraulic jump almost anywhere on the wood. The glass had to be tilted at a large angle to ensure that the water moving along it was supercritical. If I made the tilt of the wood fairly steep but still less steep than that of the glass, the jump appeared farther from the intersection of the glass and the wood. With a negligible tilt (the wood may even have been horizontal) the jump was at the intersection.

I also experimented with hydraulic jumps in tilted channels by means of a simple rig consisting of a rubber washbasin, two aluminum channels and some plastic tape. The metal channels were made out of spare aluminum strips that were three or four feet long and several inches wide. I bent both strips into a rectangular channel and taped one of them to the side of the washbasin where I had cut out part of the wall. The channels could also be made out of spare



The effects of channel slopes of increasing steepness

SAVE SAVE

LOWEST PRICES ON TRS-80 Computers

Buy Direct From

Pan American Electronics INCORPORATED

Radio Shack AUTHORIZED SALES CENTER

Dept. SA., 1117 Conway, Mission, Tex. 78572
Texas & Principal Number 512-581-2765 Telex 767339
Toll Free No. 800-531-7466

**UNIVERSITY OF SOUTHAMPTON
CHAIR OF ENGINEERING MATHEMATICS**

Applications are invited for the Chair of Engineering Mathematics which will become vacant following the retirement of Professor J W Craggs on 31 March 1981. Further particulars are available from the Secretary and Registrar, The University, Southampton, SO9 5NH, England, and applications (one copy from candidates overseas and ten copies from others) should be submitted by 13 April 1981. Please quote ref: SA.

Amateur Telescope Making

Edited by Albert G. Ingalls
Foreword by Harlow Shapley

This three-book set, published by SCIENTIFIC AMERICAN, has served many thousands of hobbyists as the authoritative reference library of the enthralling art of amateur telescope making.

BOOK ONE the basics of glass grinding and how to complete your first telescope. (510 pages, 300 illustrations)

BOOK TWO how to make finer and larger telescopes; and on to advanced optical work, including micrometers, synchronous clock, steller and solar eclipse photography. (650 pages, 361 illustrations)

BOOK THREE further fields of enterprise, including binoculars, camera lenses, spectrographs, Schmidt optics, eyepiece design, ray tracing. (646 pages, 320 illustrations)

SCIENTIFIC AMERICAN

ATM Dept., 415 Madison Ave., New York, N. Y. 10017

Please send me postpaid the following AMATEUR TELESCOPE MAKING books. My remittance of \$_____ is enclosed.

- BOOK ONE** \$10.00
- BOOK TWO** \$12.50
- BOOK THREE** \$12.50

For shipment outside U.S. add \$1.00 each.

Name _____

Address _____

City _____

State _____ Zip _____

Residents of New York City please add city sales tax.
Other NYS residents please add state sales tax plus local tax.

eaves-trough material. I installed a hose so that the basin would continually overflow. Thus a constant amount of water flowed down the channel.

At the lower end of the first channel I taped the second channel. With blocks of wood under the channels I could adjust the channel angles. The upper channel was steep, the lower one gradual. Water ran from the basin down the steep channel, down the gradual channel and then into a drain in the basement floor. I had fastened the channel to the basin with plastic tape. It was not a satisfactory arrangement because the tape eventually became loose, but I was able to make most of my observations before the structure collapsed.

Although creating a jump was easy, I could not tell what kind of jump it was. A much better arrangement would be to make the channels out of transparent plastic so that the experimenter could observe the jump from the side. Then it would be easier to determine whether the surface of the water was smooth or had relatively large waves.

To investigate the effects of a ridge or a miniature sluice gate I removed the lower channel and placed obstacles in the upper one. My gate was a metal plate almost as wide as the channel. By adjusting the slope of the channel and the depth of the plate I could generate the various sluice-gate effects. My ridge was a narrow mound of Silly Putty across the width of the channel. By adjusting the height of the ridge and the angle of the slope I again could control whether a hydraulic jump was created.

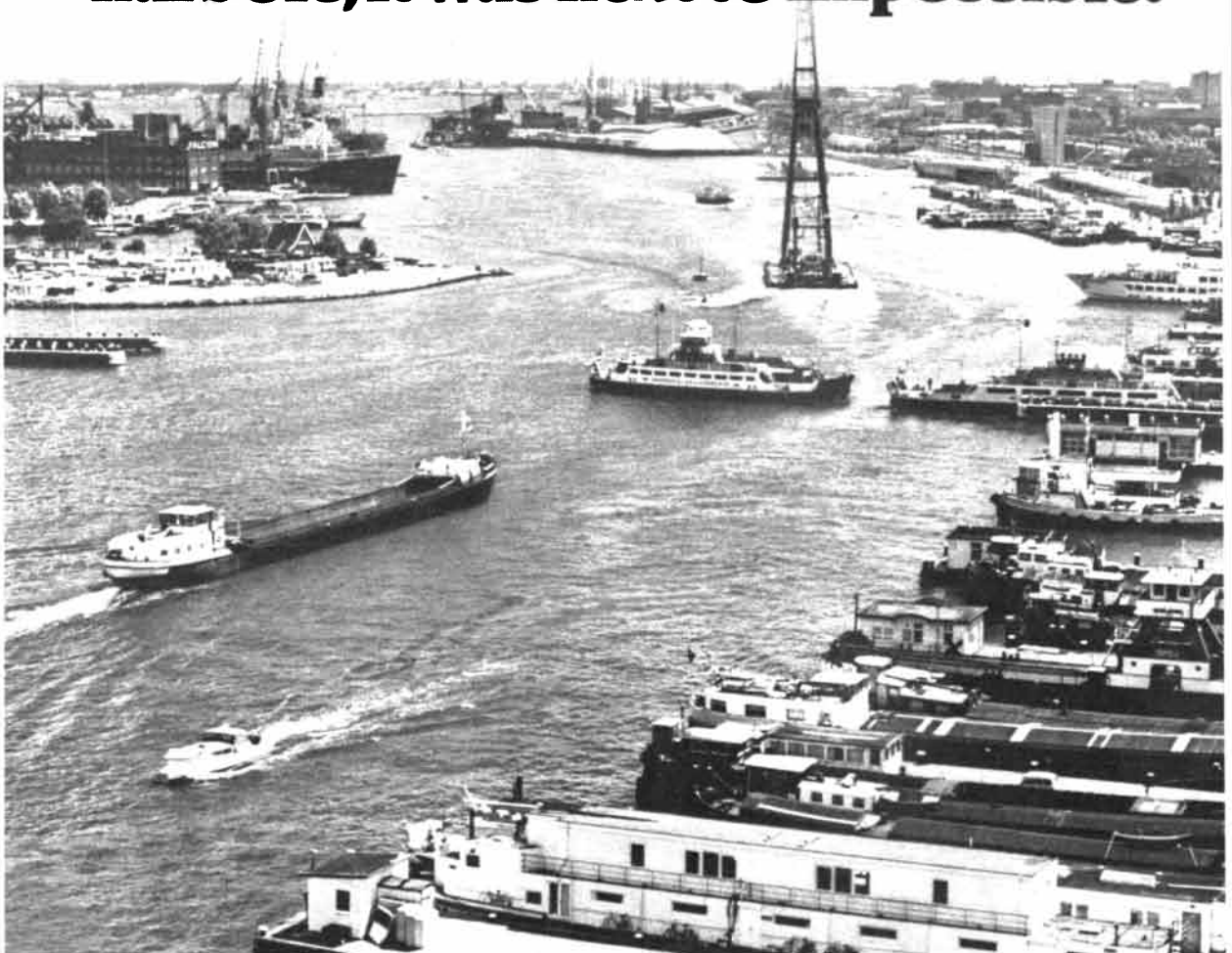
As an alternative to a ridge you can build a smooth and gradual bump in the channel bed. The theoretical analysis of a subcritical flow encountering such a bump on a gradual slope is not easy. One of two possible results is that the surface of the water may rise over the bump if the flow becomes supercritical. The other is that the surface may dip above the bump if the flow remains subcritical.

The hydraulic jump surrounding a stream of water from a faucet was apparently first reported by Lord Rayleigh. In his 1914 paper "On the Theory of Long Waves and Bores" Rayleigh organized the equations governing the kinetic energy and momentum in a bore. As a postscript he offered his simple observations of the hydraulic jump in a sink, noting that it was governed by the same principles.

Since Rayleigh's work the hydraulic jump in a kitchen sink has received little attention until quite recently. One of the most interesting discussions of the jump was published by R. G. Olsson and E. T. Turkdogan of the United States Steel Corporation. They made a stream of water fall onto a flat circular plate held perpendicular to the stream. The water formed a jump on the plate.

The water came from a tank in which

Cleanliness may be next to godliness. But in some Dutch harbors, it was next to impossible.



The well-known feeling of the Dutch for cleanliness didn't stop at the water's edge.

And what they saw in some of their harbors was disconcerting, to say the least.

Every time that a ship pumped water out of its bilge, there was a good chance waste oil would be pumped out with it.

The result was a worrisome oil slick—a slick spreading over many of the

world's harbors these days.

(Indeed, back in 1973 an international maritime ruling called on ships everywhere to monitor their oil waste, to avoid worries like these.)

What to do? Some of the people of ITT came up with an ingenious answer—an optical fiber device that carefully “watches” a ship's bilge water.

A laser beam scans the bilge waste being pumped

out. And if the oil levels are too high, an alarm goes off.

So, the pumping can be stopped.

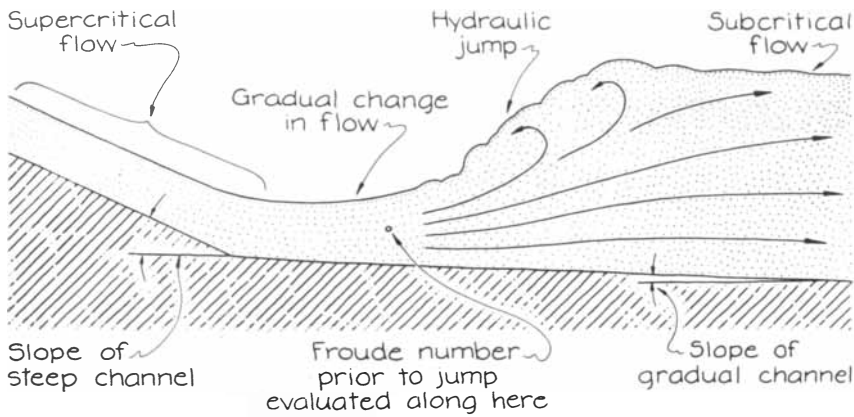
Our ITT device was the first anywhere to be government certified, meeting the required performance standards for this urgent monitoring task.

Obviously, no one expects to unpollute the world's harbors overnight.

But the least we can do is give it a good, clean try.

**The best ideas are the
ideas that help people. ITT**

© 1980 International Telephone and Telegraph Corporation 320 Park Avenue, New York, N Y 10022



The jump at a change in slope

the water level was kept constant. The diameter of the stream was controlled by an aperture. When a hydraulic jump formed on the plate, Olsson and Turkdogan measured the depth inside and outside the jump by inserting into the water a vernier height gauge with a needle pointer. (Anything wider would have disturbed the flow too much, destroying the circular jump and increasing the depth of the water close to the gauge.) The average depth inside the jump was between .1 and .9 millimeter. In the region beyond the jump the depth was between one millimeter and three millimeters.

Olsson and Turkdogan also estimated the speed of flow inside the jump by making high-speed motion pictures (2,000 frames per second) of bits of cork moving on the surface of the water. The speed was almost constant until the water reached the jump, where it began to slow down. In other experiments they replaced the water with more viscous fluids. In general the higher the viscos-

ity, the smaller the radius of the jump.

I generated hydraulic jumps with a variety of objects placed in the path of the stream from my faucet. A plate, a frying pan and even the flat side of a table knife created the full circular jump or at least sections of it. Make sure to provide drainage from a sink or a container. If the water gets too deep, the jump will be destroyed.

When I made the stream fall onto a glass windowpane, I could pour a small amount of soap powder into the area of supercritical flow. The soap was immediately whisked out to the jump, where it turned to foam in the turbulence. The undissolved soap collected just beyond the jump, left there by the sudden slowing of the water as it changed to a subcritical flow.

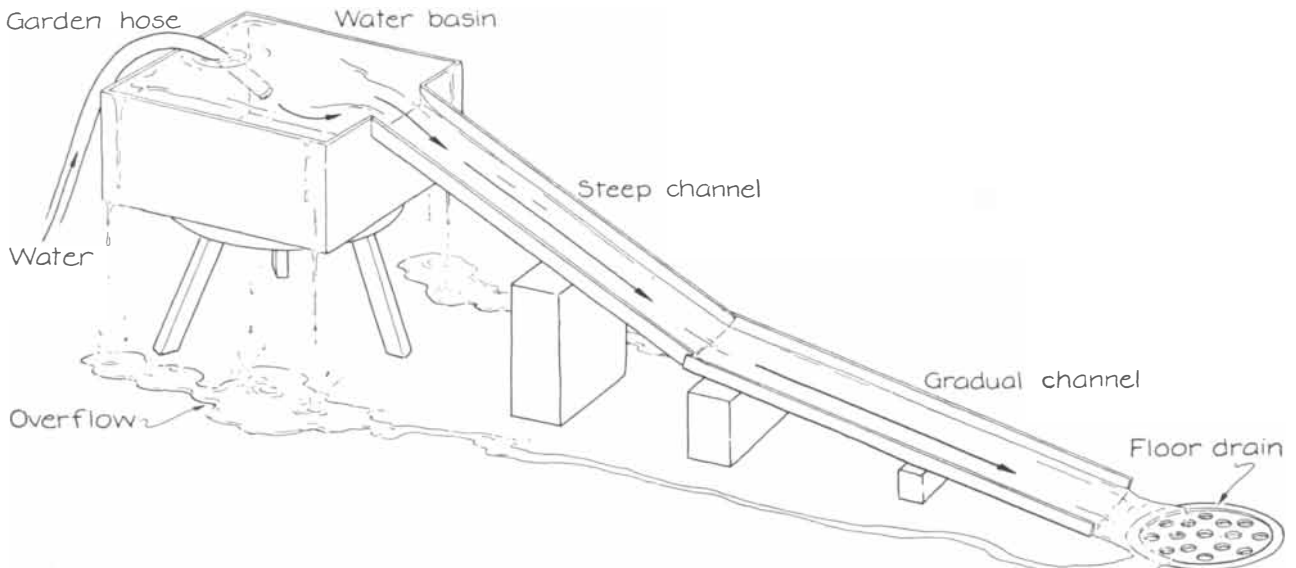
If you create a hydraulic jump in a dark sink or an iron skillet illuminated with fluorescent light, you might see colors in the supercritical area. I find circular bands of blue and yellow (or orange) around the point of impact of the falling

stream. In incandescent light or sunlight the colors are missing. A fluorescent lamp gives rise to the colors because its light is not constantly white, notwithstanding its appearance to the eye. The light consists of three main components. One component is the long-lived phosphorescence of a phosphor on the inside of the tube. The second is the short-lived phosphorescence of another phosphor. The third is the emission spectrum of the mercury excited by the current discharged through the tube. The steady emission from the long-lived phosphorescence is most intense in the yellow region of the spectrum. The periodic emission from the short-lived phosphorescence is in the blue region. Thus, although the averaged output of a fluorescent lamp looks white, it is actually a flickering blue on a constant background of yellow.

When a falling stream of water hits an object, ripples flow out to the hydraulic jump. Since the ripples change the tilt of the water surface, they also change the reflection of light from the water. As the ripples sweep outward from the impact area to the jump they reflect yellow light at times and a mixture of yellow and blue at other times. If the train of ripples is continuous, you will see concentric circles of yellow and yellow-blue around the impact point.

Many fluids found in a kitchen can be employed to generate hydraulic jumps when they are poured onto a flat, well-drained obstacle. I held a glass pane in streams of corn oil, vinegar, beer, syrup, honey and a solution of cornstarch. The fluids with relatively low viscosity showed hydraulic jumps similar to the ones formed with water.

The honey, which was at room temperature, was too viscous to create a hydraulic jump. Instead of colliding with



An arrangement for experimenting with slopes

SCIENCE/SCOPE

Pictures from space are helping mariners in the eastern Gulf of Mexico conserve fuel and travel faster in colder months by showing them where major currents are flowing. Data on the Gulf Loop Current, a circulation of water that moves roughly clockwise through the eastern portion of the Gulf, comes from a GOES (Geostationary Operational Environmental Satellite) spacecraft. An infrared sensor aboard the satellite senses the warmer waters of the current. This information is then converted into pictures and a map showing the Loop Current's coastal edge by latitude and longitude. Ships then can sail with or avoid the current, which flows up to three and one-half knots. The Hughes-built GOES satellites are operated by the National Oceanic and Atmospheric Administration.

A resilient plastic coating could lengthen the lives of infrared domes on Maverick air-to-ground missiles. Using a process called plasma polymerization, Hughes researchers placed a plastic film 5.1 micrometers thick on a curved section of an anti-reflection zinc-sulfide dome. The section, when subjected to fine-grain sandblasting, suffered a transmission loss of only 1.7 percent, compared with 18.5 percent for a noncoated sample. The film reduced the infrared transmission qualities by only 3 percent. Not only did the tests indicate the feasibility of extending dome life, but also that the frangible dome covers that are blown off in flight might even be eliminated.

Tactical cruise missiles can be guided to a target, despite electronic jamming, using signals from navigation satellites. Flight tests over nine months demonstrated extremely accurate midcourse guidance of a Navstar Global Positioning System (GPS) missile guidance system, which was mounted in a pod beneath an F-4 fighter. The system showed to be highly immune to enemy electronic countermeasures when it flew over a simulated high-power jammer without breaking its tracking lock. Tests were conducted by Hughes for the U.S. Air Force.

Hughes Research Laboratories needs scientists for a whole spectrum of long-term sophisticated programs. Major areas of investigation include: microwave devices, submicron microelectronics, GaAs integrated circuits, ion propulsion, lasers and electro-optical components, fiber and integrated optics, pattern recognition, and new electronic materials. For immediate consideration, please send your resume to Professional Staffing, Dept. SA, Hughes Research Laboratories, 3011 Malibu Canyon Road, Malibu, CA 90265. Equal opportunity employer.

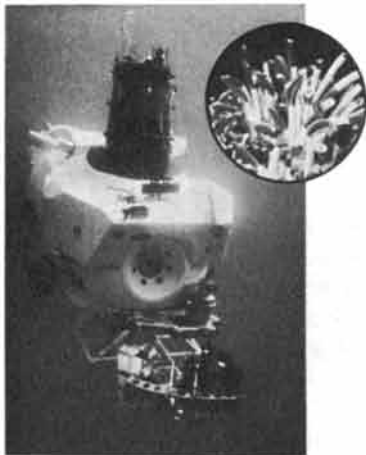
Designers of computer software systems can expect help from other computers in the near future. A computer aid being developed by Hughes serves as draftsman, librarian, and report writer of a design session. The system, appropriately called AIDES (for Automated Interactive Design and Evaluation System), converses with the designer in near English and draws charts on TV-like terminals and plotters. It also analyzes designs for soundness and testability. AIDES reduces the labor intensity associated with software design, while improving consistency and overall quality. Studies indicate the system trims design time by 30 percent and slashes costs for structure chart documentation by 95 percent.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA 90230

**SURPRISED BY LIFE
OCEANUS FOLLOWS
DEEP-DIVING SCIENTISTS
TO A NEW FIND**



Most of the ocean floor a mile and a half down is not crawling with life. But some of it is, as scientists aboard the tiny submersible *Alvin* discovered recently when they hovered over a warm-water vent in the Galapagos Rift off the western shoulder of South America. Under them were colonies of large clams, mussels, crabs and—most surprising of all—a type of giant, red-tipped seaworm never seen before. Here, near waters sailed by Charles Darwin, was a food web sufficient to support a few meters of riotous life on the otherwise desolate bottom.

This is the sort of development OCEANUS is in business to cover. Through our pages, investigators in the vanguard of the marine sciences report directly to you in language anyone seriously interested in the sea can understand. We are published by Woods Hole Oceanographic Institution, one of the leading centers of ocean research in the world—and home base for *Alvin* as well as large surface vessels designed for exploration anywhere in the ocean. We search out scientists and policy experts around the world and organize most of their material in thematic issues. Recent topics have included energy and the sea, food from the sea, pollution, climate, estuaries.

If you're drawn to the sea and what it means to all of us, you'll want to read OCEANUS, the illustrated magazine of marine research

Oceanus

**SUBSCRIPTION DEPT. BOX SC
1172 COMMONWEALTH AVE.
BOSTON, MASS. 02134**

Please enter my subscription to OCEANUS for:

- one year four issues: \$15.00
- two years save \$5.00: \$25.00
- check enclosed* (we request prepayment)
- bill me

(outside of U.S and Canada add \$2 per year to these rates)

Name _____

Address _____

City _____ State _____ Zip _____

*Please make check payable to Woods Hole Oceanographic Institution. Checks accompanying foreign orders must be payable in U.S. dollars and drawn on a U.S. bank.

the pane and then spreading out to the sides in a supercritical flow the stream of honey merged slowly with the thin layer of honey already on the pane. At the point of impact a cylindrical, thin stream wrapped itself around in a rope coil and a broad stream moved back and forth to form a ribbon.

The cornstarch solution belongs to the class of fluids called non-Newtonian. They are characterized by viscosities that can be varied not only by a change in temperature but also by stress. I described these strange fluids in this department for November, 1978. The viscosity of a cornstarch solution immediately becomes high when the solution is stressed. When the stress is removed, the viscosity returns immediately to the lower value.

I prepared a moderately thick solution of cornstarch and water. It was thicker than water but not too thick to flow. When I poured it into a large frying pan, the stream showed a hydraulic jump resembling the type seen in water. As the level of fluid in the pan rose, the hydraulic jump shrank in radius until it looked as though it would disappear. As the radius decreased, the wall of the jump became steeper but never displayed any turbulence. At the smallest radius the wall appeared to be concave. The falling stream thickened near the impact point, collided with the solution already in the pan and then spread to the sides to form the jump. Farther out, at a radius of a few centimeters, a slight increase in height was visible.

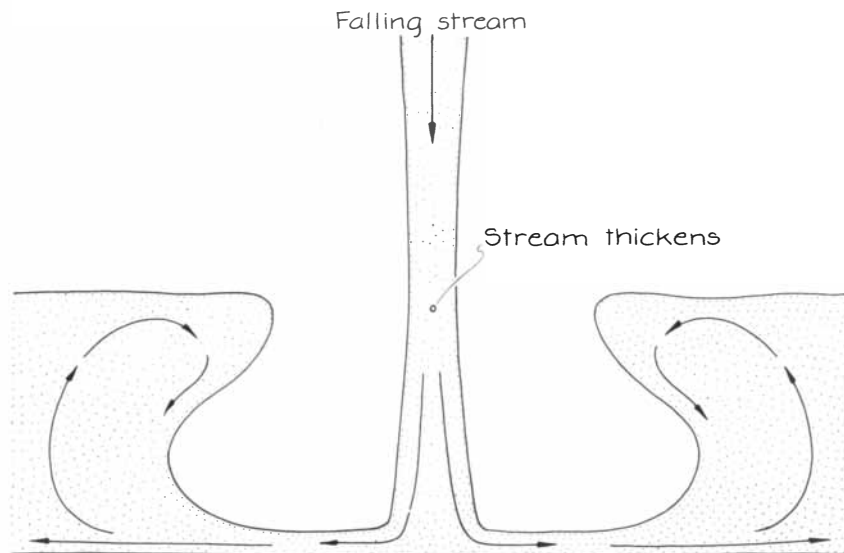
The strange appearance of the fluid around the impact area was probably due to the non-Newtonian nature of my cornstarch solution. When the solution collides with the fluid in the pan, the viscosity increases in the stream just above the impact point and in the solution just below it. As a result the stream

hits a fairly rigid surface even though the surface is only a solution of cornstarch. The fluid in the stream is sent out horizontally in a flow that is supercritical.

After the collision the fluid is still under stress, and so it maintains its relatively high viscosity. Therefore the hydraulic jump has a small radius. The fluid flows horizontally into the jump and then a backwash carries part of the solution to the top of the jump. The overhang of the rim is probably due to a relaxation of the stress and viscosity of the fluid there. As soon as the fluid is pulled back down along the wall and into the stresses of the outward horizontal flow the viscosity increases again. The result is a concave wall around the hydraulic jump.

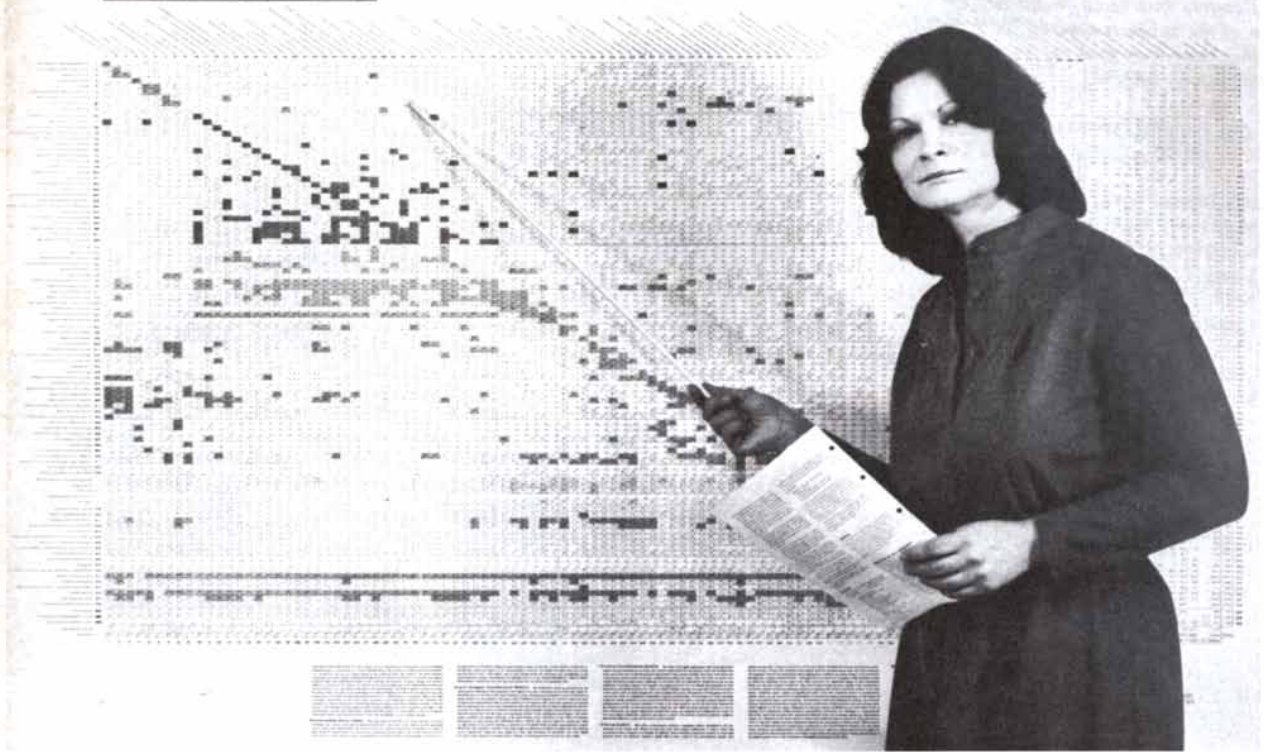
When I ran water onto the pane, it formed a circular hydraulic jump. When I tilted the pane, the jump was distorted. The uphill region moved closer to the stream and the downhill region moved farther away. When water broke through the uphill region, it did not spread out much but instead flowed downhill in the ridge of the jump.

Hydraulic jumps can develop in gases as well as in liquids. One of the most dramatic examples of an atmospheric hydraulic jump was photographed by Peter H. Hildebrand of the Illinois State Water Survey and published in the *Bulletin of the American Meteorological Society* in June, 1977. A dense cloud from a summer thunderstorm was moving to the north across Chicago toward Lake Michigan. As the high-speed cloud, which was 200 to 300 meters high, encountered less dense air it apparently broke into a hydraulic jump, with a sharp front wall followed by several waves. The jump was visible because the dense cloud was noticeably darker than the less dense air.



The pattern of a jump appearing in a solution of cornstarch and water

THE INPUT/OUTPUT STRUCTURE OF THE UNITED STATES ECONOMY



WHAT MAKES THE U.S. ECONOMY TICK?

The editors of *SCIENTIFIC AMERICAN* have prepared a wall chart displaying for the 1980's the Input/Output Structure of the U.S. Economy based on the latest interindustry study from the U.S. Department of Commerce.

The *SCIENTIFIC AMERICAN* Input/Output wall chart does for economics what the table of elements does for chemistry. It answers at a glance questions about the linkage between the microeconomics of the firm and the macroeconomics of the system; about the web of technological interdependencies that tie industry to industry; about the industry-by-industry direct and indirect consequences of swings in public and private spending; about the impact of change in technology, and about any other topic you can think of. You are rewarded by surprise as well as by confirmation of your hunches. For teaching and practical and theoretical studies, here is a powerful, graphic tool.

In the familiar format of the *SCIENTIFIC AMERICAN* Input/Output wall charts for the 1960's and 1970's, the wall chart for the 1980's measures 65" x 52" and is printed in eight colors. Each of the nearly 10,000 cells in the 97-sector interindustry matrix shows (1) the interindustry commodity flow, (2) the direct input/output coefficient and (3) the "inverse" coefficient. Where the direct input/output coefficient exceeds .01, the cell is tinted in the color code of the industrial bloc from which the input comes. This device, combined with triangulation of the matrix, brings the structure of interindustry transactions into graphic visibility.

A supplementary table displays, industry by industry, the capital stock employed; the employment of managerial, technical-professional, white-collar and blue-collar personnel; the energy consumption by major categories of fuel, and environmental stress measured by tons of pollutants.

The editors of *SCIENTIFIC AMERICAN* are happy to acknowledge the collaboration, in the preparation of this wall chart, of Wassily Leontief, originator of input/output analysis—for which contribution to the intellectual apparatus of economics he received the 1973 Nobel prize—and director of the Institute for Economic Analysis at New York University.

Packaged with the chart is an index showing the BEA and SIC code industries aggregated in each of the 97 sectors.

SCIENTIFIC AMERICAN

415 Madison Avenue, Dept. I/O, New York, N.Y. 10017

YES, send me the new *SCIENTIFIC AMERICAN* 1980's INPUT/OUTPUT wall chart—printed in eight colors on heavy paper measuring 65" x 52".

I understand my chart(s) will be shipped prepaid, unincreased*, flat-rolled and in a sturdy mailing tube. (Allow 4-6 weeks for delivery)

I enclose a check for \$ _____ for _____ chart(s) at \$35 each, saving me \$5 each in handling charges.

Bill me for \$ _____ for _____ chart(s) at \$40 each.

Name _____

Company _____

Street Address _____

City, State, Zip Code _____

ASA/81

*Charts shipped international require one fold.

Make check payable to *SCIENTIFIC AMERICAN*. Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

CATASTROPHIC RELEASES OF RADIOACTIVITY

LONG-TERM WORLDWIDE EFFECTS OF MULTIPLE NUCLEAR-WEAPONS DETONATIONS. Committee to Study the Long-Term Worldwide Effects of Multiple Nuclear-Weapons Detonations. National Research Council, 1975.

REACTOR SAFETY STUDY: AN ASSESSMENT OF ACCIDENT RISKS IN U.S. COMMERCIAL NUCLEAR POWER PLANTS (WASH-1400). U.S. Nuclear Regulatory Commission, National Technical Information Service, 1975.

THE EFFECTS OF NUCLEAR WEAPONS. Edited by Samuel Glasstone and Philip J. Dolan. United States Department of Defense and United States Department of Energy, 1977.

A UNIFIED THEORY OF ELEMENTARY PARTICLES AND FORCES

UNITY OF ALL ELEMENTARY-PARTICLE FORCES. Howard Georgi and S. L. Glashow in *Physical Review Letters*, Vol. 32, No. 8, pages 438-441; February 25, 1974.

WHY UNIFY? Howard Georgi in *Nature*, Vol. 288, No. 5792, pages 649-651; December 18/25, 1980.

SPEECH RECOGNITION BY COMPUTER

ON HUMAN COMMUNICATION: A REVIEW, A SURVEY AND A CRITICISM. Colin Cherry. The MIT Press, 1966.

TRENDS IN SPEECH RECOGNITION. Edited by W. A. Lea. Prentice-Hall, Inc., 1980.

THE ORIGIN OF GENETIC INFORMATION

SELF-ORGANIZATION OF MATTER AND THE EVOLUTION OF BIOLOGICAL MACROMOLECULES. Manfred Eigen in *Die Naturwissenschaften*, Vol. 58, No. 10, pages 465-532; October, 1971.

THE GENERAL PRINCIPLES OF SELECTION AND EVOLUTION AT THE MOLECULAR LEVEL. Bernd Küppers in *Progress in Biophysics and Molecular Biology*, Vol. 30, pages 1-22; 1975.

THE HYPERCYCLE: A PRINCIPLE OF NATURAL SELF-ORGANIZATION. PART A, EMERGENCE OF THE HYPERCYCLE; PART B, THE ABSTRACT HYPERCYCLE; PART C, THE REALISTIC HYPERCYCLE. Manfred Eigen and Peter Schuster in

Die Naturwissenschaften, Vol. 64, No. 11, pages 541-565, November, 1977; Vol. 65, No. 1, pages 7-41, January, 1978; Vol. 65, No. 7, pages 341-369, July, 1978.

PREBIOTIC EVOLUTION. Peter Schuster in *Biochemical Evolution*, edited by H. Gutfreund. Cambridge University Press, 1981.

THE SHELLS OF NOVAS

EXPLODING STARS. Robert P. Kraft in *Scientific American*, Vol. 206, No. 4, pages 54-63; April, 1962.

STRUCTURE AND EVOLUTION OF CLOSE BINARY SYSTEMS. Edited by Peter Eggleton, Simon Mitton and John Whelan. D. Reidel Publishing Company, 1976.

THEORY AND OBSERVATIONS OF CLASSICAL NOVAE. J. S. Gallagher and S. Starrfield in *Annual Review of Astronomy and Astrophysics*, Vol. 16, pages 171-214; 1978.

THE SHELL AROUND NOVA DQ HERCULIS 1934. R. E. Williams, N. J. Woolf, E. K. Hege, R. L. Moore and D. A. Kopriva in *The Astrophysical Journal*, Vol. 224, No. 1, Part 1, pages 171-181; August 15, 1978.

FILTER-FEEDING INSECTS

THE ECOLOGY OF RUNNING WATERS. H. B. N. Hynes. University of Toronto Press, 1970.

THE ROLE OF FILTER FEEDERS IN FLOWING WATERS. J. Bruce Wallace, Jackson R. Webster and W. Robert Woodall in *Archiv für Hydrobiologie*, Vol. 79, No. 4, pages 506-532; May, 1977.

AN INTRODUCTION TO THE AQUATIC INSECTS OF NORTH AMERICA. Edited by Richard W. Merritt and Kenneth W. Cummins. Kendall/Hunt Publishing Co., 1978.

FEEDING ECOLOGY OF STREAM INVERTEBRATES. K. W. Cummins and M. J. Klug in *Annual Review of Ecology and Systematics*, Vol. 10, pages 147-172; 1979.

FILTER-FEEDING ECOLOGY OF AQUATIC INSECTS. J. Bruce Wallace and Richard W. Merritt in *Annual Review of Entomology*, Vol. 25, pages 103-132; 1980.

ANCIENT OARED WARSHIPS

OARED FIGHTING SHIPS: FROM CLASSICAL TIMES TO THE COMING OF STEAM. Roger C. Anderson. P. Marshall, 1962.

GREEK OARED SHIPS, 900-322 B.C. J. S. Morrison and R. T. Williams. Cambridge University Press, 1968.

SHIPS AND SEAMANSHIP IN THE ANCIENT

WORLD. Lionel Casson. Princeton University Press, 1971.

ANOTHER PUNIC WRECK IN SICILY: ITS RAM. Lucien Basch and Honor Frost in *International Journal of Nautical Archaeology and Undersea Exploration*, Vol. 4, No. 2, pages 201-228; May, 1975.

THE HIGH SPEED CAPABILITIES OF ANCIENT BOATS. Sean McGrail and Ewan Corlett in *International Journal of Nautical Archaeology and Undersea Exploration*, Vol. 6, No. 4, pages 352-353; November, 1977.

ENGINEERING IN THE ANCIENT WORLD. J. G. Landels. University of California Press, 1978.

LITHIUM AND MANIA

LITHIUM SALTS IN THE TREATMENT OF PSYCHOTIC EXCITEMENT. J. F. J. Cade in *The Medical Journal of Australia*, Vol. 36, pages 349-352; 1949.

LITHIUM IN MEDICAL PRACTICE. Edited by F. N. Johnson and S. Johnson. University Park Press, 1978.

KINETICS AND STOICHIOMETRY OF NA-DEPENDENT LI TRANSPORT IN HUMAN RED BLOOD CELLS. B. Sarkadi, J. K. Alifimoff, R. B. Gunn and D. C. Tosteson in *Journal of General Physiology*, Vol. 72, No. 2, pages 249-265; August, 1978.

STUDIES ON LITHIUM TRANSPORT ACROSS THE RED CELL MEMBRANE, V: ON THE NATURE OF THE NA⁺-DEPENDENT LI⁺ COUNTERTRANSPORT SYSTEM OF MAMMALIAN ERYTHROCYTES. Jochen Duhm and Bernhard F. Becker in *Journal of Membrane Biology*, Vol. 51, No. 3/4, pages 263-286; 1979.

LITHIUM TRANSPORT IN HUMAN RED BLOOD CELLS: GENETIC AND CLINICAL ASPECTS. Ghanshyam N. Pandey, Elizabeth Dorus, John M. Davis and Daniel C. Tosteson in *Archives of General Psychiatry*, Vol. 36, No. 8, pages 902-908; July 20, 1979.

LITHIUM, MEMBRANES, AND MANIC-DEPRESSIVE ILLNESS. Barbara E. Ehrlich and Jared M. Diamond in *The Journal of Membrane Biology*, Vol. 52, No. 3, pages 187-200; 1980.


THE AMATEUR SCIENTIST

ON THE THEORY OF LONG WAVES AND BORES. John William Strutt, Baron Rayleigh, in *Scientific Papers: Vol. 6. 1911-1919*. Cambridge University Press, 1920.

RADIAL SPREAD OF A LIQUID STREAM ON A HORIZONTAL PLATE. R. G. Olsson and E. T. Turkdogan in *Nature*, Vol. 211, No. 5051, pages 813-816; August 20, 1966.

OPEN CHANNEL FLOW. Stephen Whitaker in *Introduction to Fluid Mechanics*. Prentice-Hall, Inc., 1968.

THE TIDES AS WE SEE THEM. Edward P. Clancy in *The Tides: Pulse of the Earth*. Doubleday & Company, Inc., 1968.



**RCA improves color TV
with an electronic
“fine tooth comb”
that separates
color from detail.
And enhances both.**

The Dynamic Detail Processor – a tiny solid state comb filter device with a computer-like “memory” – is producing sharper and clearer color TV pictures than ever before in RCA Limited Edition ColorTrak TV sets. By electronically separating the signals responsible for color and picture detail, the comb filter allows *both* signals to be fully processed by the set. As a result, minute details take on remarkable new definition and clarity, while color streaking and crawling edges (like those on checkered sport jackets) are virtually eliminated. RCA color TV sets are the first to have this solid state marvel of technology designed into them.

The Dynamic Detail Processor is a product of RCA Research – the kind of dedicated research that has been a tradition at RCA for over sixty years.

Today, many years after Nipper started listening to His Master’s Voice, research and development of electronic products are funded at the highest levels in our company’s history. Our commitment is as it has always been, to the advancement of technology – to the creation of imaginative, but practical, products that expand the human

horizon. ©1980 RCA Corporation

RCA

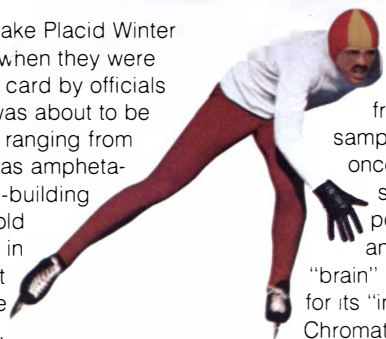
A tradition on the move.

Olympic lab gets results in record time

Athletes at the Lake Placid Winter Olympics knew when they were handed a green card by officials that their urine was about to be tested for drugs ranging from stimulants such as amphetamines to muscle-building steroids. With gold medals hanging in the balance, test results had to be fast and reliable.

Much of the drug screening was carried out by chromatography, a technique for separating mixtures of compounds so they can be identified and measured. To handle the heavy volume of test data, the Olympic laboratories were automated with a dozen Perkin-Elmer gas chromatographs linked to six Perkin-Elmer SIGMA 10 Chromatography Data Stations.

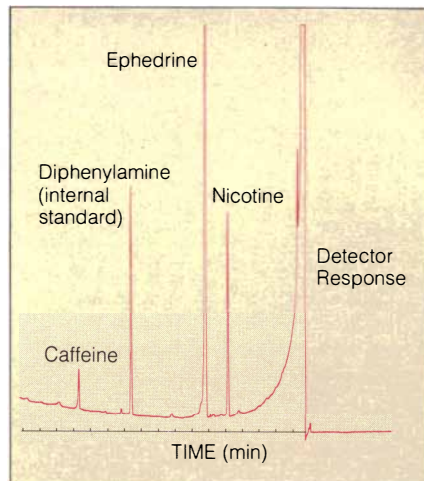
By bringing the power of the small computer into the laboratory,



the SIGMA 10 makes it possible to collect and analyze data from a number of samples in less time than it once took to perform a single analysis. With a powerful microcomputer and memory for its "brain" and versatile software for its "intelligence," a single Chromatography Data Station can guide the operator and handle up to four simultaneous analyses.

At Lake Placid, data stations controlled operation of the chromatographs, processed the data generated, and printed the final reports. In other environments, they also can function as part of a network of "distributed intelligence," trading data with a computer and interacting with other "intelligent" laboratory instruments.

Write to Perkin-Elmer for a free booklet about laboratory data handling and automation systems.



Chromatogram reveals presence of ephedrine, which dilates respiratory passages to increase oxygen intake. Other peaks indicate smoking and coffee drinking.

Data station automates trace metal search

*How much sulfur is in this coal?
Does this canned milk contain lead?
Is this well water too high in iron?*

To answer such tough questions, analytical chemists turn to atomic spectroscopy, a method of measuring traces of metals in concentrations as low as one part per billion.

The three most common atomic spectroscopy techniques are flame



atomic absorption (for rapid determination of up to six elements), graphite furnace atomic absorption (for trace and microsample analysis), and the newer inductively-coupled plasma emission (for multielement analysis and for elements that can't be readily determined by atomic absorption).

In the new ICP/5000 System, all three techniques are combined for the first time, automated by microprocessors and controlled by a Perkin-Elmer intelligent data station.

It can answer those tough questions by using the best analytical technique for each: inductively-coupled plasma emission for analyzing sulfur in coal, graphite furnace for measuring lead in canned milk, and flame for analyzing iron in well water.

Once optimum parameters have been recalled from memory by the analyst, the data station takes over. It selects wavelengths, calibrates, tells the operator when to analyze the sample, indicates any error, displays and prints the results.

With its Autosampler, the ICP/5000 System can also sequentially analyze as many as 50 samples for up to 20 elements automatically.

The results will tell you whether the coal is harmful to the environment, whether the milk is safe for your child, and whether you should drink the water.

For your copy of a free brochure on the ICP/5000 System, write to the address below.

Perkin-Elmer computers speed lab data to doctors

The clinical laboratory of Houston's M.D. Anderson Hospital and Tumor Institute, one of the largest cancer centers in the world, performed more than two million tests last year. Data from these tests and thousands more on file must be instantly available via CRT terminals to doctors throughout the hospital.

But as the volume of tests climbed over the two million mark, the computer in the clinical laboratory began to blink "overload." In a comprehensive reliability study to choose new computers, the hospital ran extensive equipment tests. Two Perkin-Elmer 32-bit "super-mini" computers were chosen on the basis of reliability, cost/performance factors and the ability to handle the hospital's growing needs.

Using system software developed by the hospital, lab data is input via 44 CRT terminals and 18 on-line instruments in the laboratory. The computers output data to the medical staff through 97



Medical team views terminal outside operating room to check patient's test file stored in lab computer.

CRT terminals throughout the hospital.

With 512 K bytes of memory each, the computers have increased total system memory tenfold. Capacity of the patient data file has been increased from 6,000 to 42,000 patients.

Despite this major upgrading in performance, the new equipment is easy for the staff to use: a few simple instructions are all that's required to enter or retrieve data from the computers.

For a free booklet about 32-bit computers, write to the address below.

For more information

If you would like to learn more about these Perkin-Elmer products for Computer Aided Chemistry, please write: Corporate Communications, Perkin-Elmer, Main Avenue, Norwalk, CT 06856.

PERKIN-ELMER

Responsive Technology

DEWAR'S® PROFILE

A thirst for living... a taste for fine Scotch.



MELVIN D. WILLIAMS

BORN: Pittsburgh, Pennsylvania.

HOME: West Lafayette, Indiana.

PRESENT TITLE: Professor of Anthropology and Director of Africana Studies and Research Center, Purdue University.

BIGGEST DISLIKE: Boredom. After 15 years of running his own business, Mel got tired of success. Against the advice of family and friends, he entered graduate school in anthropology. And now, just 7 years after getting his Ph.D., he is a full professor, as well as the author of many articles and books in his field.

POINT OF VIEW: "When there's no horizon to reach for, I have to back away and try something else. To me there's nothing more tedious than predictability."

SCOTCH: Dewar's® "White Label"® "I was introduced to Dewar's by an informant in the 'field' after a long day of interviews. I took it on the rocks then, and I still do."

Melvin D. Williams

BLENDING SCOTCH WHISKY • 86.8 PROOF
© 1980 SCHENLEY IMPORTS CO., N.Y., N.Y.